

# Riconoscimento della voce (I)

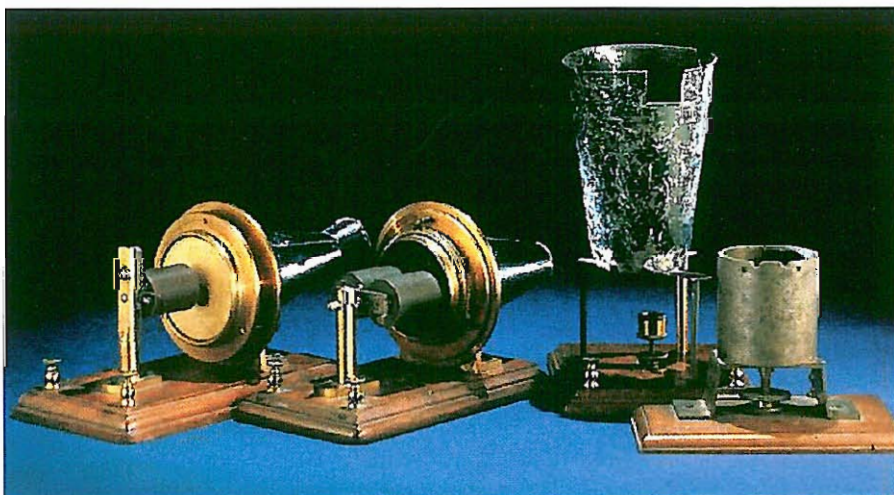
La funzione principale del riconoscimento della voce è ottenere la sequenza di parole associata a ciò che una persona pronuncia nel suo linguaggio naturale. Così, per sistema di riconoscimento automatico della voce si intende un insieme di dispositivi sia hardware che software capaci di decodificare i suoni che sono parte di un determinato segnale vocale.

Questa decodificazione può essere realizzata utilizzando diverse tecniche, e dipende da due fattori fondamentali: il numero di interlocutori, può essere sempre la stessa persona che interagisce con il sistema (dipendente), o diverse persone (indipendente), e il flusso delle parole, che è discreto quando diciamo parole separate da momenti di silenzio, ed è continuo quando parliamo come se si trattasse di una conversazione normale. In fondo, si tratta di

generare e memorizzare un insieme di matrici che possono essere comparate con il segnale acustico di ingresso che dobbiamo riconoscere, restituendo la sequenza di caratteri che con maggior probabilità lo "rappresentano". I data base dove vengono memorizzati i campioni sono denominati dizionari di modelli acustici.

## Storia del riconoscimento della voce

Possiamo dire che la storia del riconoscimento della voce iniziò nel 1870, quando Alexander Graham Bell volle sviluppare un dispositivo capace di fornire parole scritte a persone sorde. Non ebbe successo, anche se lo sforzo di questa ricerca condusse allo sviluppo del telefono.

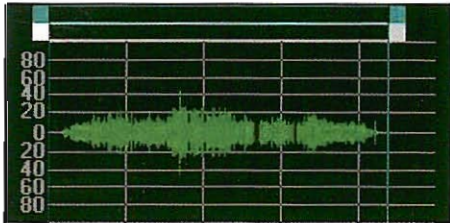
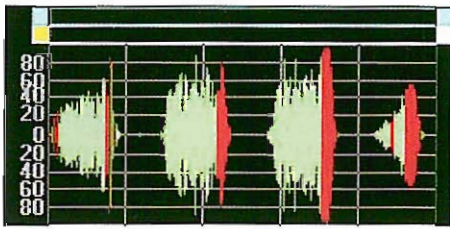
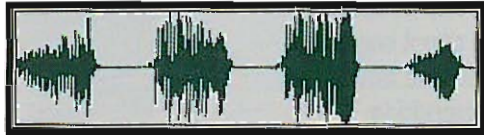
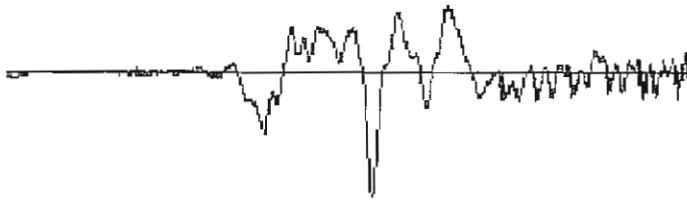
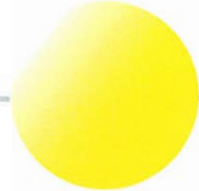


Apparato simile a quello ideato da Alexander Graham Bell.

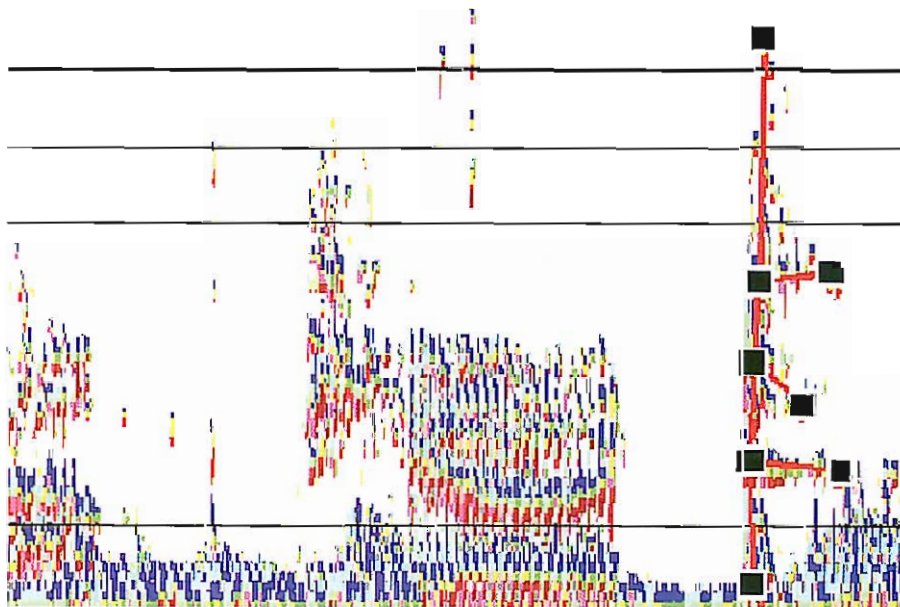


I sistemi di riconoscimento della voce possono essere molto complessi, sia per quanto riguarda il software che per l'hardware, come nel caso dello strumento nella foto che serve per controllare la disfagia.

I primi tentativi di creare una macchina per il riconoscimento della voce furono fatti nel 1950. I laboratori di AT&T idearono un sistema che doveva essere addestrato per riconoscere la voce di ogni interlocutore in modo individuale, però una volta specializzata, la macchina aveva un'affidabilità del 99%. Durante le due decadi successive, i ricercatori lavorarono su applicazioni con vocabolari ridotti, direttamente legati alla persona, che doveva parlare con piccole pause fra parole e frasi. Inoltre fu iniziata anche la ricerca sul riconoscimento della parola continua e si iniziarono a commercializzare applicazioni più specifiche. Negli anni '80 il riconoscimento della voce fu favorito dallo sviluppo dei personal computer e dai ridotti costi dei loro programmi commerciali. In questo



Differenti modi di rappresentare un'onda sonora.



Spettrogramma.

periodo si lavorò allo sviluppo di vocabolari più grandi. Nel 1985 un vocabolario di cento parole era considerato grande, ma già nel 1986 ne fu creato uno di 20.000 parole.

Dal 1990 sino ad ora, i costi delle applicazioni di riconoscimento hanno continuato a scendere, e i vocabolari si sono estesi in ogni loro nuova versione. Ora il software è indipendente dall'interlocutore e il flusso continuo delle parole non presenta alcuna difficoltà.

## Rappresentazione del segnale della voce

Come abbiamo già detto, il segnale della voce si presenta in modo continuo. Tuttavia, il linguaggio trasmesso dalla parlata, è formato da parole, che a loro volta sono divise in fonemi. I fonemi rappresentano l'unità base della parlata. I suoni consistono in variazioni della pressione dell'aria nel tempo, e a frequenze che possiamo ascoltare. Il modo più comune di rappresentare il suono è tramite un'onda (waveform).

Uno dei grandi vantaggi di questo tipo di grafica e che non occupa molto spazio nella memoria, però il suo grande svantaggio è che non descrive esplicitamente le proprietà del contenuto del segnale. Un altro modo di mostrare visivamente i suoni sono gli spettrogrammi, che contengono maggiori informazioni sui dati: sono una trasformazione che mostra la distribuzione delle componenti di frequenza del segnale. Nella figura a fianco possiamo vedere che le maggiori concentrazioni di energia, chiamate formanti, si notano per essere rappresentate con toni più scuri.

