

The ChatGPT Voice Module

Author: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Reviewer: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Learning objective

Talking is the most natural way of communicating – and AI assistants can now hold spoken conversations in real time. The aim of this material is to introduce voice-based AI interaction and its possible role in teaching and learning. By the end of the material, teachers will be able to distinguish dictation from voice conversation, describe how the current ChatGPT Voice works, identify meaningful uses for brainstorming, language practice, accessibility, quick professional assistance and reflective dialogue, and take into account practical limitations such as transcription errors, classroom noise and data protection when using voice with learners.

Introduction: two ways of using voice

ChatGPT can be used by voice in two practical forms. **Dictation** simply converts our speech into text: we speak, the system transcribes, and we send the message as if we had typed it. This is useful when we want to formulate longer thoughts quickly or typing is inconvenient, for example on a phone.

Voice conversation is a real dialogue: we speak to the AI, and it answers aloud. This allows a faster, more natural interaction – especially for brainstorming, practising or when we do not want to type continuously. The key difference: dictation turns speech into text, while voice conversation is a conversational interface in which the AI's answer also arrives as speech.

How ChatGPT Voice works today

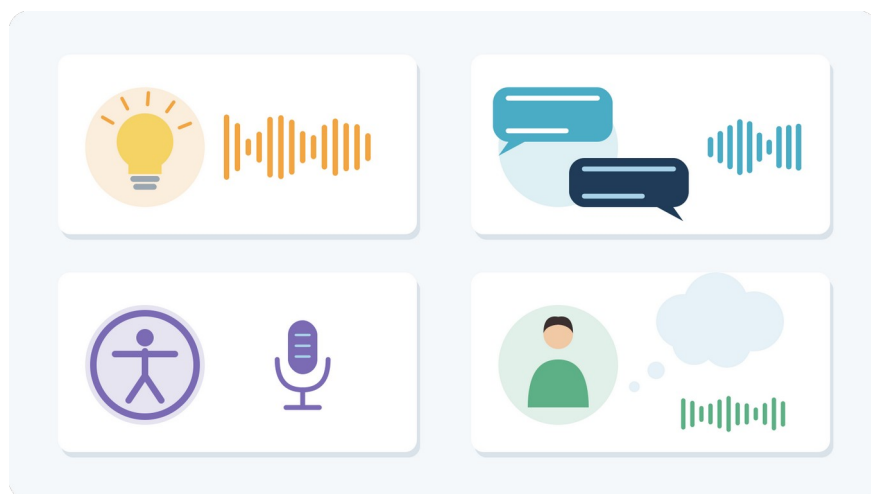
Voice has developed quickly. Since late 2025 voice conversations take place directly inside the chat, so the spoken exchange and any visual content (for example a map, a table or an image) appear in the same conversation. In July 2026 OpenAI introduced **GPT-Live**, a new generation of voice models that can listen and speak at the same time. This makes turn-taking and interruptions feel natural: we can cut in, the model can signal that it is listening, and the conversation flows more like a human dialogue.



Full-duplex voice conversation: the user and the AI can speak and listen at the same time

- **Live mode** (the default for most users) supports natural back-and-forth conversation and can use web search and memory; for complex reasoning or research it can hand over the task to a stronger text model while the conversation continues.
- **Advanced mode** keeps earlier real-time capabilities, including sharing the camera or screen on mobile devices – useful for showing an object, a device or a problem.
- Voice can also be used within **Projects**, so the spoken conversation can rely on the project's files and instructions.
- After the conversation, a **transcript** appears in the chat history. It may not exactly match what was said, especially with overlapping speech, background noise or fast conversation.
- Sessions have usage limits that depend on the subscription. The language of speech recognition can be set in the settings, which improves accuracy.

Possible roles in teaching and learning



Brainstorming, language practice, accessibility and reflective dialogue – four typical uses of voice interaction

Brainstorming and planning on the go

Many teachers have their best ideas while commuting or walking. A spoken conversation is a quick way to think aloud about a lesson: “I teach tyre changing to first-year apprentices tomorrow. Help me find a motivating opening.” Afterwards, ask for a written summary in the chat and edit it.

Language practice

Voice is especially valuable for practising professional communication in a foreign language: a receptionist greeting a hotel guest, a mechanic explaining a repair to a customer, a waiter taking an order. The AI can play the role of the partner, speak more slowly or faster, repeat, and give feedback on vocabulary. Feedback on pronunciation should be treated with caution; the teacher remains the reference.

Accessibility

Voice interaction can lower barriers for learners with reading difficulties, visual impairment or motor difficulties: they can ask questions and receive

explanations without reading or typing. For these learners, voice can be a genuine support – within the school's framework for assistive technology.

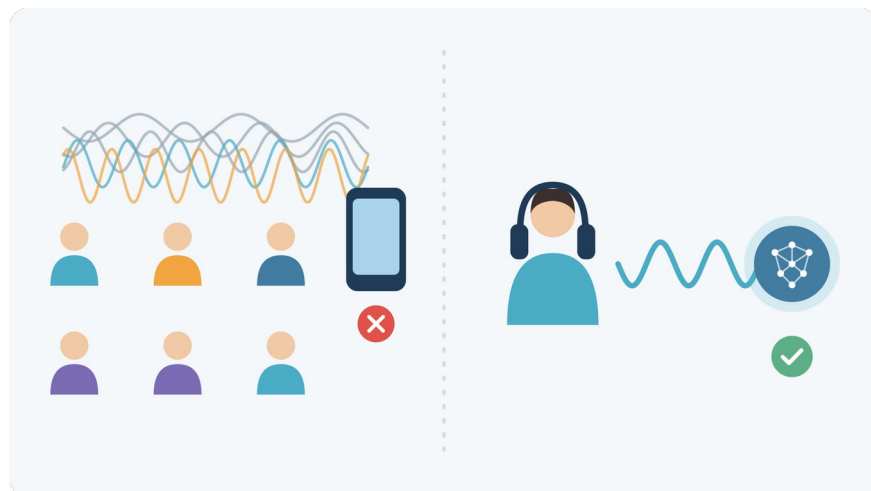
Quick professional assistance

In workshops and practical settings, hands are often busy. A spoken question ("What torque is usually specified for...?") can be faster than typing. However, safety-critical information must always be checked in the manufacturer's documentation or the relevant standard – a confident spoken answer is not more reliable than a written one.

Reflective dialogue

Voice conversation is well suited to reflection: talking through a lesson that did not go as planned, preparing for a difficult conversation with parents, or – for learners – practising a job interview. Asking the AI to take the role of a critical friend who asks questions, rather than one who simply agrees, makes the reflection more useful.

Practical limitations



In a noisy classroom, recognition suffers; voice works best one-to-one in a quiet setting

- **Transcription errors:** accents, dialects, mixed languages and specialised terminology can be misrecognised. Check the transcript when the details matter.
- **Noise:** background noise and several people talking at once can trigger unintended responses. Voice works best one-to-one; in class, use a headset or a quiet corner, and on iPhone the voice isolation setting can help.
- **No visible sources:** a spoken answer does not show its sources as clearly as a written one. For facts, ask for the sources in the chat and check them.
- **Persuasiveness:** a fluent, friendly voice can make wrong answers sound more convincing. Hallucinations do not disappear in voice mode.
- **Limits:** session length and usage limits apply, and some functions (e.g. screen sharing) are available only in certain modes.

Data protection and appropriate use with learners



Voice recordings and transcripts are personal data: know how they are stored and who can access them

- In ChatGPT, audio clips of voice conversations are stored with the transcript in the chat history and, according to the provider, retained for 30 days; they are not used for training unless the user chooses to share them or enables the corresponding setting. Check the current terms before use.
- A person's voice is personal data. Do not record learners or classmates without consent, and follow the GDPR and the school's data protection rules.
- Respect the providers' age requirements (for ChatGPT: at least 13 years, and parental consent under 18) and the school's rules. With younger learners, use teacher-led demonstrations.
- Agree on rules: no personal or sensitive information in voice conversations, and a clear purpose for each use.

Takeaways for teachers

- Distinguish dictation (speech to text) from voice conversation (spoken dialogue).
- Voice is powerful for brainstorming, language practice, accessibility and reflection.
- Set the language, use a quiet setting, and check the transcript when details matter.
- Spoken answers are not more reliable than written ones: verify facts and safety information.
- Treat voice as personal data: consent, age requirements and school rules apply.

References

- OpenAI (2026): Introducing GPT-Live. OpenAI blog, 8 July 2026.
- OpenAI Help Center (2026): ChatGPT Voice; ChatGPT release notes.
- Lengyel, Zs. (2026): Concrete applications for content generation. VITA training presentation.

Links

- ChatGPT Voice – FAQ: help.openai.com/en/articles/8400625-voice-mode-faq
- Introducing GPT-Live: openai.com/index/introducing-gpt-live

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).



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.