

Speech-to-Text, Text-to-Speech, Subtitling and Reading Support

Author: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Reviewer: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Learning objective

Converting speech into text and text into speech are among the most mature and useful applications of AI in education. The aim of this material is to show how speech-to-text, text-to-speech, subtitling and reading-support tools can make learning materials more accessible and usable in everyday teaching. By the end of the material, teachers will be able to choose and use these tools for learners with different sensory, linguistic, literacy or concentration needs, create captions for videos, audio versions of texts and transcripts of explanations, prepare simplified reading support, and check the quality of the results.

Introduction: one content, several channels

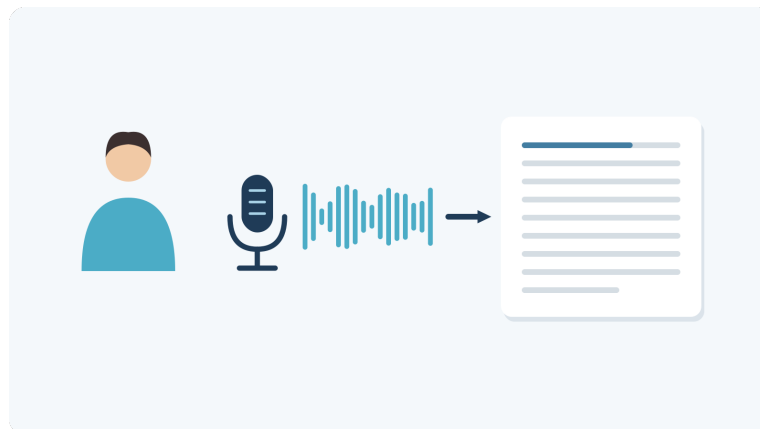
Most learning content reaches learners through a single channel: a printed text, a spoken explanation or a video. Every single channel excludes someone – the learner who cannot hear the explanation, cannot read the dense text quickly enough, or loses concentration halfway through. AI-based speech and reading tools make it easy to offer the same content through several channels at once: text that can also be listened to, speech that can also be read, video that also has captions. This is the easiest way to put the principle of "**multiple means of representation**" into practice.

UDL stands for Universal Design for Learning. It is a pedagogical framework based on the idea that there is no "average learner," so teaching should be designed from the outset to be accessible to as many people as possible, rather than being adapted to individual learners after the fact.

Speech-to-text

Speech recognition has become accurate enough – in most languages – to be used every day:

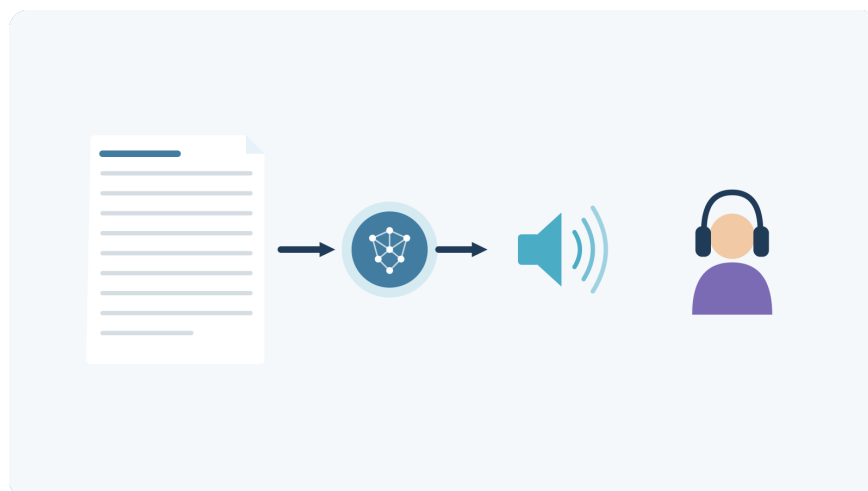
- **Dictation** is built into Windows (Win+H), macOS, Android and iOS keyboards, Google Docs and Word, and into AI assistants such as ChatGPT. Learners with writing difficulties or motor impairments can express their knowledge by speaking.
- **Transcription** of recordings or live speech: Live Transcribe on Android, the Transcribe function of Word, and meeting transcripts in Teams or GoogleMeet. Online services convert an hour of recorded speech into text within minutes at low cost.
- **Transcripts of explanations** give learners a text they can re-read, search and translate – useful for learners who are hard of hearing, who were not present, or who do not learn in their native language.



Speech-to-text turns spoken explanations into editable text

Accuracy drops with background noise, overlapping speech, dialects and specialised terminology. Speak clearly, use an external microphone where possible, and correct the key terms before sharing a transcript. Recording people requires their consent.

Text-to-speech



Text-to-speech turns any text into a natural-sounding audio version

Modern neural voices sound natural and are available in many languages, including Hungarian:

- **Read aloud** in Microsoft Edge and Immersive Reader in Word and OneNote read documents and web pages aloud and highlight the words as they are read.
- **Phone accessibility functions** (Select to Speak on Android, Speak Screen on iOS) read whatever is on the screen.
- **Voice generation services** such as ElevenLabs create high-quality narration for audio versions of learning materials.

Audio versions support learners with dyslexia or visual impairment, learners reading in a second language, and anyone who prefers listening – for example while travelling to a work placement. Listening while reading (“bimodal” reading) can improve both comprehension and stamina.

Captions and subtitles



Captions make video explanations accessible – and useful in a noisy workshop

Captions are essential for deaf and hard-of-hearing learners and are required for pre-recorded video by the **Web Content Accessibility Guidelines** (WCAG). Captions also help learners who watch in a noisy environment, who don't learn their native language, or need to see how a technical term is spelled.

- YouTube, Microsoft Stream and most video tools generate **automatic captions**; operating systems and browsers offer **live captions** for any audio.
- Automatic captions must be **checked and corrected**, especially for technical terms, numbers and safety warnings.
- Captions can be **translated** automatically, which supports learners with a different first language.
- Prefer closed captions that learners can switch on and off and resize, rather than captions burned into the image.

Reading support



Reading modes reduce visual clutter, highlight the current line and can read the text aloud

Reading-support tools reduce the effort of decoding and help learners focus on meaning:

- **Reading modes** (Immersive Reader, browser reading view) remove clutter and let users change font size, spacing, background colour and line focus; some can split words into syllables or show pictures for common words.
- **AI simplification and summaries:** a language model can rewrite a text at a lower reading level, split it into short sections, add headings and a summary – keeping the technical terms and explaining them.
- **Glossaries** of the key terms of a unit, with simple explanations and pictures, help all learners, especially those who read in a second language.

Recipe: one worksheet, several channels (about 20 minutes)

1. Check the structure: clear headings, short paragraphs, lists instead of long sentences.
2. Create an audio version with Read Aloud or a text-to-speech tool.
3. Ask the AI for a simplified version and a glossary of 8–10 key terms – then check both.
4. Add captions to the demonstration video and correct the technical terms.
5. Offer all versions to the whole group, so that learners can choose.

Limits and good practice

- AI transcripts, captions and simplifications contain errors – the teacher remains responsible for the accuracy of what is shared.
- Voice recordings and documents may be processed in the cloud: follow data protection rules and do not upload personal data. For sensitive material, tools that run locally or institutional accounts are safer choices.
- Support tools do not replace the teaching of reading and writing; they give access to content while those skills develop.
- Offer choice rather than prescribing: learners know best which format helps them.

Takeaways for teachers

- Offer the same content through several channels: text, audio and captioned video.
- Dictation and transcription help learners who find writing or listening difficult.
- Human like voices make audio versions of any text quick to produce.
- Always check automatic captions and transcripts, especially technical terms and safety information.
- Reading modes, simplified versions and glossaries reduce the effort of reading and support understanding.

References

- W3C (2023): Web Content Accessibility Guidelines (WCAG) 2.2 – Success Criterion 1.2.2 Captions (Prerecorded).
- CAST (2024): Universal Design for Learning Guidelines version 3.0.
- Microsoft Support: Immersive Reader and Read Aloud – languages and voices.

- Lengyel, Zs. (2026): Concrete applications for content generation. VITA training presentation.

Links

- Immersive Reader languages and voices: support.microsoft.com
- Live Transcribe & Sound Notifications: play.google.com
- Captions – W3C: w3.org/WAI/media/av/captions
- UDL útmutató: udlguidelines.cast.org
- W3C Web Accessibility Initiative: w3.org/WAI

Copyright © 2026 VITA Consortium

This work may be reused and adapted under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).



Images that don't have a source specified were generated by Claud for Mac.

Funded by the European Union. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the National Agency.