

# Architettura di una voce

Nell'era digitale, dove l'interazione tra uomo e macchina diventa sempre più sofisticata ed allarmante, la capacità di simulare una voce umana reale assume un'importanza dai terribili risvolti. La domanda "Quanti sono e quali sono i parametri per simulare una voce reale umana?" apre la porta a un mondo in cui la tecnologia si propone di sondare i confini tra il reale e l'artificiale. Vediamo quali sono i parametri fondamentali che consentono agli sviluppatori di creare voci sintetiche che possono ingannare l'orecchio umano, toccando temi come pitch, intonazione, durata, e molto altro ancora.

Il sempre più vertiginoso progresso della tecnologia e dell'intelligenza artificiale hanno reso possibile l'imitazione non solo delle caratteristiche vocali, ma anche delle sfumature emotive e culturali che, fino ad ora, hanno reso unica ogni voce umana.

La simulazione di una voce umana attraverso la sintesi vocale richiede, dunque, diversi parametri per essere efficace e realistica. I parametri principali possono variare a seconda della tecnologia specifica o dell'algoritmo usato.

•

**Pitch (Altezza):** Si riferisce alla frequenza fondamentale della voce. Il pitch può variare per imitare diverse caratteristiche vocali legate al genere, all'età o alle emozioni.

**Intonazione:** Riguarda il pattern di variazione del pitch lungo una frase. Aiuta a comunicare interrogativi, affermazioni, emozioni, e può caratterizzare l'accento di un parlante.

*La voce è la nostra firma*

**Durata:** Si riferisce alla lunghezza temporale dei suoni e delle pause. Influisce sulla percezione del ritmo del discorso e sulla distinzione tra le parole e le frasi.

**Intensità (Volume):** Il livello di pressione sonora del discorso, che può variare per esprimere emozioni o enfasi.

**Timbro:** Caratteristica che permette di distinguere tra voci anche se parlano alla stessa altezza e volume. Dipende dalla forma del tratto vocale e dalla modalità di vibrazione delle corde vocali.

**Qualità Vocale:** Includono caratteristiche come la voce nasale, soffiata, raucedine, ecc., che contribuiscono alla singolarità di una voce.

**Articolazione:** La precisione con cui i suoni del discorso sono prodotti. Implica il modo in cui le labbra, la lingua, e altre parti del tratto vocale sono utilizzate per formare suoni distinti.

**Fluenza:** Riguarda la fluidità con cui il discorso è prodotto. Interruzioni, esitazioni, ripetizioni possono essere simulate per rendere il discorso più naturale o per imitare certi stili di discorso.

**Prosodia:** Comprende l'insieme di caratteristiche come pitch, durata, e intensità applicate su più ampie unità di discorso come frasi o paragrafi, contribuendo a modellare l'espressione e l'interpretazione del discorso.

**Contesto Linguistico e Culturale:** La conoscenza della lingua e delle sue sfumature, degli accenti regionali, del gergo e degli usi specifici, che possono influenzare la pronuncia e l'intonazione.

L'intelligenza Artificiale con la simulazione della voce umana sta aprendo un confine sconosciuto: dalla ricreazione della voce di persone che non ci sono più, alla traduzione della voce di divulgatori stranieri per offrirli in scala globale.

Tecnologie avanzate come quelle legate al deep learning stanno migliorando la capacità di simulare voci umane in modo sempre più accurato, permettendo di personalizzare i parametri sopra menzionati per creare voci sintetiche difficilmente distinguibili da quelle reali.

E questo include l'analisi e la sintesi di emozioni, dialetti, e variazioni individuali nelle caratteristiche vocali.

Attento, allora, non solo a quello che dici, ma anche a come lo dici...

*Fonte: Edmondo Jonghi Lavarini*