Sintesi della voce (II)

I principali parametri in un sistema di sintesi vocale, implicano la naturalezza, la qualità e l'intelligibilità del linguaggio generato, la versatilità del sistema (vocabolario illimitato di fronte alla generazione dei testi senza restrizioni) e la complessità dell'elaborazione. La naturalezza del linguaggio si ottiene con una buona intonazione, cosa che in alcuni casi è necessaria per il buon intendimento del messaggio, tanto che la si considera come una delle principali responsabili della qualità. A sua volta, un sistema ideale di sintesi, dovrebbe offrire una qualità elevata negli enunciati prodotti, ed essere molto versatile, in altre parole capace di produrre qualsiasi messaggio, e per quanto riguarda la parte dell'elaborazione dovrebbe essere relativamente semplice.

Un aspetto altrettanto importante nell'intelligibilità e nella naturalezza del segnale sintetizzato, sono le regole prosodiche, visto che pur se in certa misura possono interferire nella struttura sintattica della frase, è preferibile che la macchina "sappia" ciò che sta dicendo, per generare un'intonazione adeguata; Tuttavia trasmettere emozione nella parlata sintetica è molto difficile. Nella ricerca di questa perfezione, la prosodica è un elemento più umano e complicato, e suppone una valutazione dei condizionamenti linguistici (intrinseci del linguaggio), emozionali e para-linguistici, come la qualità della voce. Il problema di generare onde sonore portatrici di messaggi, quindi, è complesso e siamo ancora lontani da un programma di sintesi che riproduca fedelmente la variazione fonetica presente nella parlata, che offra il livello di qualità e di flessibilità richiesto da molte applicazioni, pur mantenendo un'elaborazione non troppo complessa.

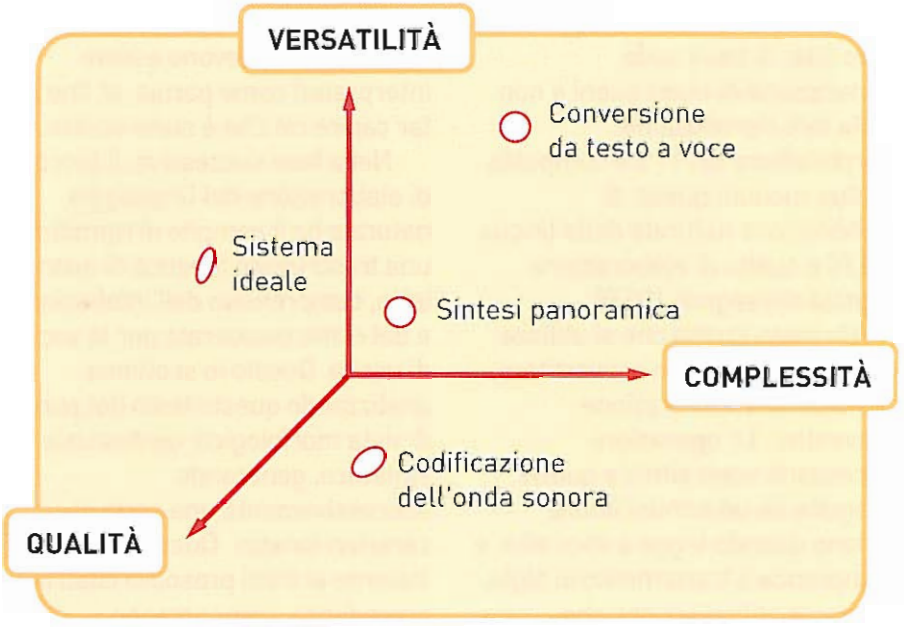
Le tecniche di sintesi

Possiamo dividerle in quattro grandi gruppi:

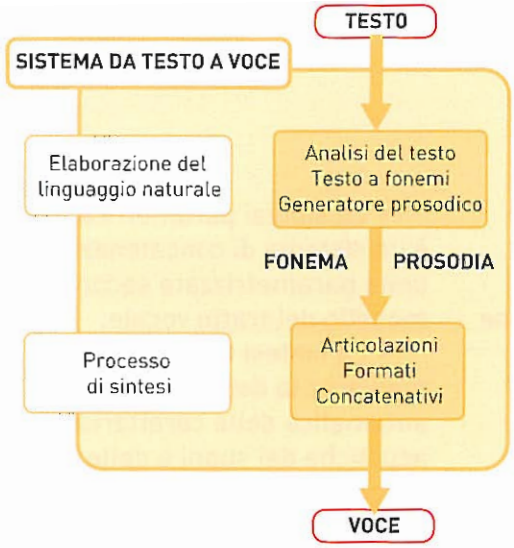
- Le tecniche di codificazione del linguaggio come la memorizzazione e la riproduzione dei messaggi mediante l'elaborazione digitale del segnale.
- La sintesi parametrica, che è un sistema di concatenazioni di unità parametrizzate secondo un modello del tratto vocale.
- La sintesi tramite regole, mediante la determinazione automatica delle caratteristiche acustiche dei suoni e delle regole di concatenazione.
- La conversione di testo in voce o trasformazione dei caratteri scritti nella loro esposizione orale.

Sistemi di sintesi della voce

Un sintetizzatore TTS (da testo a suono) è un sistema computerizzato che deve essere capace di leggere qualsiasi testo a



Tecniche di sintesi in funzione della qualità, complessità e versatilità del sistema.



Semplice diagramma funzionale di un sistema TTS.

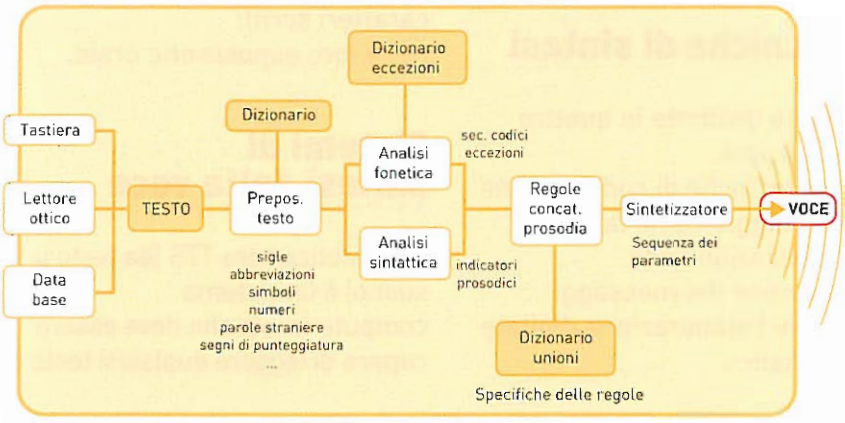
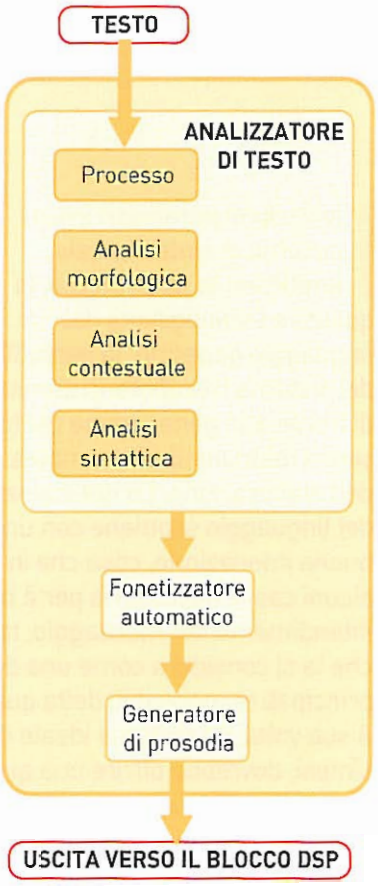


Diagramma a blocchi dettagliati di un convertitore da testo a voce.



Modulo NLP di un sistema TTS generale.

voce alta. Si basa sulla generazione di nuovi suoni e non sulla loro riproduzione. L'architettura del TTS è composta da due moduli: quello di elaborazione naturale della lingua (NLP) e quello di elaborazione digitale dei segnali (DSP).

Un testo scritto che si utilizza come ingresso di un convertitore richiede una elaborazione preventiva. Le operazioni necessarie sono simili a quelle eseguite da un annunciatore umano quando legge a voce alta, e comprende il trattamento di sigle, numeri e abbreviazioni, che devono apparire nella loro forma completa. Queste caratteristiche di ingresso consistono in un insieme di simboli che devono essere interpretati come parole, al fine di far capire ciò che è stato scritto.

Nella fase successiva, il blocco di elaborazione del linguaggio naturale ha il compito di riprodurre una trascrizione fonetica di quanto letto, comprensivo dell'intonazione e del ritmo desiderato per la voce di uscita. Questo lo si ottiene analizzando questo testo dal punto di vista morfologico, contestuale e sintattico, generando successivamente una serie di caratteri fonetici. Questi caratteri insieme ai tratti prosodici citati in precedenza, comporranno l'ingresso del modulo di elaborazione digitale del segnale (DSP) in cui si trasformerà l'informazione simbolica ricevuta dal NLP in una voce di uscita.

Le informazioni che entrano nel DSP contengono anche un insieme di simboli che sono interpretati come parole, e altri elementi linguistici differenti, come nel caso della punteggiatura che può influire nella pronuncia. Come si può facilmente intuire, le operazioni coinvolte nell'elaborazione digitale sono quelle che realizza il computer, che corrispondono al controllo dinamico dei muscoli articolatori e della frequenza di vibrazione delle corde vocali, in modo che il segnale di uscita sia adattato alle caratteristiche di quello di ingresso.