HOW TO IMPLEMENT SPEECH-TO-TEXT INTERPRETING IN LIVE EVENTS

Guidelines for making live events accessible

ILSA Interlingual Live Subtitling for Access

This project has been funded with support from the European Commission.

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

How to cite this document

Szczygielska, M., Dutka, Ł., Szarkowska, A., Romero-Fresco, P., Pöchhacker, F., Tampir, M., Figiel, W., Moores, Z., Robert, I., Schrijver, I., Haverhals, V. (2020) *How to implement speech-to-text interpreting (live subtitling) in live events. Guidelines on making live events accessible*. ILSA Project. **http://ka2-ilsa.webs.uvigo.es/guidelines**

INTRODUCTION

Live events, such as conferences, sports and cultural events, can be translated and made accessible thanks to speech-to-text interpreting, also referred to as live (sub)titling. There are different methods for producing live (sub)titles, for example with keyboards or with different degrees of automation (automatic speech recognition, machine translation).
The ILSA project and these guidelines focus on intralingual and especially **interlingual live (sub)titling** or **speech-to-text interpreting through respeaking**, where a speech-to-text interpreter known as an interlingual respeaker (or transpeaker) translates what is being said (dictating also punctuation marks) to a speech recognition software, which displays the translated words as (sub)titles on a screen.

This document is addressed to live event organisers, accessibility managers and anyone who wants the event they organise to become more accessible.

The aim of this document is to help you add live titles to your event in order to provide the maximum possible benefit to your audience.

Different terms are used to talk about the provision of live subtitling and speech-to-text services in different countries and markets. In the ILSA project, we attempt to clear up the current terminological confusion.
The table below provides the major terms and their definitions, which will be used throughout this document and in all the ILSA guidelines.

DEFINITIONS

BOOTH – sound-proof space (as used, with standardized specifications, for simultaneous interpreting)

INTERLINGUAL RESPEAKING – see 'Transpeaking'

INTRALINGUAL RESPEAKING – see 'Respeaking'

LIVE SUBTITLES – live titles displayed on a screen underneath or as part of the image

LIVE TITLES – written text produced by speech-to-text interpreting

LIVE TITLING – see 'Speech-to-text interpreting'

RESPEAKING – a method of creating live titles using speech recognition technology, whereby a person known as a respeaker repeats and/or paraphrases what is being said, dictating also punctuation marks (respeaking is also known as voice writing)

SPEECH-TO-TEXT INTERPRETING (STTI) – the production of a written version of a spoken message while it is being delivered. It is mainly performed intralingually (in the same language), using keyboards or respeaking, for the benefit of people with hearing loss, but can also be done interlingually. In different countries and contexts, speech-to-text interpreting is also referred to as live subtitling or captioning.

STENOMASK – sound-insulating cover around a microphone used by respeakers to muffle speech sounds (see Fig. 5)

TRANSCRIPT – written text representing a spoken message

TRANSPEAKING – a method of creating interlingual live titles using speech recognition technology, whereby a person known as a interlingual respeaker (or transpeaker) translates what is being said, also dictating punctuation marks

USERS

Speech-to-text interpreting will benefit many different types of live event participants, in particular:

- **participants with hearing loss**;
- other people who may experience difficulties accessing the content of talks, such as people with autistic spectrum disorders or people with dyslexia;
- **international attendees** who do not know the language of the talk well enough and can benefit from interlingual (different language) or intralingual (same language) titles delivered to them in real time;
- **hearing attendees** and the **audience at large**, who may refer to subtitles if they missed something or to check the spelling of terms.

For intralingual speech-to-text interpreting, the main target group are people who are deaf or hard of hearing. Interlingual speech-to-text interpreting is addressed to foreign participants who do not know the language of the talk as well as those who are deaf or hard of hearing.

Live titles can also be added to an online streaming of the event, this way reaching a wider audience. Some people might want or need to watch your event in noisy environments or in a context where they cannot watch the talk with the sound on. Live titles will allow them to follow your event in real time.

BENEFITS

- Live titles increase overall accessibility of and foster inclusion at your event.
- All participants may benefit from live titles as they help them to focus on the content of the talk.
- In live online streaming, adding titles will mean that people can follow the event without turning on the sound.
- Online videos on social media are often watched without sound. If you add titles to your online videos, you are likely to reach a larger number of viewers.
- If you record your events and make the recordings available afterwards, live titles can be re-synced to be used as subtitles for the video. This way the subtitles do not need to be prepared from scratch, which minimizes the cost of subtitling the video and also means that the subtitled video can be published faster.

WORKFLOWS

There are two possible types of workflows in terms of the location where speech-to-text interpreting with respeaking is done:

1. **On-site respeaking:** the respeaker works in the same room, preferably in a booth, or using a stenomask.
2. **Remote respeaking:** the respeaker works in a place different from the conference venue (for instance, from home) or in a different room.

Figures 1 and 2 below show two models of on-site respeaking.

Figure 1. Respeaker in a booth. (©Dostępni.eu)

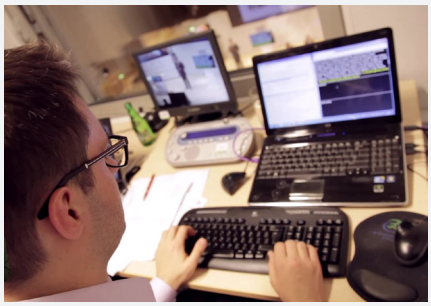


Figure 2. Respeaker and editor in a booth. (©Dostępni.eu)

The booth is sound-proof so that the respeakers can speak to the microphone without the interference of external sounds. Respeaker and editor can work in the same booth, which allows for easy communciation. They will see the event through the glass but both respeakers and editors need to have headphones with audio from the room and speakers at the event need to speak to the microphone.

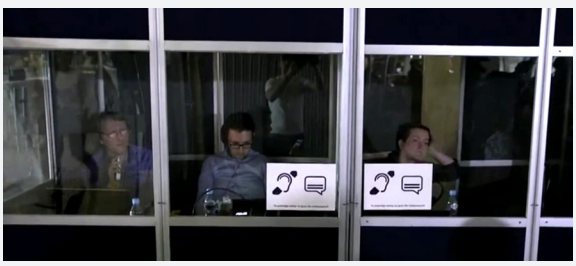


Figure 3. A team of intra- and interlingual respeakers and editors in a booth (©Dostępni.eu)

Two or more booths can be used if speech-to-text interpreting is provided in more than one language (Fig.3).

Figures 4 and 5. Remote speech-to-text interpreting. (©Dostępni.eu)

A separate, quiet room can be used instead of a booth (Fig. 4 and 5). Respeakers and editors need to receive video and audio from the event. While they might be working next door to the event, technically this does not differ much from working remotely in another location. However, working in the venue allows for transmitting audio and video locally (through cables in the building) as opposed to through the Internet and this might be preferable or easier for some events.

Figures 6 and 7 show models of on-site and remote respeaking.

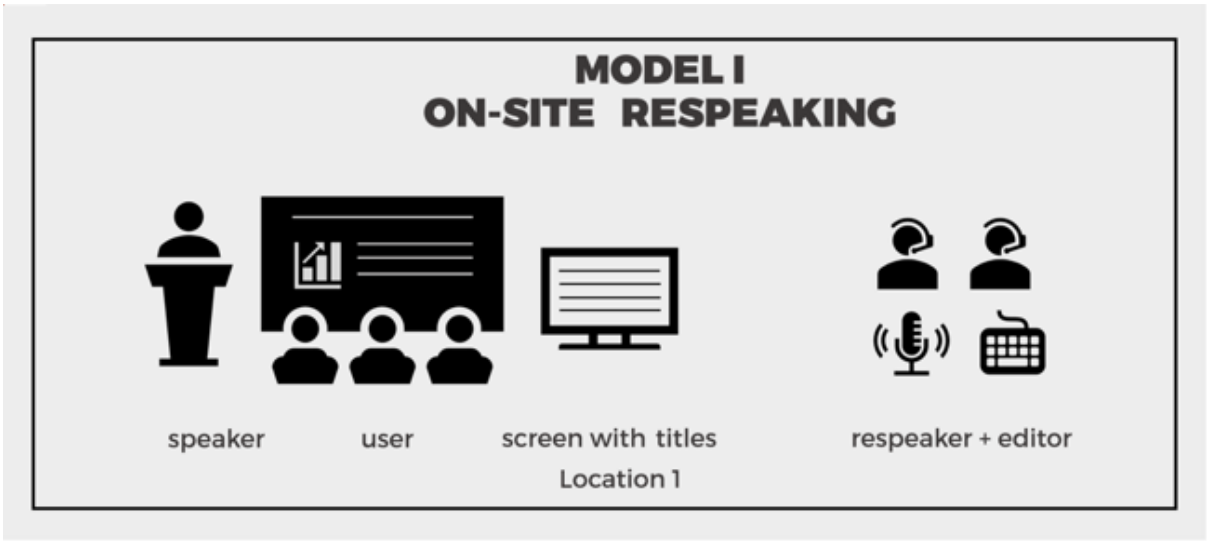


Figure 6. Model of on-site respeaking

When respeaking is done on-site, the respeaker and editor work in the same room and their workstations are connected through a cable or a wireless connection with a screen or screens in the venue that display live titles. Respeakers and editors can easily collaborate as they are sitting next to each other.

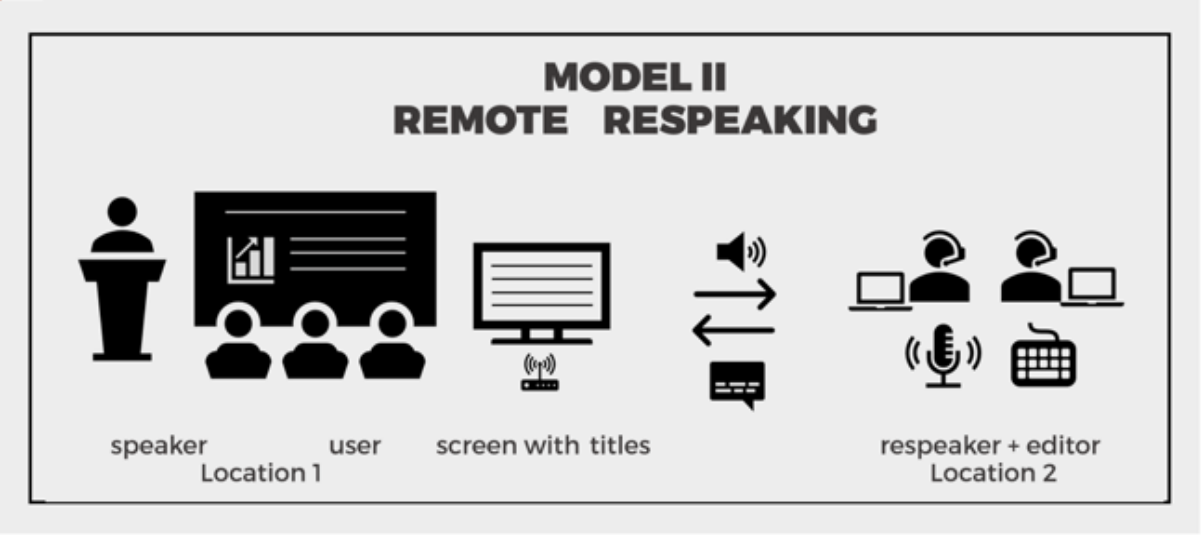


Figure 7. Model of remote respeaking

When working remotely, respeakers and editors can work together in a separate location. They can also work from home. This requires editors to have a second audio connection so that they can hear the voice of the respeaker. As for longer events there's more than one respeaker and one editor, they need to have a tool that allows them to communicate with each other and manage turn-taking.

Figures 8-11 show different models of intra- and interlingual workflows.

Figure 8. Live titling – intralingual workflow.

When working in the same language as the language of the event, the respeakers usually work with editors who correct errors in speech recognition (version A in Figure 8).
For slow-paced events and in languages where there are few errors in speech recognition, it might be possible for respeakers to self-correct their work (version B).

In the case of interlingual respeaking, several workflows are possible:

Figure 9. Live titling – interlingual workflow I with interpreter & intralingual respeaker

The workflow shown in Figure 9 requires the interpreter to hear, understand and translate first (oral language A into oral language B) before the respeaker can start respeaking (oral language B into written language B). This can create more delay but, on the other hand, the respeaker does not need to know the original language.

Figure 10. Live titling – interlingual workflow II with interlingual respeaker

Using interlingual respeakers (Fig. 10) minimizes the delay with which the live titles are displayed to users. However, it might be challening to find interlingual respeakers in some language combinations.

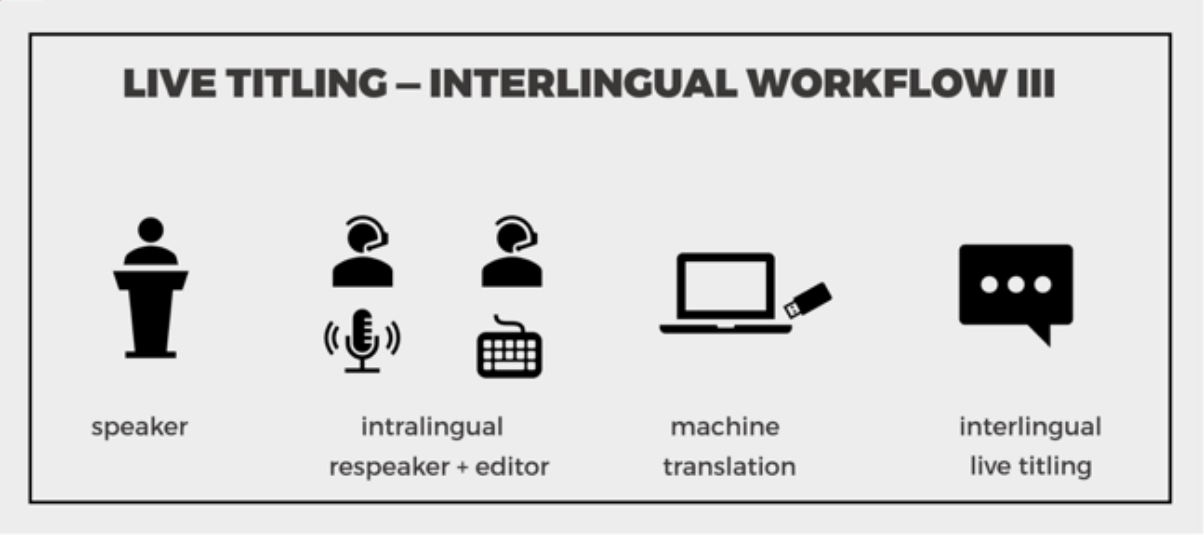


Figure 11. Live titling – interlingual workflow III with intralingual respeaker & machine translation

Using machine translation (Fig. 11) will allow live titles to be translated to multiple languages at once. Depending on the language pair, the quality of machine translation might be inadequate. The results of the machine translation output may improve if the respeakers use simple language and simplify sentence structure.

CORRECTION

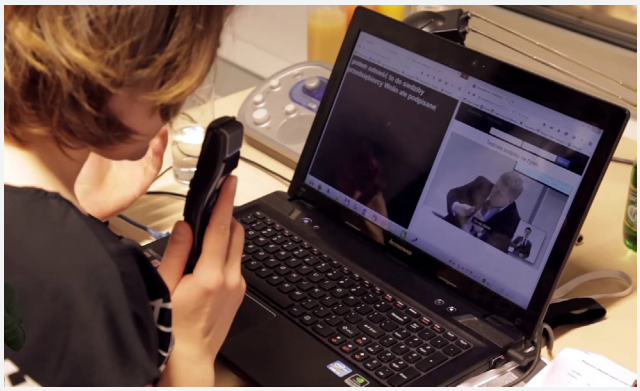
As speech recognition is hardly ever 100% accurate, the text produced through respeaking or transpeaking needs to be corrected before it is displayed as live titles. Depending on the difficulty of the topic and the number of misrecognized words, correction can be done by the respeaker or by a live editor.

1. self-correction by the respeaker
2. parallel correction by a live editor

Self-correction can sometimes be used when events are slow-paced and there are not too many errors to correct, especially in languages such as English, where speech recognition works well and produces few misrecognized words. The respeaker can stop respeaking, correct the error and start respeaking again.

Self-correction is not usually possible when transpeaking as this task is more difficult. Self-correction can also be very hard if the speech is fast or includes complex terminology. In languages where speech recognition is less developed or in flectional languages with a high number of different word forms, self-correction is not usually possible.

Parallel correction by a live editor will produce the best results, especially for fast-paced talks with complex terminology. It requires another professional, which impacts on the cost of the service but also allows for better quality and (almost) error-free live titling.



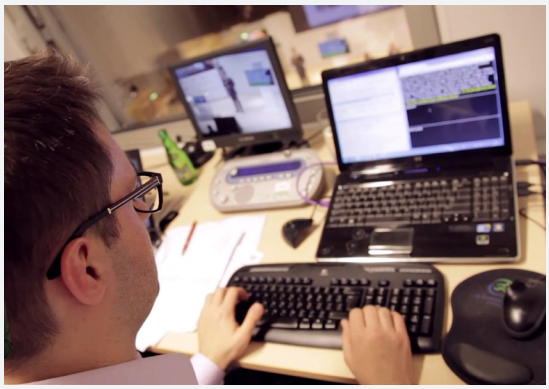


Figure 12 and 13. Respeaking with parallel correction. (©Dostępni.eu)

GOOD PRACTICES

- It is recommended that speech-to-text interpreting should be done with correction.
- The transcript of live titles can be later used as the basis for the preparation of the minutes or the transcript of the meeting. It is recommended that it should be corrected by the respeaker after the event.
- As respeaking requires high cognitive effort (and even more so when using a stenomask), it is difficult to sustain focus longer than 20-30 minutes without a break. It is recommended to include breaks for respeakers or have a team of respeakers who can take turns. An intralingual respeaker should not work for longer than 45 minutes before a colleague takes over. In interlingual respeaking, individual turns should not be longer than 30 minutes.

LIVE TITLING TOOLBOX

Depending on your needs, budget and technical infrastructure, you can choose from a number of solutions.

Speech recognition software	
desktop-based solutions (installed locally on your computer)	**cloud-based solutions** (available remotely, in the cloud)
–Dragon Professional –Newton Dictate	–Google Speech API –Microsoft Speech API –Amazon API –Dragon Anywhere

Live titling software

- Text on Top – software for producing and displaying live titles, which allows for cooperation between a number of workstations through wireless connection (using USB dongles)
- StreamText
- Various accessibility services providers have their own proprietary platforms or tools, some of which might allow for remote work.

How to display live titles

- Live titles can be displayed on one or more additional large screens in the room or the conference hall
- Alternatively, live titles can be displayed as subtitles, i.e. as two lines of text on the main screen, below the presentation
- Another option is for the audience to display live titles on their own devices or on devices provided by the event organisers, such as laptops, tablets or smartphones (this lets them customize the look of the text on their devices).

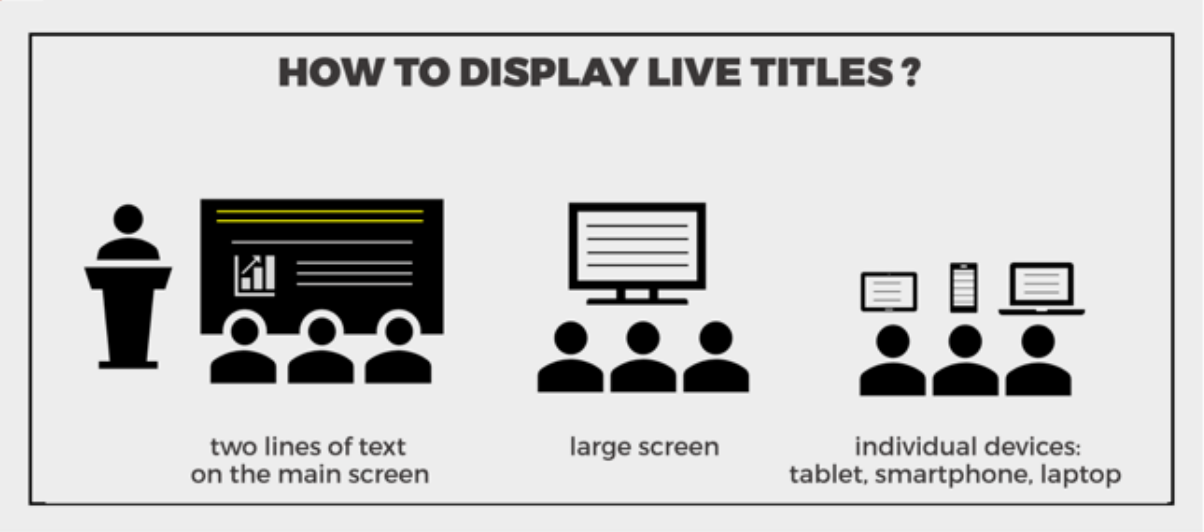


Figure 14. Various methods of displaying live titles

Figures 15, 16 and 17. Live titles displayed on LCD monitors in the conference venue. (©Dostępni.eu)

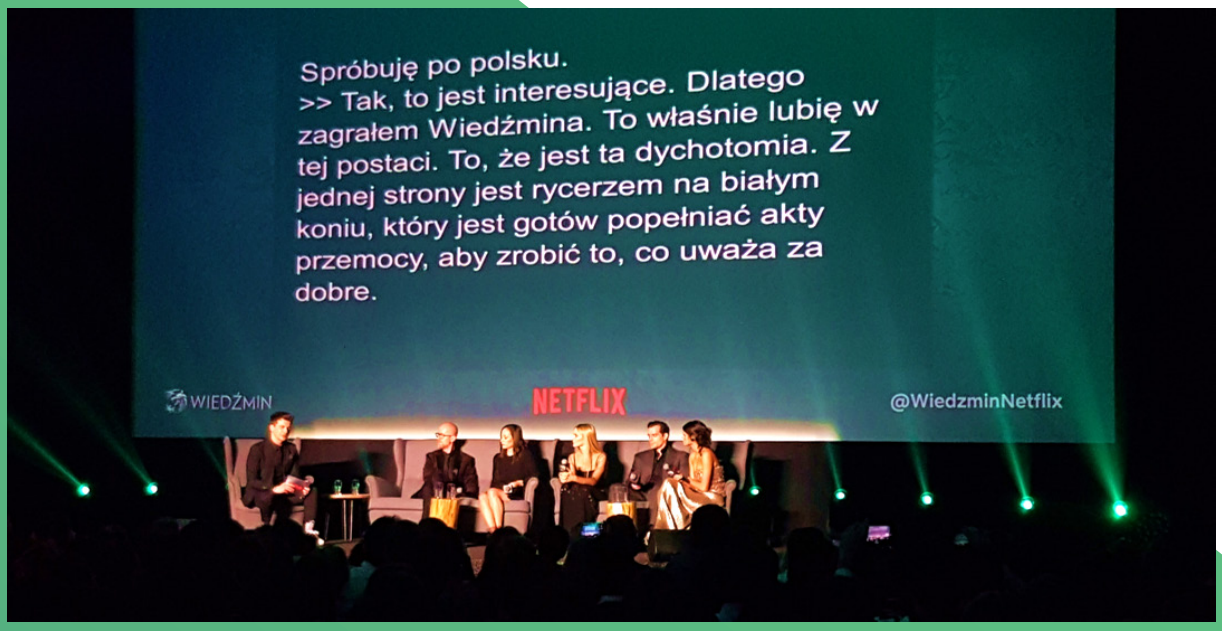


Figure 18. Live titles displayed on a cinema screen. (©Dostępni.eu)

Figure 19. Live titles displayed on individual user's mobile device in a dedicated app. (©Dostępni.eu)

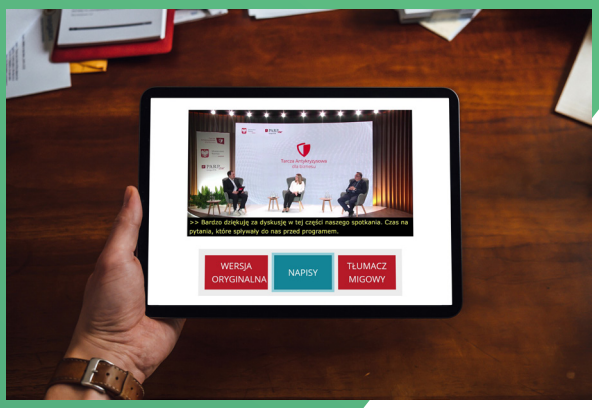


Figure 20. Live titles displayed on an individual user's mobile device in online streaming. (©Dostępni.eu)

CHALLENGES AND SOLUTIONS

- **Sound** — if the respeaker works remotely, the sound can be transmitted via a telephone, a messaging app or a video conferencing solution.
- **Display** — if you cannot provide a screen at the venue of the event, live titles can be displayed in an app or on a website which a user will open on their mobile device.
- **Space for respeakers to work** — if there are no sound-insulated rooms available for respeakers at your venue and you cannot add a booth to your venue or there is no way to transmit audio from the event, respeakers can sit in the main room so that they hear the speaker without any headphones and cable and they can use stenomasks so as not to disturb other participants.

DOs and DON'Ts	
DOs	**DON'Ts**
Let the users customize the look of the titles on their own devices by choosing the font, the size and colour of the text and the colour of the background	Do not use serif fonts to display titles
If users cannot customize the look of titles, make sure the titles are displayed in high contrast, left-aligned and the text is big enough so that it is legible	Do not justify text or use italics (as it makes it less legible)
If you have deaf participants, make sure that they are happy with live titles. Some deaf people will prefer a sign language interpreter	Do not make a single respeaker work for the whole day. Like interpreters, for longer events respeakers should take turns/breaks

CHECKLIST FOR THE ACCESSIBILITY MANAGER

- [ ] Find out what are the needs of your participants:
  - Do deaf participants prefer live titles or sign language interpreting
  - If both live titles and sign language (SL) interpreting are required, consider where the SL interpreter will be located, so that users have a clear view of both the titles and SL interpreter.
  - Do foreign attendees need a translation into their mother tongue or into English?
- [ ] Check if your venue infrastructure and Internet connection are sufficiently robust to provide the service remotely. This is also required when respeakers are working in a separate room in the venue building but not in the same room where the event is taking place.
- [ ] Choose whether live titles will be done remotely or on-site.
- [ ] Choose the language of live titles to be provided.
- [ ] Contact the speaker(s) to discuss the provision of the live titling service (share with the speaker the Guidelines for Speakers that you can find below).
- [ ] Ask the speaker(s) to provide preparatory materials in advance.
- [ ] Inform the participants that live titles will be available at the event and let them know how to access them.
- [ ] If you make a recording of your event available later, make sure it is accompanied by subtitles.
- [ ] Provide information about access services together with general information about the event.
- [ ] If you organise events with live titles on a regular basis, provide a mechanism for quality assurance and feedback from users.

GUIDELINES FOR SPEAKERS

There are a few things that you can do as a speaker to improve the experience for the audience who is accessing your talk with live titles:

- ☐ Make sure you have not prepared too many slides for your talk, which might mean that you would have to rush through them.
- ☐ Make your presentation available to the event organisers in advance (a day earlier, at the latest) so that they can share it with respeakers; preparation is vital for respeakers and the more they know about your talk, the better the live titles they can produce, and the higher the chance that your ideas reach the audience.
- ☐ Alongside the presentation, and especially if you are not planning to show any text during your talk, share with respeakers the key themes you will be talking about, numbers, and names of people or places you will refer to as well as specialised terminology that you want to use. The respeakers will add these terms to the speech recognition software wordlist. This way the terms are more likely to be correctly recognized during the event.
- ☐ As humour might be hard to capture in live titles and it is easy for part of the audience to feel excluded if they do not understand why others are laughing, let the respeakers know beforehand if you are planning to include jokes or stories in your talk so that they can make sure every member of the audience enjoy your joke equally.
- ☐ During the presentation, avoid changing slides too quickly as members of the audience will be reading live titles alongside the visual content you are talking about and live titles appear with some delay. Allow extra time for the audience to read titles and and then look at the graphics or written content you are sharing with them.
- ☐ If a microphone is provided, make sure you use it all the time as respeakers rely on clear audio.
- ☐ If there are comments or questions from audience members who do not have a microphone, repeat them to the microphone as respeakers are only able to respeak what they can hear.

CHECKLIST FOR THE RESPEAKER

- [ ] Reach out to event organisers to receive reference materials such as the schedule of the event, speakers' presentations and other materials that will be displayed or shared at the event.
- [ ] If needed, educate the organisers about the process of producing live titles: let them know that you need reference materials to do your job well, and make sure they understand that live titles will be appearing with a slight delay.
- [ ] Read through the materials provided and research the topic of the event so that you have a grasp of the ideas that will be discussed and the terminology that might be used.
- [ ] Familiarise yourself with specialised vocabulary and practise repeating it.
- [ ] Update the speech recognition software's lexicon so that you are sure that the main terms and names are included in the lexicon and will be recognized correctly.
- [ ] Look through the names of the organizers, the speakers as well as any important guests that might be mentioned. Make sure you know how to spell these names correctly.
- [ ] Before the start of the event, check if you can hear the speaker and adjust the volume of the sound.
- [ ] Stop your work and intervene or ask others to intervene on your behalf if someone is not speaking to the microphone or the audio quality deteriorates significantly.
- [ ] After the event, save all the text as a transcript in case it is needed later.

GOOD PRACTICES FOR LIVE ONLINE STREAMING

Live events are often streamed online so that they can reach a wider audience. When speech-to-text interpreting is provided in the venue, it is a good practice to also make it available as a part of the online streaming so that people following the event online can also benefit from this service. Below you can find some recommendations for what to do to make live streaming accessible.

- For technical reasons, live streaming is always delayed and the delay can be controlled. Adjust the delay of the stream to improve the synchronization of the text with the audio.
- Depending on the video player you are using, it might be possible to make the text available in the form of closed subtitles that can be turned on or off. This is a recommended option as closed subtitles are more accessible and customizable.
- If the video player you have to use does not support closed subtitles, you can included the text in the form of open subtitles, which are technically delivered not as text but as a part of the image. As a result, assistive technologies such as screen readers cannot access those subtitles. It is also impossible for users to make the font bigger.
- If you are providing interlingual subtitles and spoken language interpreting as two options at your event, do not include them in the same online stream as seeing and hearing different text makes it more difficult for users with hearing loss to follow the event. It is preferable to provide two separate streams for users to choose from.

QUALITY IN SPEECH-TO-TEXT INTERPRETING

Although there are different ways to assess the quality of live (sub)titles produced by speech-to-text interpreting, the most common method used is the NER model (Romero-Fresco & Martínez, 2015), which makes a distinction between recognition errors (caused by the interaction between the respeaker and the software) and edition errors (caused by the respeaker's incorrect decisions when omitting or changing information). These errors can in turn be minor, standard or serious depending on how they impact on viewers' comprehension. The NER model is currently being used by governmental regulators, broadcasters and universities in countries such as Spain, the UK, Belgium, Poland, Switzerland, South Africa, Australia, the US and Canada, where it has been included within the national accessibility legislation to assess live subtitling quality. More information about the NER model can be found on the website of the Galician Observatory for Media Accessibility as well as in the ILSA course.

The NER model is used for the assessment of intralingual live subtitles (those produced in the same language as the original audio). The assessment of interlingual live subtitles (those translating the original audio into another language) may be done with the NTR model (Romero-Fresco & Pöchhacker, 2017).

MORE INFORMATION

Moores, Z. (2020). Fostering access for all through respeaking at live events. *The Journal of Specialised Translation*, 33, 176-211. https://www.jostrans.org/issue33/art_moores.php

Moores, Z. (2020). Respeaking at Live Events – Information for Speakers http://galmaobservatory.webs.uvigo.es/projects/respeaking-at-live-events/

Pöchhacker, F., &Remael, A. (2019). New efforts?: A competence-oriented task analysis of interlingual live subtitling. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18, 130–143 https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/515/471

Robert, I. S., & Remael, A. (2017). Assessing quality in live interlingual subtitling: A new challenge. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 168-195.

Romero-Fresco, P. (2011). *Subtitling through speech recognition: respeaking*. Manchester: St. Jerome Publishing.

Romero-Fresco, Pablo and Eugeni, C. (2020). Live subtitling through respeaking. In Łukasz Bogucki and Mikołaj Deckert (Eds.), *Handbook of Audiovisual Translation and Media Accessibility*, Palgrave.

Romero-Fresco, P., & Martínez, J. (2015). Accuracy Rate in Live Subtitling: The NER Model. In J. Díaz-Cintas & R. Baños Piñero (Eds.), Audiovisual Translation in a Global Context. Mapping an Ever-changing Landscape (pp. 28-50). London: Palgrave.

Romero-Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 16, 149–167. https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/438

Szarkowska, A., Dutka, Ł., Krejtz, K., & Pilipczuk, O. (2017). Respeaking crisis points. An exploratory study into critical moments in the respeaking process. In M. Deckert (Ed.), Audiovisual Translation – Research and Use (pp. 179-201). Bern: Peter Lang.

Szarkowska, A., Dutka, Ł., Szychowska, A., & Pilipczuk, O. (2018). Visual attention distribution in intralingual respeaking. In C. Walker & F. Federici (Eds.), Eye Tracking and Multidisciplinary Studies on Translation (pp. 185-201). Amsterdam/New York: John Benjamins.

Szczygielska, M., & Dutka, Ł. (2019). Historia napisów na żywo tworzonych metodą respeakingu w Polsce [The history of live subtitling through respeaking in Poland]. In K. Hejwowski, K. Dębska, & D. Urbanek (Eds.), Tłumaczenie wczoraj, dziś i jutro (pp. 129-164). Warsaw: Institute of Applied Linguistics University of Warsaw.

ILSA Interlingual Live Subtitling for Access

This project has been funded with support from the European Commission.

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

universität wien · Universidad de Vigo · vrt · Universiteit Antwerpen · dostępni.eu · University of Warsaw

With the support of the Erasmus+ Programme of the European Union