

**Le voci sintetiche
promettono di migliorare i
risultati dell'apprendimento**





Nel panorama tecnico odierno l'intelligenza artificiale, gli esseri umani virtuali e la tecnologia vocale stanno assumendo un ruolo sempre più importante nella tecnologia didattica. Storicamente le voci sintetiche, ovvero quelle generate al computer, sono state considerate inferiori alle voci umane per quanto riguarda i risultati di apprendimento. Tuttavia studi recenti hanno dimostrato che, grazie ai continui progressi della tecnologia vocale e all'abbinamento di un essere umano virtuale, le voci sintetiche moderne possono effettivamente produrre risultati di apprendimento migliori rispetto alle voci umane o ai vecchi motori di sintesi vocale.

Secondo lo studio *'Riesame dell'effetto vocale quando si apprende da un essere umano virtuale'* condotto da Scotty D. Craig dell'Arizona State University e da Noah L. Schroeder della Wright State University, **"il motore vocale moderno ha generato un apprendimento significativamente maggiore sui risultati di trasferimento, ha dimostrato una maggiore efficienza di addestramento ed è stato valutato allo stesso livello di un agente con voce umana per quanto riguarda il facilitare l'apprendimento e la valutazione di credibilità superando al contempo in efficienza i motori vocali di vecchia generazione. Questi risultati mettono in discussione sia i risultati precedenti in cui si usavano motori vocali di vecchia concezione che le affermazioni sull'esistenza del cosiddetto effetto vocale."** ⁽¹⁾

Man mano che le innovazioni tecnologiche entrano in classe si sviluppa un'attività di ricerca in continua crescita riguardante un'efficace progettazione e implementazione delle tecnologie educative. In linea di massima, si è scoperto che le tecnologie di apprendimento diventano più efficaci quando vengono proposte tramite esseri umani virtuali o personaggi a schermo di aspetto simile all'umano. ⁽²⁾

Gli esseri umani virtuali vengono utilizzati in ambienti di apprendimento multimediale e sistemi di tutoraggio intelligenti come istruttori, i personaggi sono presenti nei videogiochi didattici o come agenti pedagogici. In particolare, i personaggi aiutano nel processo di apprendimento dando dei segnali, motivando, assumendo ruoli di role-playing, come facilitatori o modellatori di strategie di apprendimento.

Le voci sintetiche promettono di migliorare i risultati dell'apprendimento

“The modern voice engine produced significantly more learning on transfer outcomes, had greater training efficiency, and was rated at the same level as an agent with a human voice for facilitating learning and credibility while outperforming the older speech engine.”

Come si può ragionevolmente intuire, i ricercatori hanno dimostrato che la specifica progettazione dell'agente pedagogico ossia la sua voce, i suoi schemi linguistici o il suo aspetto, influenza l'efficacia con cui l'agente facilita l'apprendimento ⁽³⁾. Questi risultati evidenziano l'importanza di una «progettazione degli agenti pedagogici mirata e basata sui dati». ⁽⁴⁾

Secondo ciò che è noto tradizionalmente come *effetto vocale o principio vocale*, «in una situazione di apprendimento multimediale l'apprendimento tende a migliorare quando la narrazione viene fornita da una voce umana con intonazione standard (standard-accented) mentre non migliorerà se la voce è stata generata tramite computer, ossia con una cosiddetta «voce robotica». ⁽⁵⁾

Mayer trovò all'epoca delle prove convincenti a sostegno di questa conclusione. Tuttavia, quattro di questi studi sono stati pubblicati almeno 10 anni fa. Da allora la tecnologia vocale ha fatto rapidi progressi e il software di sintesi vocale è notevolmente migliorato.

La ricerca di Craig e Schroder del 2017, *'Riesame dell'effetto vocale quando si apprende da un essere umano virtuale'*, analizza le implicazioni dell'effetto vocale associato ad un essere umano virtuale sui risultati dell'apprendimento, sul carico cognitivo e su come l'agente viene percepito dal discente.

Storicamente, i ricercatori ritennero che l'apprendimento tramite una voce artificiale determinasse un ulteriore carico cognitivo su chi studia, oltre a causargli delle concrete distrazioni. Le prime ricerche di Mayer e altri nel 2003, e quelle successive nel 2005, hanno dimostrato che le voci umane erano nettamente superiori a quelle sintetiche. Tuttavia, i risultati di uno studio simile condotto nel 2012 da Mayer e DaPra utilizzando una tecnologia vocale più avanzata non hanno evidenziato alcuna differenza nell'apprendimento tra gruppi che disponevano di agenti con voci umane e gruppi che lavoravano con voci sintetiche. Questo stava ad indicare il fatto che, forse, l'effetto vocale non sussisteva più.

Nella loro ricerca del 2017 Craig & Schroder hanno utilizzato come base il motore vocale di Microsoft in quanto motore vocale classico e **la voce «Kate» di NeoSpeech (ora sotto il marchio ReadSpeaker)** come rappresentante di un motore vocale moderno. Come controllo di fascia alta è stata utilizzata una voce umana. Tutte e tre le voci sono state allocate a un essere umano virtuale femminile.

Una selezione casuale di partecipanti è stata valutata su percezione, carico cognitivo, domande a scelta multipla e ritenzione. «Per la prima e la seconda delle domande in cui si articolava la ricerca (apprendimento e carico cognitivo) sono stati trovati risultati coerenti che o non hanno evidenziato differenze tra le diverse condizioni oggetto della ricerca o hanno dimostrato che la presentazione dell'agente con un motore vocale moderno era più efficace rispetto alla presentazione con motore vocale più vecchio o alla versione con voce umana. Ciò fornisce una prova coerente contro l'esistenza dell'effetto vocale.» ⁽⁶⁾ Per contro, non sono state osservate differenze statisticamente significative per quanto riguarda i valori misurati nelle domande a scelta multipla, nei test di

Le voci sintetiche promettono di migliorare i risultati dell'apprendimento

ritenzione e nelle altre misurazioni di efficienza.

Confrontando i risultati del text-to-speech moderno con quelli ottenuti con le voci umane registrate si può dunque concludere che il tipo di voce usato non è più così importante per i risultati di apprendimento come una volta si riteneva e che i moderni motori vocali possono essere altrettanto efficaci di una voce umana registrata. Analogamente, non si sono osservate differenze nelle valutazioni date dai partecipanti al test sulla capacità dell'agente di facilitare l'apprendimento e la credibilità percepita.

In definitiva, lo studio di Craig & Schroder, che ha utilizzato una tecnologia vocale più avanzata, non solo smentisce il mito secondo cui le voci umane sarebbero superiori negli ambienti di apprendimento ma sottolinea anche il fatto che le moderne voci sintetiche possono persino produrre risultati migliori delle voci umane.

È dunque possibile che l'idea di lunga data secondo cui gli esseri umani virtuali saranno in grado di migliorare l'apprendimento è al momento assolutamente fondata e continuerà ad esserlo in misura ancora maggiore in futuro. ⁽⁷⁾

(1) - Craig & Schroeder, 2017

(2) - Dehn & Van Mulken, 2000; Graesser, McNamara, & VanLehn, 2005; Graesser & McNamara, 2010; Johnson & Lester, 2016

(3) - Baylor & Kim, 2004, 2009; Clark & Choi, 2005; Domagk, 2010; Kim & Wei, 2011; Moreno & Flowerday, 2006; Ozogul, Johnson, Atkinson, & Reisslein, 2013; Schroeder, Romine, & Craig, 2017; Veletsianos, 2010

(4) - Craig & Schroeder, 2017

(5) - Mayer, 2014b, p. 358

(6) - Craig & Schroeder, 2017

(7) - Johnson & Lester, 2016

ReadSpeaker

pioneering voice technology

ReadSpeaker è un fornitore globale di voci text-to-speech (TTS) di alta qualità in grado di supportare decine di lingue e offrire una pluralità di voci sintetiche dal timbro vocale pressoché identico a quello umano. Utilizzando tecnologie proprietarie leader del settore, l'azienda offre le voci sintetizzate dal suono più naturale disponibili sul mercato e si avvale della tecnologia Deep Neural Network (DNN) di nuova generazione per migliorare strutturalmente la qualità delle voci a tutti i livelli. ReadSpeaker è una società del gruppo HOYA Corporation, con uffici in 15 paesi, oltre 10.000 clienti in 65 nazioni, in grado di proporre un'offerta completa di soluzioni text-to-speech sia come Software-as-a-Service (SaaS) che come licenze software utilizzabili per applicazioni offline. Unico fornitore di sintesi vocale che controlli interamente la catena di produzione e distribuzione delle voci TTS, ReadSpeaker integra tutte le tecnologie vocali innovative di HOYA (NeoSpeech, Voiceware, VoiceText e rSpeak) offrendo soluzioni TTS a 360 gradi tramite una vasta gamma di applicazioni vocali per tutte le piattaforme e i dispositivi utilizzati nei vari settori industriali, per organizzazioni pubbliche e private, per esigenze di vocalizzazione online, offline, embedded, su server o desktop, per app vocali, produzione di audio parlato, voci customizzate e altro ancora. Con oltre 20 anni di esperienza, ReadSpeaker è sinonimo di «tecnologia vocale pionieristica» all'avanguardia nel settore del text-to-speech.

 contact@readspeaker.com

 www.readspeaker.com

 [company/readspeaker](https://www.linkedin.com/company/readspeaker)

 [readspeaker](https://www.facebook.com/readspeaker)

 [@readspeaker](https://twitter.com/readspeaker)

To learn more about ReadSpeaker, contact us at www.readspeaker.com/contact-us