INDICE

INTRODUZIONE

In questo lavoro di tesi verrà presentato un algoritmo per il riconoscimento vocale di parole isolate (isolated speech recognition), e verrà descritto come programmare ed implementare un sistema di questo tipo utilizzando il software MATLAB.
Il riconoscimento vocale (speech recognition) è il processo mediante il quale il linguaggio orale umano viene riconosciuto ed elaborato da un computer. Le applicazioni sono molteplici, dalla telefonia cellulare, alla domotica, fino ai sistemi di dettatura, metodi che consentono la trasposizione scritta del parlato umano tramite il riconoscimento dei fonemi, delle singole parole o tramite associazioni comuni fra parole.
La tecnologia del riconoscimento vocale vede la luce negli anni cinquanta, le tecniche e le metodologie sfruttate sono oggi arrivate a garantire velocità e precisione sempre maggiore.

L'obiettivo di questa tesi è, costruito un database contenente segnali vocali di diverse parole registrate da voci maschili e femminili, di implementare un algoritmo capace, attraverso diverse tecniche e funzioni matematiche, di riconoscere una parola pronunciata, questa parola deve essere una delle voci presenti nel database.

Nel Capitolo 1 verrà trattata l'elaborazione dei dati e delle informazioni nel campo delle telecomunicazioni e dell'informatica, il processo mediante il quale viene effettuata la conversione dei dati in informazioni. Per dato si intende un insieme di numeri o lettere che descrivono il comportamento di un sistema reale. L'informazione non è altro che la risposta ad una determinata serie di dati. Verranno trattati i vari tipi di compressione audio, tecniche che agiscono sulle dimensioni del file, riducendole. Questo procedimento permette la riduzione della memoria necessaria alla registrazione dei dati su di un supporto di memorizzazione e allo stesso tempo riduce i tempi di trasmissione dell'informazione. Verranno descritti poi in maniera approfondita i segnali, sia analogici che digitali, che non sono altro che le grandezze fisiche che permettono la trasmissione e l'elaborazione delle informazioni. Dal momento che l'obiettivo di questa tesi è, dato un segnale analogico (la voce umana), intervenire con strumenti matematici

e informatici allo scopo di identificare una corrispondenza fra più segnali vocali, verrà trattata la tematica della conversione analogico/digitale. E' necessario che il segnale analogico venga trasformato in segnale digitale in maniera tale da poter essere analizzato e processato con l'utilizzo di MATLAB e in generale che sia trasformato in un formato riconoscibile da un computer. Si tratteranno i metodi di campionamento (Naturale ed Istantaneo) e la modulazione con codice a impulsi. Verrà poi descritta l'operazione della Trasformata di Fourier, in quanto è necessario ottenere informazioni sul contenuto frequenziale dei segnali e di conseguenza verrà trattata l'analisi di spettri, attraverso l'utilizzo della funzione che determina lo spettrogramma, che è la rappresentazione grafica dell'intensità di un suono in funzione del tempo e della frequenza, ossia la rappresentazione grafica della funzione reale i (intensità) e delle variabili reali t (tempo) ed f (frequenza).. Successivamente si parlerà di autocorrelazione, che è uno strumento matematico usato nella teoria dei segnali per l'analisi di funzioni o di serie di valori. Il procedimento prevede che il segnale all'istante t venga confrontato con un altro valore di se stesso ritardato di una quantità τ per verificare quanto si somigli o, più precisamente, quanto si correli all'avanzare del tempo. L'autocorrelazione contiene l'informazione relativa alle variazioni sull'asse dei tempi. Questa tecnica si utilizza per cercare pattern che si ripetono all'interno di un segnale, in maniera tale da determinare un segnale periodico che è stato sepolto da un rumore o identificare la frequenza fondamentale di un segnale. Verrà trattata la correlazione incrociata, una tecnica matematica grazie alla quale, attraverso il confronto di due segnali diversi, è possibile ricavarne delle similitudini.

Il Capitolo 2 tratta invece l'elaborazione dei segnali tramite approcci informatici. Uno dei problemi nella cattura di un segnale vocale è spesso legato al rumore, che consiste in un insieme di segnali indesiderati che si sovrappongono all'informazione interessata. Nelle telecomunicazioni esistono diversi metodi per l'eliminazione del rumore, alcuni di questi sono stati trattati in questa tesi. Per la cattura di un file audio da utilizzare in un sistema di riconoscimento vocale viene utilizzato un microfono, il segnale registrato sarà dunque soggetto a distorsioni e disturbi, verranno illustrate tecniche per la pulizia dei segnali da ronzii e rumori di fondo, quali equalizzatori e filtri, strumenti utili all'eliminazione di frequenze indesiderate. Se si riesce ad identificare la frequenza del segnale rumore, e se queste frequenze non sono simili alle frequenze della voce umana,

è possibile ottenere in uscita al filtro un segnale uguale il più possibile a quello originale.

Un altro importante argomento è l'eliminazione del silenzio all'interno dei file audio, trattato anch'esso trattato nel Capitolo 2. A tal proposito verrà descritto un algoritmo che, attraverso un ciclo di iterazioni, dividerà il segnale il più frame; se all'interno del frame considerato il segnale presenterà un'ampiezza minore o uguale ad un'ampiezza impostata precedentemente, l'algoritmo andrà a cancellare quella porzione di frame considerato non utile in quanto contenente silenzio. Successivamente verranno illustrati gli strumenti matematici e le funzioni utilizzate in MATLAB che permettono l'audio processing per l'elaborazione e l'analisi dei segnali, quali DTF, FFT e la funzione che permette di ricavare lo spettro di un segnale audio.

Nel Capitolo 3 verranno descritti i principali metodi di riconosicmento vocale, dalla nascita ad oggi, le principali applicazioni, le differenziazioni fra diversi tipi di riconoscimento vocale, gli algormtmi generali utilizzati in MATLAB. Verrà descritto come sviluppare un generico sistema per il riconoscimento vocale isolato in MATLAB. Attraverso l'acquisizione e l'analisi dei dati e le successive fasi di teaching-training-testing, verrà descritto come implementare lo scheletro per un sistema di riconoscimento vocale. Verranno descritte le tecniche e le metodologie moderne basate su sistemi più complessi che utilizzano i modelli di Markov, sistemi dotati di un numero finito di *n* stati e regolati da probabilità.

Verrà descritta nel dettaglio la tecnica del Dynamic Time Warping, un algoritmo che si occupa dell'allineamento temporale fra i segnali. Con questa tecnica è possibile comparare due segnali trovando corrispondenze molto precise, questo metodo è basato su un algoritmo di allineamento di serie temporali. Esso mira ad allineare due sequenze di vettori tramite una deformazione dell'asse temporale in modo iterativo fino a quando non si trova una corrispondenza ottimale tra le due sequenze.
Verranno trattati i sistemi di riconoscimento vocale basati sulle Reti Neurali, modelli che mirano a riprodurre il comportamento del cervello umano. Imitando il comportamento e la fisionomia di un neurone come una funzione a più variabili o comunque un sistema con più ingressi e uscite. Una volta costruita questa rete si ha bisogno di effettuare ripetuti test per la fase di addestramento. Il sistema dovrà

presentare una data risposta in base agli ingressi immessi nel sistema. Dopo la fase di addestramento il sistema diventa indipendente e accurato.

Nel Capitolo 4 verrà descritto l'algoritmo realizzato in questa tesi per il riconoscimento vocale. Si parte dalla creazione del database, una serie di segnali vocali che contengono diverse informazioni, grazie alle analisi effettuate sul database il sistema può essere in grado di riconoscere una determinata parola. L'informazione viene trattata dal software MATLAB sotto forma di vettore di elementi. Verranno trattare poi le tecniche che il sistema utilizza, quali la divisione in frame e l'utilizzo della correlazione incrociata. Dati due segnali, questi verranno suddivisi in frame. La divisione in frame viene effettuata tramite il richiamo di un particolare metodo, che attraverso una serie di iterazioni, andrà a suddividere il segnale in tante porzioni di uguale lunghezza e andrà a compararli attraverso la funzione della correlazione incrociata. I valori ricavati dalla correlazione di ogni singolo frame confrontato verranno sommati di volta in volta in valore assoluto e salvati. Quando il sistema avrà concluso la comparazione della parola di riferimento con tutti i segnali contenuti all'interno del database, si andranno a fare delle considerazioni in base ai risultati ottenuti. Il risultato di tutte le correlazioni è un vettore di dimensione pari al numero di parole all'interno del database, ad esempio, se il database è formato da 10 parole diverse, ognuna delle quali sarà pronunciata 50 volte, il vettore finale avrà dimensione 10. Gli elementi del vettore sono i risultati delle varie correlazioni. Il sistema andrà ad identificare il minimo valore contenuto nell'array e lo associerà alla parola contenuta nel database corrispondente, in quanto il minimo valore rappresenta la migliore corrispondenza che il sistema ha identificato. Infine verranno commentati i test effettuati e i risultati ottenuti.