Trascrizione vocale, sintesi vocale, sottotitolazione e supporto alla lettura

Autore: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Revisore: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Obiettivo di apprendimento

La conversione della voce in testo e del testo in voce è tra le applicazioni dell'IA più mature e utili nell'istruzione. Questo materiale mostra come gli strumenti di trascrizione vocale, sintesi vocale, sottotitolazione e supporto alla lettura possano rendere i materiali didattici più accessibili e fruibili nella didattica quotidiana. Al termine del materiale, i docenti saranno in grado di scegliere e usare questi strumenti per studenti con diversi bisogni sensoriali, linguistici, di alfabetizzazione o di concentrazione, di creare sottotitoli per i video, versioni audio dei testi e trascrizioni delle spiegazioni, di preparare un supporto alla lettura semplificato e di verificare la qualità dei risultati.

Introduzione: un contenuto, più canali

La maggior parte dei contenuti didattici raggiunge gli studenti attraverso un unico canale: un testo stampato, una spiegazione orale o un video. Ogni canale singolo esclude qualcuno: lo studente che non riesce a sentire la spiegazione, che non riesce a leggere abbastanza in fretta un testo denso o che perde la concentrazione a metà. Gli strumenti vocali e di lettura basati sull'IA permettono di offrire facilmente lo stesso contenuto su più canali contemporaneamente: un testo che si può anche ascoltare, un discorso che si può anche leggere, un video con i sottotitoli. È il modo più semplice per mettere in pratica il principio dei "**molteplici mezzi di rappresentazione**".

UDL sta per Universal Design for Learning (Progettazione universale per l'apprendimento). È un quadro pedagogico basato sull'idea che lo "studente medio" non esiste: la didattica va quindi progettata fin dall'inizio per essere accessibile al maggior numero possibile di persone, anziché essere adattata ai singoli studenti a posteriori.

Dalla voce al testo (speech-to-text)

Il riconoscimento vocale è diventato abbastanza preciso – nella maggior parte delle lingue – da poter essere usato ogni giorno:

- La **dettatura** è integrata in Windows (Win+H), nelle tastiere di macOS, Android e iOS, in Google Docs e Word e negli assistenti IA come ChatGPT. Gli studenti con difficoltà di scrittura o disabilità motorie possono esprimere le proprie conoscenze parlando.
- **Trascrizione** di registrazioni o di parlato dal vivo: Trascrizione istantanea (Live Transcribe) su Android, la funzione Trascrivi di Word e le trascrizioni delle riunioni in Teams o Google Meet. I servizi online convertono un'ora di parlato registrato in testo in pochi minuti e a basso costo.

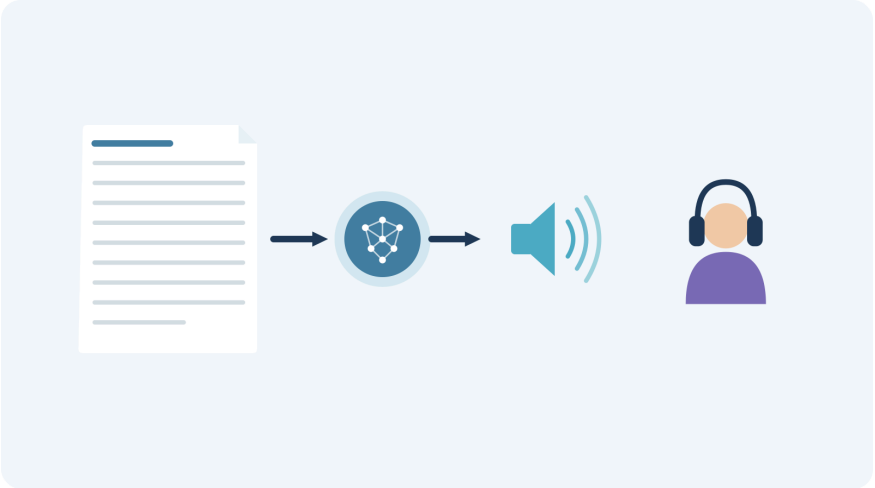
- Le **trascrizioni delle spiegazioni** offrono agli studenti un testo da rileggere, consultare e tradurre: utile per chi ha problemi di udito, per chi non era presente o per chi non studia nella propria lingua madre.



La trascrizione vocale trasforma le spiegazioni orali in testo modificabile

La precisione della trascrizione diminuisce in presenza di rumori di sottofondo, voci sovrapposte, dialetti e terminologia specialistica. Parlate in modo chiaro, utilizzate un microfono esterno quando possibile e correggete i termini chiave prima di condividere la trascrizione. Per registrare le persone è necessario ottenere il loro consenso.

Dal testo alla voce (text-to-speech)



La sintesi vocale trasforma qualsiasi testo in una versione audio dal suono naturale

Le moderne voci neurali hanno un suono naturale e sono disponibili in molte lingue, tra cui l'italiano:

- **Leggi ad alta voce** in Microsoft Edge e lo Strumento di lettura immersiva in Word e OneNote leggono ad alta voce documenti e pagine web ed evidenziano le parole man mano che vengono lette.
- Le **funzioni di accessibilità dello smartphone** (Seleziona per ascoltare su Android, Leggi schermo su iOS) leggono tutto ciò che compare sullo schermo.
- I **servizi di generazione vocale** come ElevenLabs creano narrazioni di alta qualità per le versioni audio dei materiali didattici.

Le versioni audio dei materiali didattici sono utili agli studenti con dislessia o disabilità visive, a coloro che leggono in una seconda lingua e a chiunque preferisca ascoltare, ad esempio durante il tragitto verso la sede di tirocinio.L'ascolto accompagnato dalla lettura («lettura bimodale») può migliorare sia la comprensione sia la capacità di mantenere la concentrazione durante la lettura per periodi prolungati.

Sottotitoli per non udenti e sottotitoli tradotti



I sottotitoli rendono accessibili le spiegazioni video, e sono utili in un'officina rumorosa

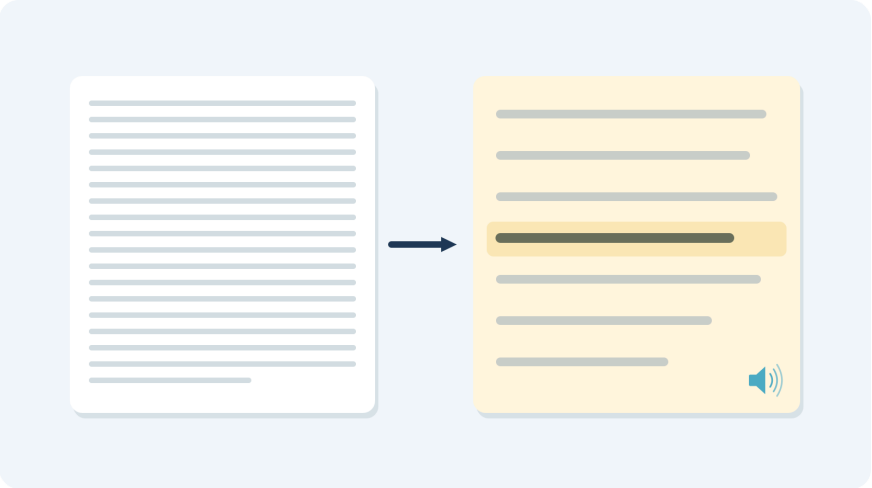
I sottotitoli sono indispensabili per gli studenti sordi e con ipoacusia e le **Web Content Accessibility Guidelines** (WCAG) li richiedono per i video preregistrati. I sottotitoli sono utili anche agli studenti che guardano video in ambienti rumorosi, che apprendono in una lingua diversa dalla propria lingua madre o che hanno bisogno di vedere come si scrive un termine tecnico.
YouTube, Microsoft Stream e la maggior parte degli strumenti video generano **sottotitoli automatici**; i sistemi operativi e i browser offrono **sottotitoli in tempo reale** per qualsiasi audio.

- I sottotitoli automatici vanno **controllati e corretti**, soprattutto per termini tecnici, numeri e avvertenze di sicurezza.

- I sottotitoli possono essere tradotti automaticamente, favorendo così l'apprendimento degli studenti la cui lingua madre è diversa da quella utilizzata durante le lezioni.

- - Privilegiare i sottotitoli attivabili e disattivabili (closed captions), che gli studenti possono visualizzare o nascondere e di cui possono modificare le dimensioni, rispetto ai sottotitoli incorporati direttamente nel video.

Supporto alla lettura



Le modalità di lettura riducono il disordine visivo, evidenziano la riga corrente e possono leggere il testo ad alta voce

Gli strumenti di supporto alla lettura riducono lo sforzo di decodifica e aiutano gli studenti a concentrarsi sul significato:

- Le **modalità di lettura** (Strumento di lettura immersiva, visualizzazione di lettura del browser) eliminano gli elementi di disturbo e permettono di modificare dimensione del carattere, spaziatura, colore di sfondo e messa a fuoco della riga; alcuni strumneti possono dividere le parole in sillabe o mostrare immagini per le parole di uso comune.
- **Semplificazione e riassunti con l'IA:** un modello linguistico può riscrivere un testo a un livello di lettura più semplice, dividerlo in brevi sezioni, aggiungere titoli e un riassunto, mantenendo i termini tecnici e spiegandoli.
- I **glossari** dei termini chiave di un'unità, con spiegazioni semplici e immagini, aiutano tutti gli studenti, in particolare chi legge in una seconda lingua.

Ricetta: una scheda di lavoro, più canali (circa 20 minuti)
1. Controllate la struttura: titoli chiari, paragrafi brevi, elenchi invece di frasi lunghe.
2. Create una versione audio con Leggi ad alta voce o con uno strumento di sintesi vocale.
3. Chiedete all'IA una versione semplificata e un glossario di 8–10 termini chiave, poi verificateli entrambi.
4. Aggiungete i sottotitoli al video dimostrativo e correggete i termini tecnici.
5. Offrite tutte le versioni all'intero gruppo, in modo che gli studenti possano scegliere.

Limiti e buone pratiche

- Trascrizioni, sottotitoli e semplificazioni generati dall'IA contengono errori: il docente resta responsabile dell'accuratezza di ciò che condivide.
- Le registrazioni vocali e i documenti possono essere elaborati nel cloud: vanno rispettate le norme sulla protezione dei dati e non

caricare dati personali. Per i materiali sensibili, gli strumenti che funzionano in locale o gli account istituzionali sono scelte più sicure.
- Gli strumenti di supporto non sostituiscono l'insegnamento della lettura e della scrittura, ma facilitano l'accesso ai contenuti mentre gli studenti sviluppano queste competenze.
- Offrire possibilità di scelta invece di imporre: gli studenti sanno meglio di chiunque quale formato li aiuta.

Indicazioni per i docenti

- Offrire lo stesso contenuto su più canali: testo, audio e video sottotitolato.
- Dettatura e trascrizione aiutano gli studenti che trovano difficile scrivere o ascoltare.
- Le voci simili a quelle umane permettono di produrre rapidamente versioni audio di qualsiasi testo.
- Controllare sempre sottotitoli e trascrizioni automatici, soprattutto i termini tecnici e le informazioni sulla sicurezza.
- Modalità di lettura, versioni semplificate e glossari riducono lo sforzo della lettura e favoriscono la comprensione.

Bibliografia

- 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. Presentazione della formazione VITA.

Link

- Lingue e voci dello Strumento di lettura immersiva: support.microsoft.com
- Trascrizione istantanea e Notifiche sonore: play.google.com
- Sottotitoli – W3C: w3.org/WAI/media/av/captions
- Linee guida UDL: udlguidelines.cast.org
- W3C Web Accessibility Initiative: w3.org/WAI

Copyright © 2026 Consorzio VITA

Quest'opera è concessa in licenza secondo i termini della Creative Commons Attribuzione-Condividi Allo Stesso Modo 4.0 Internazionale (CC BY-SA 4.0).

Le immagini senza indicazione della fonte sono state generate con Claude per Mac.

Finanziato dall'Unione europea. Le opinioni espresse sono esclusivamente quelle degli autori e non riflettono necessariamente quelle dell'Unione europea o dell'Agenzia nazionale.