

Sintesi della voce (I)

La sintesi della voce è una delle principali tecnologie del linguaggio, che assieme al riconoscimento della voce, costituiscono il nucleo della comunicazione uomo-macchina.

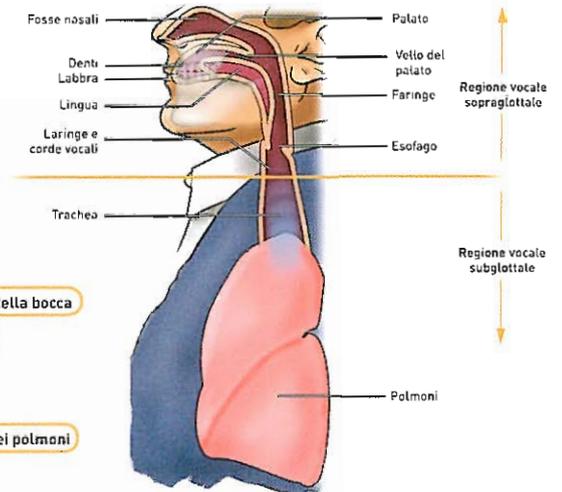
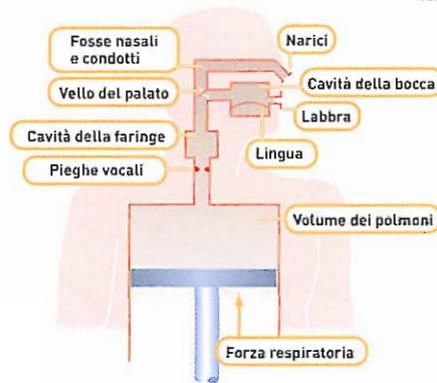
Il processo di sintesi consiste nel creare una replica di generazione automatica del segnale della voce, dotando il computer della capacità di produrre messaggi vocali non registrati precedentemente.

All'interno dello studio di questa materia possiamo riconoscere due correnti di pensiero.

La prima presuppone il processo di generare il suono a partire da diversi fattori base, che sono dei caratteri scritti (il testo) e una serie di parametri estratti dall'analisi della parola naturale e del risultato della decodificazione di onde precedentemente memorizzate.

Questo servirà per creare la voce artificiale: date le parole scritte, il sintetizzatore si incarica di pronunciarle.

La seconda corrente di pensiero concepisce la sintesi come un modello dell'attuazione umana. Spiegherebbe come utilizziamo le risorse computazionali di cui disponiamo per aumentare la conoscenza del processo di produzione della parola nella nostra specie.



Gli organi della parola e il modello funzionale della regione vocale.

Breve storia

Sono più di duecento anni che l'uomo si interessa di questo argomento. Nel 1791 Kempelen pubblicò il meccanismo di una macchina per parlare.

Dudley (Bell) - Voder implementò nel 1939 una sintesi spettrale della parola e nel 1952 Cooper (Bell) fece alcune variazioni su di essa.

Fu però negli anni Sessanta, con l'arrivo dei computer digitali, che si poté simulare un sintetizzatore (il più importante è quello di Klatt). Poco dopo la microelettronica rese possibile costruirli a basso costo e, negli anni Novanta furono creati i primi sintetizzatori di alta qualità, ai quali veniva richiesto un linguaggio non robotizzato.

Limitazioni tecnologiche

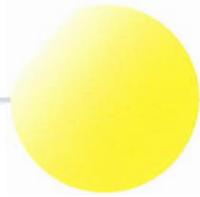
Le difficoltà con cui si scontrarono i ricercatori nella sintesi della voce erano la

dimensione della memoria e la velocità del computer.

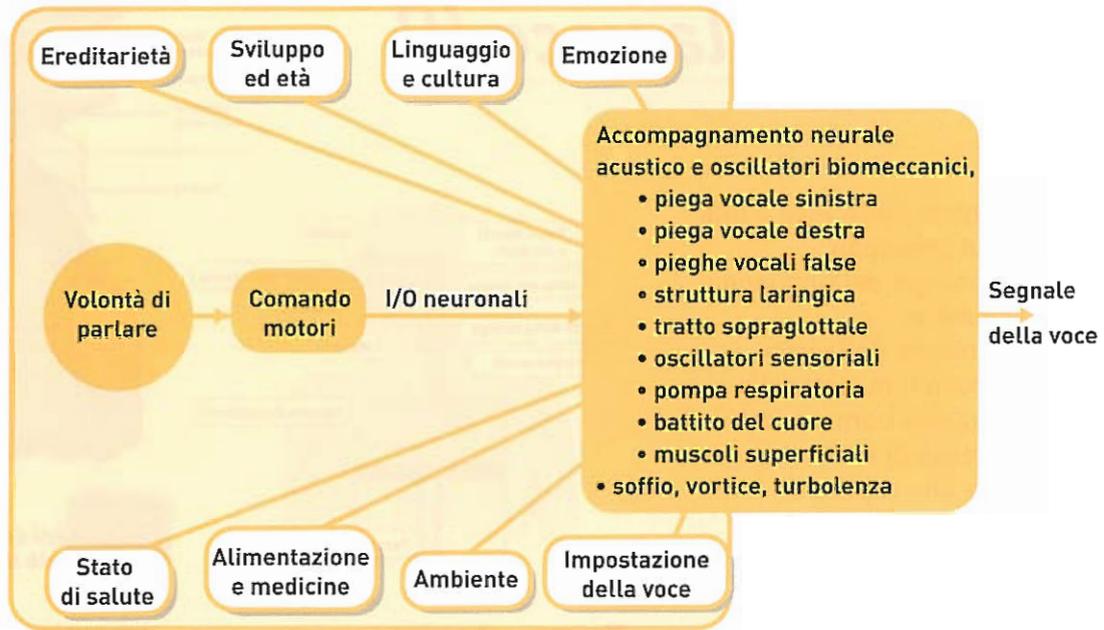
Il problema maggiore consisteva nell'ottenere una voce naturale (rendere la parola "più umana"). Quindi l'utilizzo di queste macchine restò riservato a sistemi costosi e sofisticati, tuttavia, più tardi, grazie al progressivo calo di prezzi di computer e schede audio, così come allo sviluppo della tecnologia multimediale, gli studi su questa tecnica ebbero nuovi impulsi.

Difficoltà nella ricerca e nello sviluppo

La chiave sta nel creare una macchina che imiti il processo naturale di sintesi degli umani: la generazione di suoni vocali che si sviluppano nel tempo insieme alla pressione dei polmoni, alla tensione della glottide e alla configurazione vocale e nasale, controllati dal nostro cervello per estrarre l'essenza delle parole.



Fattori che influenzano la generazione del segnale della voce. Si possono dividere in fattori propri della persona, come le caratteristiche della voce, ecc. e fattori esterni, dell'ambiente che la influenzano direttamente.



Concetti fondamentali di suono e voce

Per poter sintetizzare qualche cosa bisogna sapere prima come funziona, e questo richiede studi sul suono, sulla voce e sul campionamento dei segnali. Per tanto saremo obbligati a studiare il processo di creazione del messaggio vocale dal punto di vista acustico e linguistico.

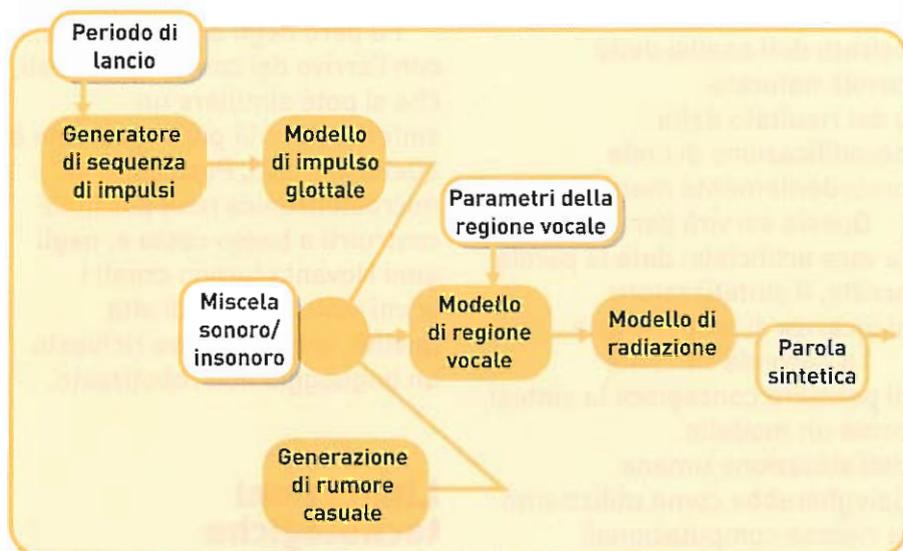
Questo ci porta a contemplare le caratteristiche della voce che comprende lo studio dei diversi suoni, dei componenti più significativi della parola, delle relazioni strutturali di queste ultime e del loro significato, tenendo presente le differenze fondamentali fra il supporto vocale e il supporto scritto. Come principio fondamentale, assumiamo che la parola umana si forma mediante la generazione di suoni nel sistema vocale, insieme alle proprietà dell'udito e della percezione.

Dobbiamo comprendere il

comportamento fisico dell'apparato fonetico umano e come elabora i suoni il sistema uditivo per sviluppare un modello matematico del medesimo. Considerando la sua struttura linguistica, si estrae dal testo l'informazione necessaria per controllare il modello

matematico e in questo modo, a partire dai caratteri scritti, generare la voce.

Riassumendo, la sintesi consiste nel produrre nuove sequenze orali e non nel riprodurle, questo costituisce l'interfaccia orale della comunicazione fra una macchina e il suo utente.



Modello di produzione di parola sintetica. Nella generazione della voce a partire dal testo bisogna tenere conto di molti fattori diversi. In questo grafico sono riportate le fasi necessarie per completare il processo.