

Introduzione al suono delle macchine



Nella vita quotidiana tutti siamo abituati a interagire con le macchine, in molti modi diversi. Alcune di queste macchine, ad esempio i televisori e i videoregistratori, ricevono gli ordini dall'utente mediante i pulsanti di un comando a distanza; altri come le lavatrici, i forni a microonde, ecc. tramite un pannello di controllo con pulsanti e manopole, i computer principalmente con la tastiera e il mouse. Però queste non sono le uniche vie di intercomunicazione fra la macchina e la persona.

La comunicazione tramite il suono è un fattore che si sta studiando e che possiamo trovare già applicato: sulle strade esistono semafori che avvisano i non vedenti quando è possibile attraversare la strada senza pericolo, centraline telefoniche che rispondono e rinviano automaticamente le chiamate,

telefoni cellulari che chiamano un numero sentendo il nome del destinatario della chiamata, ecc.

Le ricerche su questa tecnologia hanno reso possibile ai computer di poter ricevere comandi per mezzo della voce umana e anche rispondere con una voce che sembra umana. Queste interfacce basate sulla voce coinvolgono principalmente due tecnologie: riconoscimento e sintesi della voce.

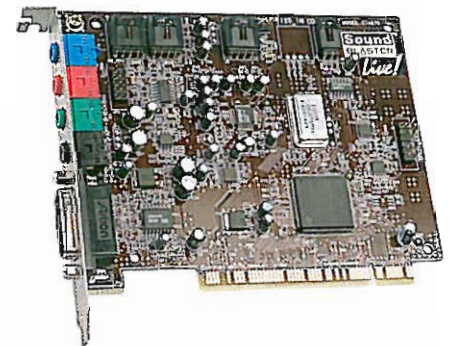
Il riconoscimento della voce è il processo di trasformazione di un segnale sonoro in testo.

La sintesi della voce o Tts (Text to speech) è il processo di trasformazione del testo in una sequenza di suoni. Grazie al veloce sviluppo dei computer, i diversi campi di elaborazione della voce renderanno possibile in un prossimo futuro una vera e totale comunicazione fra l'essere umano e il computer.

Componenti del sistema

Per poter utilizzare il suono come mezzo di comunicazione con l'utente, il sistema deve poter disporre di un microfono, un altoparlante, una scheda audio e una applicazione software.

Ognuno di questi componenti svolge il suo compito all'interno del processo, dalla cattura della



Scheda audio.



Cuffia da telefonista.

voce da parte del microfono sino all'emissione del suono generato dagli altoparlanti, tuttavia chi determina in maggior misura il rendimento e l'affidabilità del sistema è senza dubbio il programma che si sta utilizzando.

Esistono molte similitudini nei concetti del trattamento del suono e del trattamento delle immagini, anche se normalmente il volume dei dati da processare è significativamente minore quando parliamo di sintesi o riconoscimento della voce.

Come si ottiene?

Ottenere la sintesi della voce è relativamente facile, il modo più semplice è memorizzare tutti i possibili messaggi completi, dopo di che si riproducono al momento del bisogno. Questa è la tecnica utilizzata dai distributori di sigarette, dai mezzi di trasporto come il treno o la metropolitana per annunciare le stazioni, o i distributori di benzina.

Normalmente un programma di sintesi vocale contiene un data base con i diversi fonemi della lingua, e li riproduce seguendo il

testo in modo da formare le parole che compongono la frase.

Il riconoscimento della voce è un lavoro più complesso.

Esistono diversi metodi, quali la comparazione di campioni di voce, l'utilizzo di modelli automatici parametrizzati, o le tecniche di estrazione delle caratteristiche fonetiche. Tutti questi metodi si basano sull'acquisizione del segnale vocale e sulla ricerca, all'interno di un data base, del campione che più si avvicina a quello acquisito.

Applicazioni di questa tecnologia

Le applicazioni più comuni nell'ambito domestico sono l'esecuzione di comandi mediante la voce e la dettatura automatica. Queste operazioni richiedono un certo addestramento del programma da parte dell'utente, per cui all'inizio il rendimento non è molto elevato. Nell'ambito professionale, soprattutto nei servizi di assistenza ai clienti, sono molto comuni le segreterie automatiche che interagiscono con l'utente mediante un'approssimazione del dialogo fra due persone. Negli ultimi anni si sta insistendo nella ricerca sulla sintesi e sul riconoscimento della voce, grazie alle molteplici possibilità che offre, soprattutto per aiutare le persone con handicap fisici.

Una di queste applicazioni già in uso è quella che permette l'accesso a Internet alle persone non vedenti, che accedono alle informazioni e

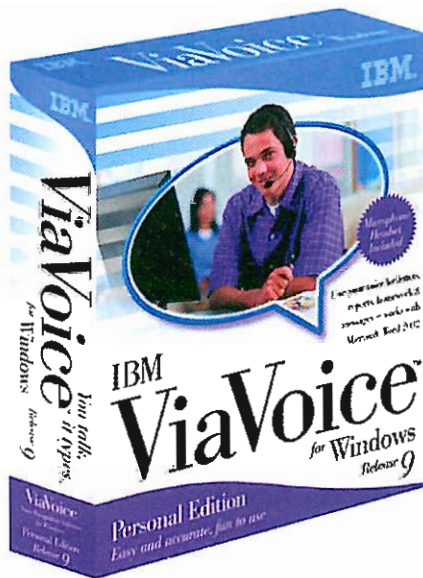
navigano mediante la loro voce, o nella fabbricazione di apparati che permettono di comunicare meglio alle persone laringectomizzate.

Programmi di sintesi e riconoscimento

Esempi di programmi di sintesi e di riconoscimento vocale, sono il Via Voice di IBM, molto noto per l'alto rendimento che offre, sia nelle versioni per privati sia per uso professionale, o i prodotti della serie Naturally Speaking di Dragon Systems.

Questi programmi sono disponibili per tutte le versioni dei sistemi operativi di Microsoft e anche per il Mac.

Per gli altri sistemi operativi, tipo Linux, i software non sono ancora così numerosi, anche se



esistono applicazioni meno conosciute come gli strumenti SFS (Speech Filling System), le OGI Speech Tools o il Festival Speech Sintesis.