UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI INGEGNERIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

TESI DI LAUREA SPECIALISTICA IN

INGEGNERIA INFORMATICA

Un Sistema Biometrico Per la Verifica di Identità Basato sul Riconoscimento del Parlato

Laureando: KAMGANG SIMEU CHRISTIAN

Relatore: Prof. Carlo Ferrari

Correlatore: Prof. Federico Avanzini

Anno Accademico 2012-2013

"C'è vero progresso solo quando i vantaggi di una nuova tecnologia diventano per tutti". ***Henry. Ford***

A mia Madre.
Alla mia famiglia
A Patricia
Agli amici
Grazie del cuore.

Christian Simeu

Indice

Introduzione

Le tecniche d'identificazione tradizionale si basino essenzialmente su qualcosa di cui si è in possesso, come ad esempio una carta magnetica, un dispositivo di memoria oppure su qualcosa che si conosce, come ad esempio una password o un codice pin; nessun di tale sistema di autenticazione può dirsi realmente efficace se non con la stretta collaborazione dell'utente. E' infatti cura dell'utente il ricordo dell'eventuale password, la conservazione del badge e il non comunicare ad altri gli estremi del proprio sistema di autenticazione.

Con la crescente richiesta di sicurezza, si fa strada a due concetti fondamentali: quello dell'identità e quello della non ripudiabilità. In altri termini, se la tecnica di identificazione è sufficientemente sicura, allora oltre a garantire l'identità di colui che accede, implicherà che quest'ultimo non potrà negare di aver avuto accesso alle risorse che l'autenticazione gli garantiva.

Le tecniche biometriche di identificazione sono infatti finalizzate a identificare un individuo sulla base delle sue peculiari caratteristiche fisiologiche o comportamentali, difficili da alterare o simulare. Un sistema biometrico è in quindi in grado di riconoscere una persona (cioè verificare se un individuo è veramente colui che dichiara di essere) sulla base di caratteristiche fisiologiche (impronta digitale, forma del volto o della mano, retina, iride, timbro di voce, vene...) e/o comportamentali (calligrafia, stile di battitura...).

I sistemi di riconoscimento biometrico si stanno rapidamente diffondendo a livello globale in quanto sono in grado di offrire maggiore sicurezza rispetto ai sistemi di autenticazione tradizionale, e sono ancora costosi. I costi sono principalmente legati alla tecnologia usata per la realizzazione del dispositivo di rilevazione dei dati biometrici.

Con questa tesi ho voluto esplorare ed approfondire il riconoscimento biometrico basato sul parlato perché ormai la tecnologia utilizzata per l'implementazione di questi sistemi biometrici è alla portata di tutti, c'è la possibilità di implementare tale sistema su larga scala per l'identificazione dell'utente sia al livello privato che al livello pubblico. Inoltro, i sistemi biometrici di verifica della voce godono di alto grado di apprezzamento rispetto ad altri sistemi biometrici. In questa tesi, verrà presentato gli algoritmi per la realizzazione di un sistema biometrico per la verifica di identità basato proprio sui dati acquisiti dal microfono dei nostri personal computer oppure smartphone.

Nel primo capitolo si effettua una panoramica sulle principali tecniche biometriche introducendo le motivazioni che ci spingono verso un approccio biometrico nel settore della sicurezza e le problematiche che ne derivano.

Nel secondo capitolo si approfondisce la tecnica di verifiche basato sulla voce.

Nel terzo capitolo e quarto capitolo si presentano i vari processi e algoritmi utilizzati per l'implementazione di questo progetto, approfondiremmo l'argomento sull'estrazione delle impronte vocali mediante i mel frequency ceptrum coefficients e la misura di similarità tra due impronte vocali attraverso il dynamic time warping.

Nel quarto capitolo e ultimo capitolo, si analizzano le prestazioni della nuova applicazione biometrica, si prospettano possibili scenari futuri e vengono fatte alcuni considerazioni sulle prospettive di sviluppo.

1 Biometria e Sistemi di Riconoscimento del Parlatore

1.1 Biometria

Oggi viviamo in una società, dove c'è sempre più gente disperata e pericolosa di cui non possiamo fidarci basandosi sui documenti dell'identificazione come la carta di identità, la patente di guida, in quanto questi documenti possono essere compromessi. Uno dei limiti dei sistemi informatici è l'incapacità di riconoscere con certezza se l'utente al quale viene concesso un certo diritto è effettivamente colui che lo sta usufruendo. Il furto d'identità, le frodi, la pirateria informatica e i virus stanno ponendo enormi difficoltà altrettanto ai singoli che alle aziende e i governi costringendoli ad adottare soluzioni efficaci per proteggere i dati dal furto.

Le tecniche d'identificazione più comune fano riferimento:

- Su qualcosa di cui si è in possesso, come ad esempio una carta magnetica, un dispositivo di memoria oppure,
- su qualcosa che si conosce, come ad esempio una password o un codice pin;

Questi metodi possono essere anche sovrapposti per formare un sistema più forte come nel caso per esempio del bancomat, in cui è previsto il possesso di una carta e la memoria di un pin.

Però nella vita di tutti i giorni, sono così numerose le password da memorizzare (carta di credito, accesso a porte, controllo del computer dell'auto...). Gli utenti del Web hanno una media di ventuno password; 81% degli utenti selezionano una password comune (per esempio, nome più anno di nascita) e 30% annotano le loro password.

Tali dati (carta magnetica, password …) possono essere non soltanto sottratti oppure letti da una terza persona ma anche essere perse o dimenticate dall'utente stesso. C'è la necessità di risolvere questo problema, infatti, la maggior parte degli sforzi della comunità scientifica e della ricerca industriale si è orientata allo studio di quelle variabili che permettono l'identificazione affidabile degli individui. Le tecniche

biometriche mirano ad aggirare queste limitazioni sfruttando il rilevamento delle caratteristiche umane ritenute uniche da individuo a individuo.

"La biometria (dalle parole greche bìos = "vita" e metros = "conteggio" o "misura") è la disciplina che ha come oggetto di studio la misurazione delle variabili fisiologiche o comportamentali tipiche degli organismi, attraverso metodologie matematiche e statistiche". La biometria è quindi la disciplina che studia la misurazione dei tratti fisiche o comportamentali tipiche degli esseri umani.

Le caratteristiche biometriche possono essere divise in due classi principali:

a) Fisiologico; collegato con la forma del corpo. Impronte digitali, il volto, la geometria della mano e l'iride sono alcuni esempi di questa classe.
b) Comportamentale, collegato al comportamento di una persona. Alcuni esempi in questo caso sono: la firma, la scrittura, la camminata e la voce. A volte la voce è considerata come un tratto biometrico fisiologico.

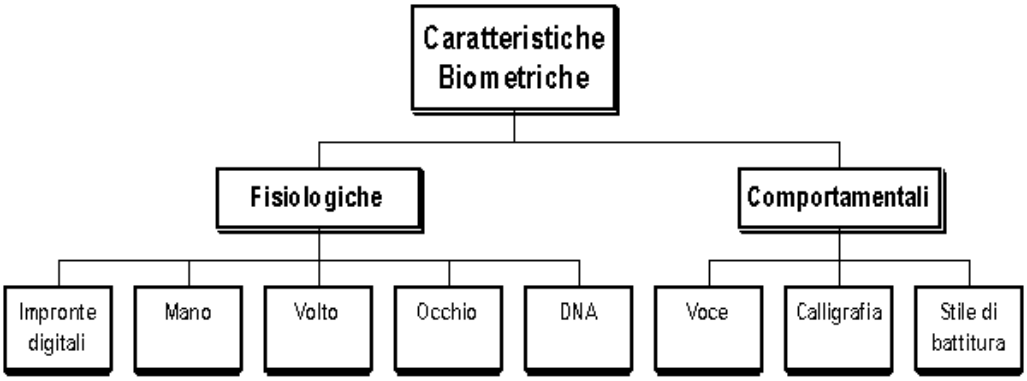

Figura 1.1 Classificazione delle caratteristiche biometriche

Figura 1.2 Esempi di tratti biometrici

I dati biometrici di un essere umano sono derivabili dalla misurazione di varie caratteristiche del corpo o del comportamento, perciò sono più difficilmente falsificabile rispetto ad una password o ad una chiave. Inoltre, i tratti biometrici sono caratteristiche peculiari di ogni individuo e non possono essere dimenticati o rubati e usati da altre persone.

In ogni caso, le caratteristiche che deve avere una variabile per essere ritenuta "biometrica" sono:

- Universalità: tutti devono averla;
- Unicità: due o più individui non possono avere la stessa uguale caratteristica;
- Permanenza: questa caratteristica biometria deve rimanere invariante nel tempo;
- Collezionabilità: deve essere misurata quantitativamente;

I sistemi di riconoscimento biometrico si stanno rapidamente diffondendo a livello globale poiché sono in grado di offrire maggiore sicurezza rispetto ai sistemi di autenticazione tradizionali.

1.2 Storia della Biometria

Per migliaia di anni gli uomini hanno istintivamente utilizzato alcune caratteristiche fisiche (come il volto, la voce, il portamento ecc.) per riconoscersi gli uni con gli altri. Circa a metà dell'800, A. Bertillon, allora capo della sezione di identificazione criminali della polizia di Parigi, sviluppò l'idea di usare alcune misure del corpo umano (altezza, lunghezza delle braccia, piedi, dita, ecc.) per identificare i responsabili dei crimini. Verso la fine del XIX secolo, questa idea di partenza fu ulteriormente sviluppata grazie alla scoperta (dovuta agli studi di F. Galton e E. Henry) del carattere distintivo delle impronte digitali, ovverosia del fatto che queste individuano biunivocamente una persona.

Subito dopo questa scoperta, le polizie di tutto il mondo cominciarono ad acquisire e memorizzare in appositi archivi le impronte digitali di criminali, detenuti e sospetti. Inizialmente, le impronte erano "registrate" su supporto cartaceo, inchiostrando i polpastrelli dei soggetti in questione e realizzando il "timbro dell'impronta". Subito dopo questa fase, le forze di intelligence e di pubblica sicurezza perfezionarono le loro tecniche per il rilievo, sulle scene del crimine, delle impronte digitali lasciate dai malviventi.

In questi anni, la polizia comincia a fare sempre più affidamento su tecniche di indagine scientifiche, che si affiancano e quelle tradizionali (logica deduttiva) nelle investigazioni. Segni evidenti di questo nuovo "approccio scientifico" nel condurre le indagini si riscontrano anche in alcuni famosi personaggi della letteratura poliziesca (per tutti, Sherlock Holmes).

La scienza biometrica comincia, quindi, a essere impiegata nell'attività giudiziaria e anticrimine, così come in applicazioni inerenti alla sicurezza di un numero sempre crescente di persone. Oggi, in piena era digitale, un numero elevatissimo di persone utilizza tecniche di riconoscimento biometrico, non solo nel campo della giustizia,

ma anche in applicazioni civili e militari. Nel futuro, si prevede che la maggior parte degli abitanti della Terra avrà a che fare, episodicamente o in maniera continua, con le tecniche di riconoscimento biometrico.

1.3 Sistemi Biometrici

Un sistema di riconoscimento biometrico vuole garantire l'unicità della persona, infatti, è un sistema informatico che ha la funzionalità e lo scopo di riconoscere una persona sulla base di una o più caratteristiche fisiologiche e/o comportamentali, confrontandole con i dati, precedentemente acquisiti e presenti nel database del sistema, tramite degli algoritmi e di sensori di acquisizione dei dati in input.

Secondo la caratteristica biometrica in esame, i sistemi biometrici sono divisi in due classi:

- Sistemi che utilizzano caratteristiche fisiologiche: quindi sfruttano le peculiarità fisiche dell'individuo quali, le impronte digitali, la geometria della mano, la fisonomia del volto.
- Sistemi che utilizzano caratteristiche comportamentali quali la voce, lo stile di battitura della tastiera e la firma;

I sistemi biometrici basati sulle caratteristiche fisiologiche sono solitamente più affidabili di quelli basati sulle caratteristiche biometriche del tipo comportamentali poiché le caratteristiche fisiologiche sono più difficili da ripetersi e non sono influenzati dalle condizioni psicologico/fisiche dell'individuo quali lo stress o la malattia.

1.3.1 Funzionalità dei Sistemi Biometrici

A secondo della loro finalità, i sistemi biometrici possono operare in due modalità differenti: verificazione o addestramento.

- In verificazione, l'obiettivo del sistema è di confermare o di negare l'identità dichiarato dall'utente;
- In identificazione, l'obiettivo del sistema è di riconoscere un individuo all'interno di un insieme di N identità.

Durante il processo di verifica il soggetto dichiara la sua identità, il sistema quindi effettua un confronto tra l'immagine rilevata in tempo reale e quella di riferimento presente nell'archivio del sistema.

Nella fase di identificazione, l'immagine acquisita in tempo reale viene confrontata con tutte le immagini presenti nel database del sistema e viene poi associata a quella con le caratteristiche più simili. Nell'identificazione si faccia uso di uno schema di confronto uno a molti, mentre nella verifica è sufficiente quello uno a uno. Dal punto di vista della sicurezza, l'identificazione è diversa dalla verifica. Per esempio presentare il passaporto all'imbarco di un aeroporto è un processo di verifica - il personale confronta la faccia dell'individuo con la fotografia nel documento.

Viceversa il poliziotto che confronta l'identikit di un malvivente con un database di criminali precedentemente archiviato è un processo di identificazione. Nelle applicazioni forensi è comune effettuare prima il processo di identificazione, per creare una lista di migliori candidati e quindi una serie di processi di verifica per determinare il risultato finale.

1.3.2 Architettura dei Sistemi Biometrici

Ogni dispositivo di riconoscimento biometrico ha bisogno di una fase di addestramento o di registrazione detta “enrollment” nella quale il sistema identifica le caratteristiche biometriche dell’utente e lo memorizza nel database del sistema in forma di template; e di una fase di autenticazione nella quale si confronta la nuova impronta biometrica o con un’altra di riferimento presente nel database se si tratta di un sistema di verifica; o con tutte quelle presenti nel database se sistema di identificazione biometrico.

Nell’addestramento, l’utente che desideriamo inserire nel sistema si identifica e tramite il sensore (laser, scanner, telecamere, microfoni…) si acquisisce un’immagine o un suono che si riferisce al tratto biometrico dell’individuo, la qualità dei dati acquisiti è cruciale per le successive autenticazione; dunque durante questa fase si può avere bisogno della presenza di un personale specializzato che dia indicazioni agli utenti, che avendo in generale poca dimestichezza con questo tipo di rilevazioni, si mostrano diffidenti verso le apparecchiature compromettendo il buon esito della rilevazione. I dati rilevati dal sensore sono poi elaborati grazie ad un algoritmo di estrazione delle caratteristiche pertinenti che varia da sistema a sistema per così creare una impronta biometrica chiamato “biometric template”. Il template viene quindi memorizzato nel sistema in modo tale da essere utilizzato come riferimento durante le successive fasi di autenticazioni.

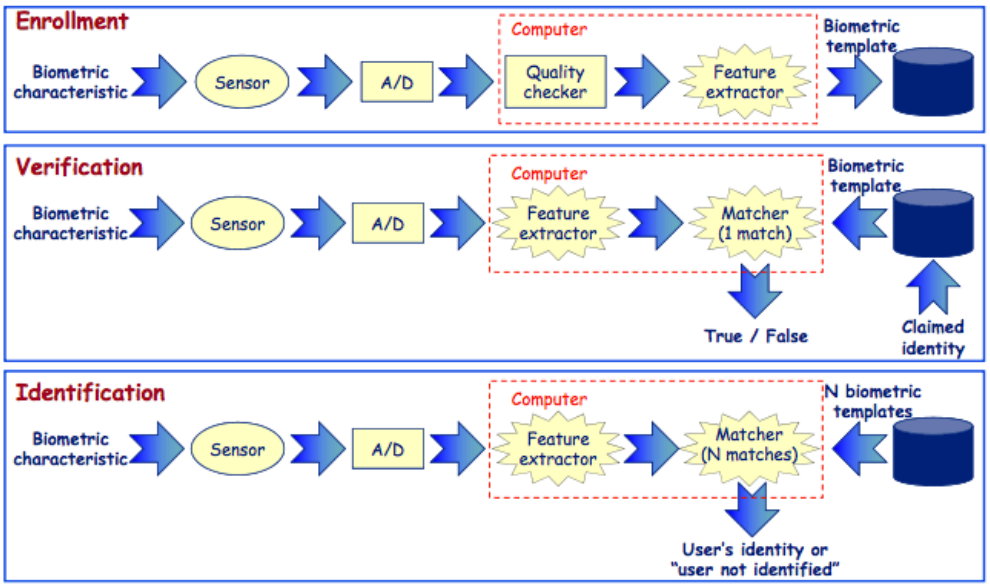

Figura 1.3 Architettura di un sistema biometrico

Nella seconda fase, mi riferisco a quella di autenticazione, si procede come nella fase di enrollment alla creazione di una nuova impronta biometrica a partire dell'immagine acquisita all'istante, la nuova impronta, nel caso di identificazione è confrontata con tutti template presenti nella database del sistema per individuare gli utenti registrati avente la stessa caratteristica biometrica; nel caso di verificazione di identità, si confronta la nuova impronta con un'altra presente nel sistema e di cui l'utente si dichiara l'identità.

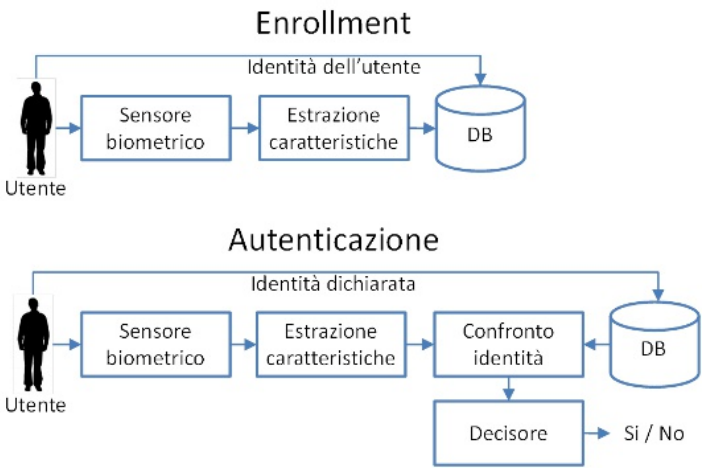

Figure 1.4: Moduli di un sistema di riconoscimento biometrico

Nella figura 1.4, viene rappresentato il modo in cui sono ordinati i differenti moduli che costituiscono un sistema di riconoscimento, i moduli sono principalmente cinque: il sensore, il modulo di estrazione delle caratteristiche, il modulo di confronto, la base di dati e il decisore.

- Il sensore designa un dispositivo di input come scanner, fotocamere, microfono, con scopo di rilevare l'immagine o il segnale della nostra variabile biometrica.
- Il modulo di estrazione delle caratteristiche si occupa di elaborare l'immagine o il segnale acquisito dal sensore, ed estrarne le caratteristiche pertinenti;
- Il modulo di confronto delle identità esegue il confronto tra due immagini e restituisce una misura di somiglianza;
- Il decisore che prende una decisione booleana in base a una soglia prefissata dal sistema, infatti nel caso della verifica di identità, la decisione booleana è positiva se la misura di somiglianza è minore o uguale alla soglia e negativa nel caso contrario, se invece siamo di fronte ad sistema di identificazione, il decisore ci può anche restituire tutti gli utenti di cui la misura di somiglianza è minore della soglia;
- Il database contiene tutti i modelli dei dati, chiamati più comunemente templates, su cui eseguire il confronto. Un tipico problema nella progettazione di un sistema consiste nello scegliere il tipo di supporto e la struttura dati destinati alla memorizzazione dei dati, in quanto il riconoscimento viene effettuato confrontando i dati biometrici della persona in questione con quelli memorizzati.

1.3.3 Principali Tratti biometrici

I principali tratti biometrici attualmente sono le impronte digitali, le caratteristiche del volto, la geometria della mano, le caratteristiche dell'iride e della retina.

I sistemi biometrici basati sulle impronte digitali sono quelli maggiormente diffusi, oltre ad essere stati i primi ad essere impiegati su larga scala. L'impronta digitale è rappresentata da un pattern di creste e valli presente sulle dita, che si sviluppa da una configurazione causale già presente dall'embrione. Il riconoscimento biometrico è effettuato confrontando caratteristiche come la tipologia globale dell'impronta, la posizione e la tipologia di alcuni punti distintivi, l'orientamento e frequenza delle creste, la posizione ed il tipo delle minuzie (terminazioni, biforcazioni delle creste).

Il riconoscimento del volto rappresenta uno dei sistemi biometrici più apprezzati dagli utenti. Ciò è dovuto alla scarsa invasività del processo di acquisizione del tratto biometrico ed all'abitudine umana di riconoscere un individuo in base all'osservazione del suo viso. I sistemi biometrici basati sul volto possono utilizzare caratteristiche globali o misurazioni locali. Esempi di caratteristiche globali sono le autofacce, ottenute come le differenze tra l'immagine del volto acquisita ed il volto medio di una base di dati biometrica. Caratteristiche locali sono invece le informazioni geometriche, ottenute misurando le distanze relative tra punti distintivi come occhi, bocca e naso.

L'iride è considerato il tratto biometrico più accurato. L'iride umana è infatti caratterizzata da un pattern casuale, stabile per l'intera durata della vita di un individuo e dotato di numerose caratteristiche distintive. Oltre ad un'elevata accuratezza, il processo di riconoscimento dell'iride è molto veloce. Ciò ha consentito una rapida diffusione dei sistemi di riconoscimento basati su questo tratto

biometrico in contesti applicativi caratterizzati da un elevato numero di utenti, come frontiere o aeroporti. Un limite alla diffusione di questa tecnologia consiste nel fatto che il processo di acquisizione delle immagini iridee viene considerato invasivo e pericoloso per la vista da parte di numerosi utenti. I sistemi biometrici basati sull'iride sono inoltre relativamente costosi. La maggior parte di questi sistemi biometrici è basata sul calcolo di una stringa binaria che ne incorpora le caratteristiche distintive, chiamata Iriscode.

I sistemi biometrici basati sulla geometria della mano, pur fornendo un'accuratezza inferiore a quella dei sistemi basati su impronte digitali o iride, sono apprezzati dagli utenti e ritenuti poco invasivi. Il metodo più diffuso per il riconoscimento biometrico è basato sull'acquisizione di una fotografia della mano mentre essa è posizionata su un supporto (eventualmente con l'ausilio di pioli per aiutare il corretto posizionamento). Successivamente, sono effettuate misurazioni delle dimensioni della mano, come, ad esempio, la lunghezza e larghezza delle dita e del palmo .

I sistemi basati sulla retina sfruttano l'unicità dei pattern delle vene presenti sulla zona posteriore del bulbo oculare per effettuare il riconoscimento biometrico. La distribuzione dei vasi sanguinei sulla retina è infatti principalmente casuale ed univoca per ogni individuo. Tra i principali vantaggi, è da annoverare che questo tratto biometrico è difficilmente falsificabile, in quanto la parte esaminata si trova all'interno dell'occhio. Per lo stesso motivo, però, la scansione della retina viene vista come intrusiva e potenzialmente dannosa.

Esistono inoltre sistemi biometrici in grado di sfruttare contemporaneamente informazioni inerenti a differenti tratti. Questa tipologia di sistemi biometrici consente di ottenere una maggiore accuratezza nel riconoscimento e risulta maggiormente robusta a tentativi di frode rispetto ai sistemi basati su un unico tratto biometrico.

1.4 Valutazione delle Prestazione dei Sistemi Biometrici

Nell'identificazione infatti, possiamo trovarci di fronte a due situazioni:

1. Riconoscimento positivo, il quale prevede le due possibili situazioni: la persona è vera oppure è un impostore;
2. Riconoscimento negativo, e in questo ultimo caso il sistema o ha sbagliato dando un falso allarme oppure la persona è realmente un impostore.

Di conseguenza abbiamo due tipologie di errore:

- FRR (False Rejection Rate) è la percentuale di falsi rifiuti, cioè utenti autorizzati ma respinti per errore, in pratica il sistema non riesce a riconoscere le persone autorizzate.
- FAR (False Acceptance Rate) è la percentuale di false accettazioni, per utenti non autorizzati ma accettati per errore, il sistema quindi accetta le persone che non sono autorizzate.

FAR e FRR sono due grandezze strettamente correlate dalla seguente proprietà: al diminuire dell'una cresce l'altra. Ogni Sistema biometrico offre la possibilità di regolare il rapporto FRR/FAR e quindi di aumentare o diminuire la sensibilità del sistema.

Definiamo la variabile t come il grado di tolleranza del sistema che serve a definire la bontà del sistema in termini di sicurezza. Con un basso grado di tolleranza si ha un numero elevato di false accettazioni, mentre con un alto grado di tolleranza si ha un numero elevato di falsi rifiuti.

Una volta definito t costruiamo le funzione FAR(t) (monotona non crescente) e FRR(t)(monotona non decrescente) tramite le quali è possibile calcolare ERR (Equal Error Rate) che rappresenta l'errore intrinseco del sistema, per il quale FAR(t*) = FRR(t*) = ERR

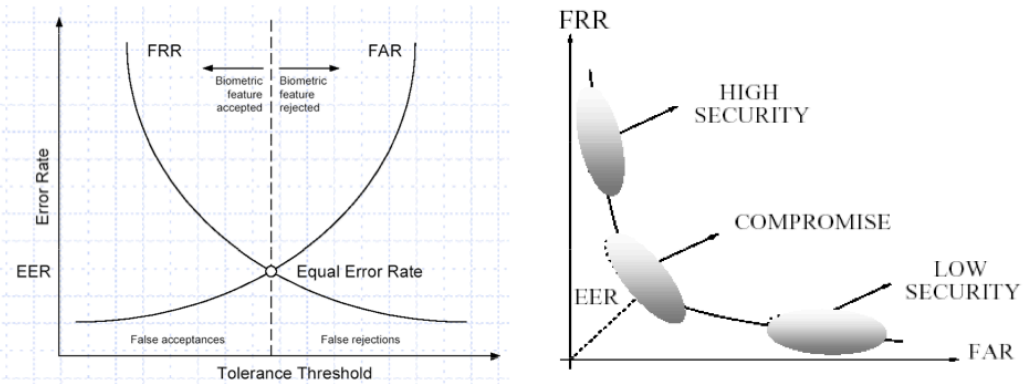

Figura 1.3: Curve di errori dei sistemi biometrici

Il grado di tolleranza t* rappresenta il punto di equilibrio del sistema attraverso il quale è possibile regolare il rapporto FRR/FAR. Infatti nel punto t* tale rapporto risulta essere pari ad 1; quindi per valori t>t* tale rapporto diminuisce mentre per valori t<t* tale rapporto si incrementa. Per questo motivo, a seconda della tipologia di applicazione, è opportuno valutare questo aspetto. Ad esempio al tornello di ingresso di una banca posso tenere alto il rischio di false accettazioni per non fare perdere tempo ai clienti, mentre nel caveau della banca è conveniente aumentare il rischio di false rifiuti per aumentare la sicurezza e ridurre al minimo il false accettazioni. Nelle applicazioni reali i valori di tolleranza si trovano al di sotto di t* per garantire un numero ridotto di false accettazioni.

Più selettivo è il riconoscimento per ragioni di sicurezza più alta è la probabilità che una persona autorizzata venga respinta. Il punto in cui le due curve caratteristiche relative all'accettazione degli impostori ed al rifiuto delle persone autentiche s'intersecano, ossia quando i due tassi di errore si equivalgono, corrisponde all'Equal Error Rate (ERR).

1.5 Applicazioni

Le tecnologie biometriche vengono sempre più frequentemente usate per lo sviluppo di soluzioni di identificazione e verifica dell'identità altamente sicure. L'utilizzo delle tecniche biometriche avviene attraverso un Sistema di riconoscimento biometrico, questi sistemi si stanno espandendo in diversi settori, con costi ancora abbastanza elevati.

Le tecniche biometriche di verificazione dell'identità possono essere applicate sia al controllo dell'accesso a luoghi ed informazioni, sia all'autenticazione di informazioni, in sostituzione di sistemi nome utente/parola chiave, o di dispositivi elettronici o meccanici aventi funzione di chiave. Possono essere utilizzata da sola o integrata con altre tecnologie - come le smart card, le chiavi di crittografia e le firme digitali - la tecnologia biometrica è destinata a permeare molti aspetti dell'economia e della vita quotidiana. Attualmente sono infatti sempre più numerosi i prodotti dell'elettronica di consumo, come portatili, PDA, cellulari o lettori MP3, che utilizzano sistemi biometrici integrati per il controllo e l'identificazione.

Esistono richieste pratiche per le soluzioni biometriche? Al giorno d'oggi, gli utenti non ha alcun problema a usare sistemi di riconoscimento basati su tratti biologici anziché su password. Nella vita di tutti i giorni, sono così numerose le password da ricordare che è spesso molto più rapido e semplice strisciare un dito su un pannello che dover ricordare e immettere un'ulteriore password. Ciò spiega perché l'uso della biometria per l'autenticazione si sta rapidamente imponendo rispetto ad altri metodi attuali come le password o le smart card.

Tra i principali tratti biometrici attualmente utilizzati, è possibile annoverare le impronte digitali, le caratteristiche del volto, la geometria della mano, le caratteristiche dell'iride e della retina. La firma e la voce rappresentano senza dubbio quelli maggiormente impiegati nei sistemi biometrici dinamici.

Abbiamo già detto che le caratteristiche che una grandezza deve avere per essere ritenuta "Biometrica" sono: l'universalità, l'unicità, la permanenza e la collezionabilità; Ci sono, inoltre, altre importanti parametri che possono incidere nella scelta di un sistema piuttosto che di un altro:

- Performance: Indica le risorse richieste, le modalità e l'ambiente che determina una più accurata identificazione.
- Grado di Gradimento: Indica il grado con il quale ogni individuo accetta la metodologia biometrica impiegata.

Caratteristiche biometriche	Universalità	Unicità	Persistenza	Collezionabilità	Prestazioni	Accettabilità	Contraffazione
DNA	H	H	H	L	H	L	L
Orecchio	M	M	H	M	M	H	M
Volto	H	L	M	H	L	H	H
Termogramma facciale	H	H	L	H	M	H	L
Impronta	M	H	H	M	H	M	L
Andatura	M	L	L	H	L	H	M
Geometria della mano	M	M	M	H	M	M	M
Vene della mano	M	M	M	M	M	M	L
Iride	H	H	H	M	H	L	L
Stile di battitura	L	L	L	M	L	M	M
Odore	H	H	H	L	L	M	L
Retina	H	H	M	L	H	L	L
Firma	L	L	L	H	L	H	H
Voce	M	L	L	M	L	H	H

Figura 1.4: Confronto fra le principali caratteristiche biometriche.

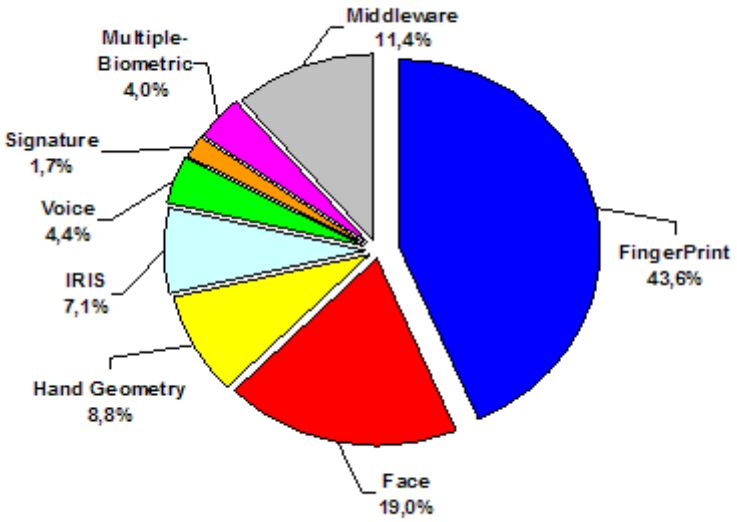

Figura 1.5 Suddivisione del mercato biometrico in base alla tecnologia (2006)

Quale tecnica biometrica deve essere usata per una data applicazione? La scelta scaturisce da un'analisi delle richieste che applicazioni richiede, delle caratteristiche dell'applicazione e delle particolari proprietà della tecnologia biometrica.

Si deve potere individuare il proprio bisogno rispondendo alle seguenti domande:

- L'applicazione necessita di un percorso di identificazione o di uno di autenticazione?
- Il processo deve essere semiautomatico o completamente automatico?
- Gli utenti sono "abituabili"? Cioè accettano di buon grado la particolare tecnologia scelta?
- Qual è la richiesta di memoria richiesta? Ogni differente applicazione impone limiti di grandezza nella rappresentazione interna delle caratteristiche biometriche scelte.
- Quanta rigorosità è richiesta nell'identificazione/autenticazione? E' chiaro che più l'applicazione richiede un alto grado di precisione e più è necessario individuare caratteristiche biometriche uniche.

In conclusione, la scelta del sistema biometrico più idoneo in una applicazione è un compromesso tra una serie di fattori non ultimi quelli economici in rapporto alle prestazioni.

1.6 Sicurezza e Privavy

Come per molti altri sistemi informatici provvisti di database e dati sensibili di utenti, un sistema di riconoscimento biometrico può essere soggetto ad attacchi che possono minare uno o più dei tre requisiti classici della sicurezza informatica ovvero la confidenzialità dei dati, l'integrità dei dati e la disponibilità dei dati minando quindi anche il funzionamento stesso del sistema di riconoscimento.

In generale, le procedure di acquisizione dei dati biometrici sono relativemente ben accettate dalla popolazione, che percepisce il maggiore grado di sicurezza offerto da tali tecnologie. La maggiore sicurezza si può tradurre, però, nella sensazione di una riduzione della propria privacy, in quanto tali tecnologie consentono il riconoscimento univoco di una persona e possono provocare in alcuni individui la sensazione di essere schedati come avviene nelle attività di polizia investigativa e/o costantemente monitorati. Alcuni tratti biometrici, inoltre, possono mettere in luce alcune informazioni personali o problemi di salute. Le immagini dell’iride, per esempio, possono rilevare alcune malattie ed il DNA può rivelare informazioni come sesso e parentele di un individuo.

A differenza di una password, non è possibile modificare i propri dati biometrici. Nel caso in cui tali dati vengano acquisiti da malintenzionati, le problematiche inerenti a furti di identità risultano quindi amplificate. Il medesimo tratto biometrico di un individuo, inoltre, potrebbe essere utilizzato al fine di consentire l’accesso a differenti sistemi o edifici e potrebbe essere legato a documenti di identità del possidente. Una persona in possesso dei dati biometrici di un individuo, potrebbe quindi impersonare quest’ultimo in un elevato numero di situazioni. Un’ulteriore problematica consiste nel fatto che risulta estremamente complesso ripudiare le azioni svolte quando l’autenticazione viene effettuata utilizzando sistemi di riconoscimento biometrico.

Per questi motivi, risulta necessario utilizzare una serie di accorgimenti nella progettazione e nella gestione dei sistemi biometrici. Nel momento in cui un utente viene registrato in una base di dati biometrica, egli dovrebbe essere consapevole degli obiettivi e delle funzionalità del sistema stesso. Tali obiettivi e funzionalità, inoltre, non dovrebbero essere successivamente estesi senza il consenso esplicito dell'utente. I dati biometrici non dovrebbero infatti essere diffusi o utilizzati per fini non esplicitamente dichiarati.

Un ulteriore accorgimento consiste in una gestione accorta dei dati biometrici memorizzati. Un sistema biometrico dovrebbe infatti memorizzare la minima quantità di informazione necessaria. Ciò implica che i campioni del tratto biometrico non dovrebbero essere conservati, in quanto risulta sufficiente memorizzarne il template. Secondo questo principio, inoltre, le informazioni relative ai dati biometrici ed agli accessi degli utenti non dovrebbero essere collegabili. Ulteriormente, il periodo temporale di memorizzazione dei dati biometrici dovrebbe essere limitato e noto all'utente. Infine, gli utenti i cui tratti biometrici sono registrati nel sistema dovrebbero essere sempre in grado di modificare o eliminare i propri dati.

Anche la scelta del tratto biometrico utilizzato da un sistema di riconoscimento risulta importante in termini di protezione della privacy. Tratti biometrici che non mutano per lunghi periodi di tempo e garantiscono elevata accuratezza nel riconoscimento dovrebbero essere utilizzati solo per applicazioni che richiedono elevati standard di sicurezza. In situazioni di minore criticità, dovrebbero essere utilizzati tratti maggiormente mutevoli o in grado di fornire minore accuratezza. In un aeroporto, ad esempio, potrebbe essere necessario utilizzare un sistema di riconoscimento dell'iride. In uno stadio o in un parco di divertimenti, invece, sarebbe preferibile gestire gli accessi utilizzando sistemi basati sul volto o sulla geometria della mano.

Durante la fase di progettazione di un sistema biometrico, inoltre, risulta necessario valutare con attenzione le caratteristiche del contesto applicativo. Sistemi posti all'aperto, ad esempio, possono essere maggiormente soggetti ad attacchi rispetto a sistemi posti in ambienti protetti. Allo stesso modo, sistemi sorvegliati fisicamente risultano più sicuri rispetto a sistemi non sorvegliati. La modalità di accesso utilizzata, inoltre, influisce sulla sicurezza del sistema. In generale, i sistemi che effettuano l'autenticazione risultano più difficilmente frodabili rispetto a quelli che effettuano l'identificazione.

Infine, lo scambio e la memorizzazione dei dati biometrici dovrebbero essere protetti da eventuali attacchi informatici attraverso l'uso di tecniche crittografiche e di protezione delle infrastrutture di rete. Al fine di garantire un maggiore livello di protezione dei dati biometrici, sono state sviluppate apposite tecniche di cifratura dei template biometrici. Queste tecniche permettono di effettuare il confronto tra template nel dominio cifrato, senza la necessità di decrittare i dati. Tra le più famose tecniche di cifratura di dati biometrici presenti in letteratura, si possono distinguere: metodi di biohashing, trasformazioni non reversibili, tecniche di cifratura omomorfiche, metodi basati su tecniche fuzzy e metodi basati su tecniche di intelligenza computazionale.

Attualmente, al fine di garantire la protezione della privacy degli utenti, in tutto il mondo sono state emanate leggi per la regolamentazione dell'utilizzo dei sistemi biometrici e per la gestione dei dati sensibili. In Italia, ad esempio, l'utilizzo dei dati biometrici è regolato da decreti del Garante della Privacy.

1.7 Sistemi biometrici di riconoscimenti della voce

Con sistema di riconoscimento della voce o del parlatore, dall'inglese speaker recognition, si intende il processo di validazione dell'identità che un utente dichiara, utilizzando le caratteristiche estratte dall'analisi della sua voce.

C'è una differenza fra riconoscimento del parlatore (riconoscere chi sta parlando) e riconoscimento vocale (riconoscere cosa viene detto). Questi due termini sono confusi frequentemente. C'è anche una differenza fra l'atto di autenticare un utente (a cui ci si riferisce spesso col termine autenticazione del parlatore, verifica del parlatore o, più spesso, con i termini inglesi speaker verification e speaker authentication) e quello di identificare l'utente (a cui ci si riferisce solitamente col termine identificazione del parlatore o con l'inglese speaker identification). C'è spesso confusione anche con il processo di speaker diarisation (riconoscimento di quando interviene il medesimo parlatore).

Il riconoscimento del parlatore ha una storia lunga quattro decadi e utilizza le caratteristiche acustiche del parlato che si è scoperto caratterizzare al meglio i diversi individui (cioè che differiscono maggiormente al variare dell'individuo). Queste caratteristiche riflettono sia quelle dell'anatomia (come la dimensione e la forma del collo e della bocca) che quelle comportamentali (come l'altezza della voce o la cadenza del parlato). La speaker verification ha guadagnato il titolo di misurazione biometrica al riconoscimento del parlatore.

Ci sono due principali applicazione delle tecnologie e delle tecniche di riconoscimento del parlatore. Se un parlatore afferma di possedere una certa identità e la voce è utilizzata per validare questa affermazione, il processo è detto di verifica o di autenticazione. Viceversa l'identificazione è il processo di determinare l'identità di un parlatore sconosciuto. In altre parole la verifica del parlatore è un confronto 1:1, dove la voce di un parlatore è confrontata con un'unica impronta vocale (o "modello del parlatore"), mentre l'identificazione è un confronto 1:N dove la voce è confrontata con N modelli distinti.

La verifica del parlatore può essere impiegata per l'accesso a sistemi sicuri in aggiunta ad altre tecniche di accesso. Questi sistemi generalmente operano con la consapevolezza dell'utente e richiedono la loro cooperazione. I sistemi di identificazione del parlatore sono realizzati solitamente senza prevedere la cooperazione del parlatore.

Ogni sistema di riconoscimento del parlatore ha due fasi: una fase di raccolta dati (enrollment) e una fase di verifica. Durante la fase di raccolta dati la voce del parlatore viene registrata e da essa vengono estratte un certo numero di caratteristiche per formare un impronta vocale, o modello. Nella fase di verifica un

campione vocale è confrontato con l'impronta vocale precedentemente creata. Per i sistemi di identificazione, i campioni vengono confrontati con varie impronte vocali, per trovare i risultati più simili, mentre nei sistemi di verifica i campioni sono confrontati con una sola impronta vocale. Per questo motivo la verifica è solitamente più veloce dell'identificazione.

I sistemi di riconoscimento del parlatore si suddividono in due categorie: dipendenti dal messaggio (o text-dependent) e indipendenti dal messaggio (text-independent), a seconda che il messaggio pronunciato durante la fase di raccolta dati debba coincidere o meno con quello pronunciato durante la fase di verifica.

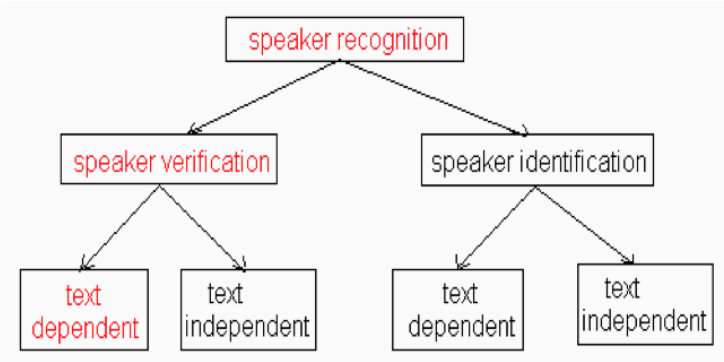

Figura 1.6 Albero sistema di riconoscimento del parlatore.

Nei sistemi dipendenti dal messaggio il messaggio può essere comune a tutti i parlatori (per esempio una parola d'ordine comune) o univoco. In aggiunta è possibile usare delle informazioni segrete condivise (o shared-secrets, come ad esempio parole d'ordine o PIN) o delle informazioni basate sulla conoscenza, al fine di creare scenari di autenticazione a più fattori.

Nei sistemi di identificazione si utilizzano più spesso sistemi indipendenti dal messaggio, poiché non richiedono la collaborazione del parlatore. In questo caso il messaggio pronunciato nella fase di identificazione è diverso da quello utilizzato in fase di raccolta ed entrambe le fasi possono avvenire senza la consapevolezza del parlatore, come nel caso di alcune applicazioni forensi.

Poiché le tecnologie indipendenti dal messaggio non possono confrontare direttamente quello che viene detto nelle due fasi di raccolta e verifica, le applicazioni di verifica che ne fanno uso spesso impiegano anche sistemi di riconoscimento vocale per determinare cosa viene detto in fase di autenticazione.

2 Verifica di Identità basato sul Parlato

Un sistema per la verifica di identità basato sul parlato si riferisce ad un sistema informatico il cui compito è quello di verificare l'identità dichiarata da una persona, il riconoscimento si effettua esclusivamente tramite un codice vocale condiviso tra l'utente e il sistema.

Questo sistema automatico prevede due modalità di funzionamento:

- La registrazione in cui un utente viene registrato nel sistema, in questa fase, l'utente deve inoltre registrare un codice vocale.
- L'autenticazione in cui si deve verificare l'autenticità della voce, basandosi sul codice vocale registrato al momento, il sistema deve decidere se il timbro della voce è uguale a quello percepito durante la fase di registrazione.

Il nostro obiettivo non è assolutamente quello di verificare se i codici registrati nella fase di registrazione e di autenticazione sono identici, ma quello di decidere se sì o no è lo stesso individuo ad aver pronunciato entrambi i codici vocali.

In altre parole, si tratta di un sistema biometrico per la verifica del parlatore dipendente dal messaggio proprio perché come spiegato prima l'esito della verifica, è condizionata dalla parola "chiave" registrata durante la fase di addestramento o registrazione.

La realizzazione di un tale sistema di riconoscimento si basa sulla possibilità di identificare un individuo (riconoscere chi sta parlando) attraverso l'analisi di un suo campione di voce. La voce è il mezzo di comunicazione il più utilizzato dagli esseri umani, però oltre a trasportare il contenuto informativo che si vuole divulgare, è carico di informazioni non soltanto sul nostro stato d'animo (siamo contenti, tristi,…), ma contiene anche informazioni sull'identità della persona che lo ha emesso come la sua età , il suo sesso, la sua origine... Infatti, al telefono per esempio, è lo strumento che un essere umano usa per riconoscere i propri cari.

Il riconoscimento del parlatore ha una storia lunga quattro decadi e utilizza le caratteristiche acustiche del parlato che si è scoperto caratterizzare al meglio i diversi individui (cioè che differiscono maggiormente al variare dell'individuo). Queste caratteristiche riflettono sia quelle dell'anatomia (come la dimensione e la forma del

collo e della bocca) che quelle comportamentali (come l'altezza della voce o la cadenza del parlato).

2.1 La voce come variabile biometrica

Il segnale vocale non è soltanto qualcosa che è presente in ogni individuo, ma è anche qualcosa di unico. L'unicità della voce è determinata dagli elementi che possono essere del tipo fisiologico (come la dimensione del tratto vocale), oppure del comportamentali (come l'intonazione).

La voce è il suono emesso dall'essere umano parlando o cantando oppure urlando. Un problema molto interessante è quello di individuare quali sono le caratteristiche che rendono un segnale vocale unico tra altri segnali e che cosa rende un segnale vocale differente da un altro. Questi includono informazioni specifiche sul tratto vocale, sulla sorgente di eccitazione (corde vocali) e tratti comportamentali del parlatore.

Il segnale vocale nasce dalle vibrazioni delle corde vocali, il segnale deve attraversare tutto il tratto vocale prima di essere emessa nell'aria (vede figura 2.1).

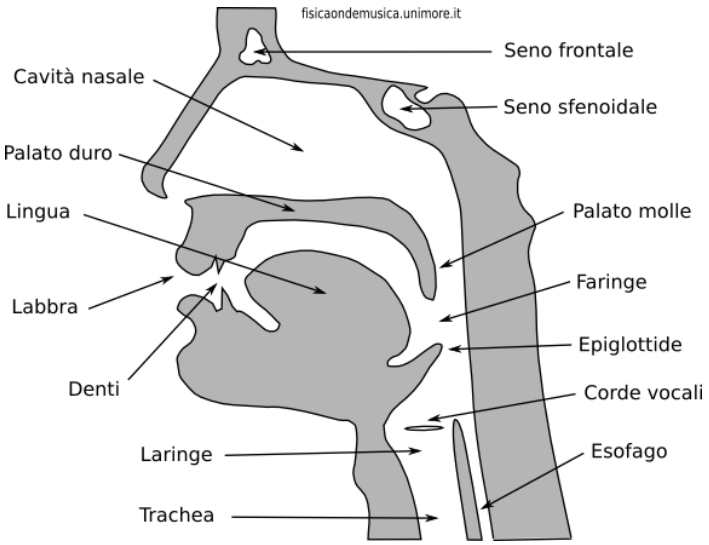

Figura 2.1 Schema del tratto vocale umano

A causa della pressione della glottide e dell'aria spinta dai polmoni, le corde vocali possono aprirsi e chiudersi molto rapidamente, il che genera vibrazioni nell'aria. La vibrazione è modulata dalle risonanze della cavità faringea e/o nasale e/o orale, formando timbro diverso delle nostre voci.

La struttura fisica e la dimensione del tratto vocale, nonché della sorgente di eccitazione, sono unici per ogni individuo. Quest'unicità è incorporata nel segnale vocale durante la produzione vocale e può essere utilizzato per il riconoscimento del parlatore.

Inoltre, le caratteristiche comportamentali come il modo in cui viene controllato il tratto vocale e le corde vocali durante la produzione della voce sono unici per ogni individuo. Le informazioni sul tratto comportamentale sono anche incorporati nel segnale vocale e possono essere utilizzati per il riconoscimento parlatore.

In altre parole:

La frequenza di vibrazione delle corde vocali determina l'altezza della voce.
Le posizioni e le forme delle labbra, lingua e naso determinano il timbro.
La compressione dei polmoni determinare il volume della voce.

Uno dei parametri più importanti di un segnale audio è la sua frequenza, infatti, i segnali audio sono discriminati l'uno dall'altro tramite il valore delle loro frequenze. Quando la frequenza di un suono aumenta, il suono è acuto e irritante. Quando la frequenza di un suono diminuisce, il suono si approfondisce. La frequenza di un segnale vocale corrisponde alla frequenza di vibrazione delle corde vocali. Il più alto valore della frequenza che un essere umano può produrre è di circa 10kHz, e il valore più basso è circa 70 Hz. Questo intervallo di frequenza cambia di persona a persona.

2.2 Struttura di un Sistema Biometrico per la Verifica di Identità

In generale, un sistema biometrico è costituito essenzialmente da cinque moduli principali:

- Il microfono per rilevare il segnale vocale;
- Un modulo di estrazione delle caratteristiche, che si occupa di elaborare il segnale acquisito dal sensore, ed estrarne le caratteristiche pertinenti;
- Un modulo di confronto delle identità
- Un decisore
- Un database contenente i modelli, chiamati più comunemente templates, dei dati su cui eseguire il confronto.

Questa struttura non prevede che tutti moduli si trovino essenzialmente su un unico dispositivo, infatti, tramite il telefono si potrebbe acquisire il segnale vocale, estrare le caratteristiche distinguibili del locutore e inviarle in modo remoto a un server per la memorizzazione oppure per l'elaborazione, dopodiché il server invierà il risultato dell'operazione al cliente.

I sistemi di verifiche prevedono due fasi essenziali:

1. La registrazione: memorizzazione dei tratti pertinenti del locutore
2. La verificazione: controllo dei diritti di accesso accordati alla persona autenticata.

La fase di registrazione prevede in primo luogo l'identificazione formale della persona interessata, seguita dalla registrazione biometrica.

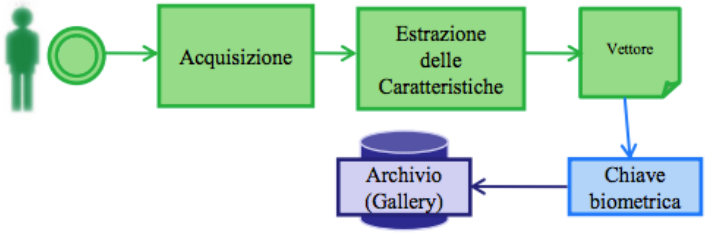

Figura 2.2: Fase di registrazione.

Come evidenziato nella figura 2.2, durante fase di registrazione, tramite il microfono, si acquisisce il segnale audio relativo al codice vocale dell'individuo, la qualità dei dati acquisiti è cruciale per le successive verificazione e dipende da molti fattori in particolare il rumore dell'ambiente di cui la persona si trova e lo stato d'animo dell'utente; questa fase necessita la collaborazione dell'utente infatti l'utente non deve cercare di ingannare il sistema riproducendo un codice vocale diverso da quello richiesto oppure modificare il suo comportamento cambiando il tono di voce oppure l'intonazione. Il segnale acustico acquisito dal microfono è poi processato dal modulo di estrazione delle caratteristiche per la creazione di un vettore contenente le caratteristiche pertinenti che costituiscono la chiave biometrica del locutore. Tale chiave viene memorizzata nell'archivio del sistema. Questa fase è eseguita una volta dall'utente ma può essere ripetuta con periodo di tempo relativamente lunghi.

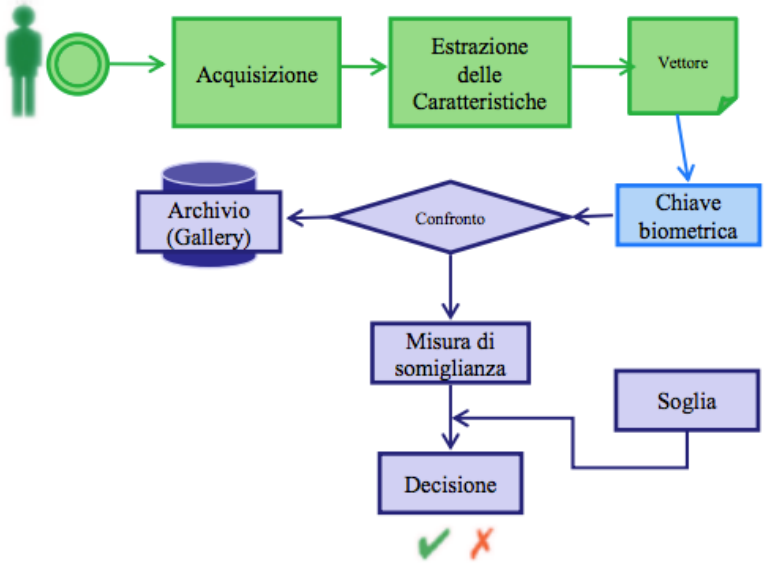

Figura 2.3: Fase di verificazione

La figura 2.3 rappresenta l'algoritmo che viene eseguito durante la fase di verificazione, questa fase si ripete ogniqualvolta un individuo richiede un certo diritto al sistema e quindi l'utente dichiara la sua identità individuando il suo template di riferimento presente in archivio, come nella fase di registrazione si registra il segnale vocale del parlato da cui si estrae una sequenza delle

caratteristiche, dopodiché questa sequenza è confrontata con quella di riferimento. Dato che le due sequenze sono state estratte dallo stesso parlato detto con istanti e toni diversi, durante il confronto si cerca di trovare una misura di similarità (non una distanza) che ci descrive quanto le nostre due sequenze sono simili tra loro; in generale, due sequenze sono simili se la misura di somiglianza tra loro definita in maniera opportuna, è piccola. In base a la soglia prefissata dal sistema, il sistema decide di accettare l'utente concedendoli i diritti richieste se la misura di somiglianza trovata è minore della soglia; nel caso contrario i diritti richiesti sono negati presumendo che si tratta di un impostore.

In linea di massima il processo di riconoscimento automatico del parlatore può essere implementato tramite i cinque sotto-processi seguenti: acquisizione, estrazione delle caratteristiche, modello, confronto tra modelli e decisione.

Di seguito verrà fornito una panoramica delle principali tecniche sviluppate in ciascuna di queste sotto-processi. Tale riesame aiuta a comprendere le scelte disponibili delle tecniche nella letteratura.

2.2.1 Acquisizione del segnale vocale

I segnali audio rappresentano la pressione dell'aria in funzione del tempo, è una continua nel tempo e in ampiezza del segnale.

Inizialmente, l'onda acustica emesso nell'aria è trasformata dal microfono in un segnale numerico di voce adatto a essere elaborare. Questo segnale analogico è condizionato dal filtro di antialiasing (e possibilmente un filtro supplementare per compensare qualsiasi danni del canale) Il filtro antialiasing limita la larghezza di banda del segnale approssimativamente alla frequenza di Nyquist (metà della frequenza di campionamento) prima il campionamento.

Il segnale analogico è campionato per formare un segnale digitale da un convertitore analogico-digitale (A/D); nel corso di quest'operazione, ci sono due parametri da impostare:

- La frequenza di campionamento: è il numero di punti di campionamento per secondo, in unità di Hertz (abbreviato Hz). Una frequenza di campionamento superiore indica migliore qualità del suono, ma lo spazio di archiviazione è anche più grande. Le frequenze di campionamenti comunemente utilizzati sono elencate accanto:
  - 8 kHz: La qualità della voce per i telefoni e giocattoli.
  - 16 KHz: utilizzato solitamente per il riconoscimento vocale.
  - 44.1 KHz: qualità CD.
- Risoluzione di bit: è il numero di bit utilizzati per rappresentare ciascun campione del segnale audio. I numeri comunemente usati sono:
  - 8-bit: l'intervallo corrispondente è 0 ~ 255 o -128 ~ 127.
  - 16 bit: Il campo corrispondente è -32768 ~ 32767.

In altre parole, ciascun campione del segnale audio acquisito è rappresentato da un numero intero da 8 bits o 16 bits.

Nelle applicazioni locali di verifiche del locutore, il canale analogico è semplicemente il microfono e il suo cavo; quindi, il segnale numerico risultante può essere di molto alta qualità che quelli segnali analogici trasmessi sopra via telefono.

Generalmente, osservando il segnale audio acquisito, si vede che è composto in due parti, una regione di bassa ampiezza che può corrispondere al silenzio oppure al rumore ambientale, la parte di alta intensità può a sua volta corrisponde a un segnale non vocalizzato come un fischio, battimenti delle mani, … oppure a un segnale vocale vocalizzato.

I sistemi di riconoscimento del locutore producono buone prestazioni quando il segnale audio è pulito. In pratica, come detto in precedenza, il segnale acquisito è rumoroso non solo a causa dell'ambiente in cui ci si trova ma anche a causa del microfono, e quindi degrada le prestazioni. Migliorare le prestazioni in queste condizioni è una questione importante. La soluzione proposta è individuare le regioni di disturbi ed elaborare solo sui segnali vocali vocalizzati.

Possiamo riassumere tutto questo discorso dicendo che il segnale registrato dal microfono è una somma due segnale: un segnale vocalizzato carica d'informazioni e un segnale di disturbo (rumore ambientale, segnale non vocalizzato) privo d'informazioni. Nei sistemi di riconoscimento del parlatore, è opportuno individuare le regioni di disturbi ed elaborare solo sui segnali vocali vocalizzati.

2.2.2 Estrazione delle caratteristiche

Il processo di estrazione delle caratteristiche si occupa di elaborare il segnale acquisito dal sensore, ed estrarne le caratteristiche pertinenti in forma di vettori delle caratteristiche (features vectors); il feature vector rappresenta informazioni specifiche sul locutore su uno o più tratti seguenti: il tratto vocale, la sorgente di eccitazione e i tratti comportamentali.

La forma del tratto vocale è un fattore di distinzione fisico importante della voce. Mentre l'onda acustica attraversa il tratto vocale, il suo contenuto di frequenza (spettro) è alterato dalle risonanze del tratto vocale. Le risonanze del tratto vocale sono chiamate formanti; la forma del tratto vocale può essere stimata dalla forma spettrale (per esempio, energie, i formanti, lo spettro e i coefficienti MFCCs) del segnale di voce. I sistemi biometrici di riconoscimento della voce usano tipicamente le caratteristiche derivate soltanto dal tratto vocale.

Il meccanismo vocale umano è anche determinato da una sorgente di eccitazione, che egualmente contiene informazioni specifiche sul locutore. L'eccitazione è generata flusso d'aria dai polmoni, portati dalla trachea tramite le corde vocali. Ci sono diversi tipi di eccitazione; ognuno di loro danno luogo a un suono differente: un fonema, un sussurro, un fischio, una vibrazione o una qualche combinazione di questi.

L'eccitazione di fonema accade quando il flusso dell'aria è modulato dalle corde vocali. La frequenza di oscillazione delle corde vocali è chiamata la frequenza

fondamentale e dipende dalla lunghezza, dalla tensione e dalla massa delle corde vocali; quindi frequenza fondamentale è un'altra caratteristica di distinzione del tipo fisico. Le altre tipe di eccitazione provocano un flusso d'aria turbolento che ha una caratteristica a larga banda di disturbo. La voce prodotta dall'eccitazione fonato è detta vocalizzata, e la voce prodotta da altri tipi di eccitazione è chiamata voce non-vocalizzata.

Le caratteristiche della sorgente di eccitazione possono essere stimate calcolando la frequenza fondamentale, la variazione di pitch, e i coefficienti di predizione lineare della sorgente (LPCCs).

Altre proprietà fisiologiche comprendono la capacità vitale (il volume massimo di aria che un individuo può espirare in seguito ad un'inspirazione massima), il tempo massimo fonatorio (la durata massima che può essere sostenuta una sillaba), il quoziente fonatorio (rapporto tra capacità vitale a tempo massimo fonatorio) e la pressione aerea sottoglottica (quantità di aria che passa attraverso le corde vocali).

Altri aspetti di produzione della voce che potrebbero essere utili per discriminare fra gli individui sono caratteristiche comportamentali, compreso la frequenza delle parole, il dialetto, l'intonazione.

Tra queste caratteristiche pertinenti per l'identificazione del locutore, quelle maggiormente utilizzate sono le caratteristiche spettrali, in particolare, MFCCs e LPCCs. Le ragioni principali sono dovute alla poca variabilità intra-locutore e anche alla disponibilità di ricchi strumenti di analisi spettrale. Tuttavia, informazioni specifiche sull'oratore basato sulla sorgente di eccitazione e sul tratto comportamentale rappresentano aspetti diversi d'informazioni; quindi la fase di estrazione potrà essere migliorata utilizzando delle tecniche di estrazione delle caratteristiche per la sorgente di eccitazione e per i tratti comportamentali, tuttavia, il principale limite per entrambi le tecniche è la non disponibilità di strumenti adeguati per l'estrazione di queste caratteristiche.

2.2.3 Il modello

L'obiettivo della tecnica di modellazione è di generare modelli per il parlatore a partire dei vettori delle caratteristiche specifici del locutore ottenuto nella fase di estrazione delle caratteristiche.

I modelli consentono di ridurre la dimensione dei dati estratti senza però rinunciare alla qualità sfruttando i principi di funzionamento delle tecniche di modellazione. Lo stato dell'arte dei sistemi di riconoscimento del locutore utilizza diverse tecniche di modellazione, che sono brevemente descritte in questa sezione. La maggior parte di queste tecniche può essere sostanzialmente raggruppata in tipi generativi e discriminativi. Studi precedenti in materia di riconoscimento del parlatore usano modello diretto ovvero il confronto diretto dei dati estratti nella fase di registrazione con quelli ottenuti nella fase di autenticazione.

Nel modello diretto, vettori delle caratteristiche sono direttamente confrontati tramite una misura di similarità. Per misurare di similarità, viene utilizzato una delle tecniche come la distanza euclidea o spettrale o distanza di Mahalanobis.

Furui ha introdotto il concetto di dynamic time warping DTW (deformazione dinamico temporale) per i sistemi di verifica di identità basati sul parlato come nel nostro progetto. Tuttavia, questo algoritmo è stato originariamente sviluppato per il riconoscimento vocale.

L'approccio dynamic time warping è soprattutto utilizzato nei sistemi di riconoscimento dipendenti dal parlato e cerca principalmente di trovare un allineamento ottimo tra le due sequenze di vettori delle caratteristiche ottenute nelle fasi di registrazione e di verificazione attraverso una distorsione non lineare rispetto alla variabile indipendente tempo.

Lo svantaggio del modello diretto è il tempo di calcolo, infatti, risulta molto elevato quando il numero di vettori cresce. Per questo motivo, è comune per ridurre il numero di vettori delle caratteristiche mediante una tecnica di modellazione come nella tecnica di quantizzazione vettoriale (VQ).

La quantizzazione vettoriale (VQ) è una tecnica di quantificazione che viene spesso utilizzata nella compressione dei dati per produce una rappresentazione più compatta dei dati. Si tratta evidentemente di una tecnica di compressione che non conserva tutto il contenuto informativo (Lossy Data Compression). L'idea che sta alla base di questa tecnica è di codificare i valori di un spazio vettoriale multidimensionale con i valori di un sotto-spazio discreto di dimensione inferiore. Ovviamente, il vettore dello spazio piccolo necessita meno spazio per la memorizzazione. Il cambiamento di spazio si realizza tramite l'uso di un dizionario (codebook); Il più noto algoritmo di generazione di codebook nei sistemi di verifica automatica di identità è il LBG algoritmo.

Per modellare le variazioni statistiche, si utilizza il modello nascosto di Markov (HMM). Gli HMM sono utilizzabili sia per il riconoscimento della voce dipendente dal parlato che per il riconoscimento del locutore indipendentemente dal parlato. Sono modelli matematici che descrivono le probabilità di trovare una data sequenza in un database conoscendo il contenuto del database.

In HMM, i parametri temporali ottenuti (features vectors) sono considerati come simboli di osservazione, HMM hanno la caratteristica di poter determinare i parametri ignoti (hidden) in base ai parametri osservabili. Viene generato misure di probabilità continue utilizzando i modelli di misture gaussiani (GMM). Il presupposto principale di HMM è che lo stato futuro dipende dallo stato attuale e non dagli stati precedenti. In HMM, anche la funzione di generazione delle features in corrispondenza alle transizioni è un processo stocastico: a ogni stato è associata una certa densità di probabilità e la feature emessa in corrispondenza di una certa transizione viene generata casualmente in accordo con la relativa densità di probabilità.

2.2.4 Autenticazione, Confronto e Decisione

Il processo di autenticazione di un sistemi di verifica di identità è composto di due fase:

Durante la prima fase, l'individuo dichiara la sua identità e il sistema va a recuperare il suo modello di riferimento nel database.

La seconda fase consiste ad acquisire il codice vocale del soggetto e a generare un suo modello di voce come nella fase di registrazione.

L'ultimo passo consiste nel prendere una decisione logica: accettare l'individuo affermando quindi l'identità del soggetto è vera; rigettare l'individuo quindi il sistema attesta che il soggetto è un impostore. Questa decisione viene presa confrontando il modello generato con quello di riferimento, il risultato di questo confronto è in generale un punteggio che rappresenta quanto vicino sono i due modelli.

La decisione dipende di un valore di soglia prefissato dal sistema di riconoscimento biometrico, infatti, se il punteggio rilevato è minore o uguale a tale soglia, si risponde affermativamente. La risposta è negativa nel caso in cui il punteggio rilevato è superiore alla soglia prefissata.

Le prestazioni di un sistema verifica di identità sono principalmente dovuti al fatto si prende una decisione in base ad una soglia infatti, può succedere che il sistema possa scambiare una persona per un'altra (False Acceptance) oppure che il sistema non riconosca una persona (False Reject).

E di conseguenza abbiamo due tipologie di errore:

- FRR (False Rejection Rate) è la percentuale di falsi rifiuti, cioè utenti autorizzati ma respinti per errore, in pratica il sistema non riesce a riconoscere le persone autorizzate.
- FAR (False Acceptance Rate) è la percentuale di false accettazioni, per utenti non autorizzati ma accettati per errore, il sistema quindi accetta le persone che non sono autorizzate.

FAR e FRR sono due grandezze strettamente correlate dalla seguente proprietà: al diminuire dell'una cresce l'altra. Ogni Sistema biometrico offre la possibilità di regolare il rapporto FRR/FAR e quindi di aumentare o diminuire la sensibilità del sistema

3 Implementazione

Un sistema di verifica di identità facendo uso dei metodi e tecniche presentati nel capitolo precedenti può essere tranquilamente realizzato mediante un software su un computer personale. Infatti, si è scelto di implementare questo progetto tramite MATLAB versione R2011b.

MATLAB (Matrix Laboratory) è un ambiente per il calcolo numerico e l'analisi statistica che comprende anche l'omonimo linguaggio di programmazione creato dalla MathWorks. MATLAB consente di manipolare matrici, visualizzare funzioni e dati, implementare algoritmi, creare interfacce utente, e interfacciarsi con altri programmi. MATLAB è usato da milioni di persone nell'industria e nelle università per via dei suoi numerosi tools a supporto dei più disparati campi di studio applicati e funziona su diversi sistemi operativi, tra cui Windows, GNU/Linux, Unix e Mac OS, infatti è proprio su questo ultimo che abbiamo scelto di realizzare il detto sistema; versione Mac OS X Lion 10.7.3, processore 2,4 intel core 2 Duo.

Il principale vantaggio dei sistemi biometrici di riconoscimento del locutore è la versatilità: l'autenticazione della voce è una tecnologia semplice da usare e non intrusiva, quindi, di norma, facilmente accettata dagli utenti. Rispetto alle altre tecnologie biometriche, è molto accurata e non richiede l'uso di apparecchiature specifiche. Ogni essere umano ha una sua impronta vocale univoca che si può facilmente acquisire tramite il microfono di un elaboratore (computer, smartphone, tablet…) e processare il segnale acustico in modo locale oppure in modo remote nel caso in cui ci si tratta di un sistema decentralizzato.

Come abbiamo visto nel capitolo precedente, un sistema di verifica di identità è costituito di cinque moduli ordinati cronologicamente; nella figura 3.1 viene illustrata i processi che andremmo a realizzare, i loro ingressi, le loro uscite.

Il processo "voiceAcquisition" implementato tramite il file voiceAcquisition.m acquisisce il segnale vocale dal microfono e chiama il metodo "voice detection" (voiceDetection.m) per elaborare il segnale ottenuto e restituisce il segnale corrispondente all'istante di inizio e fine parlato. Questo segnale viene a sua volta trasmesso al processo mfccs extration di cui compito, è di analizzare il segnale e restituire i vettori di coefficienti MFCCs.

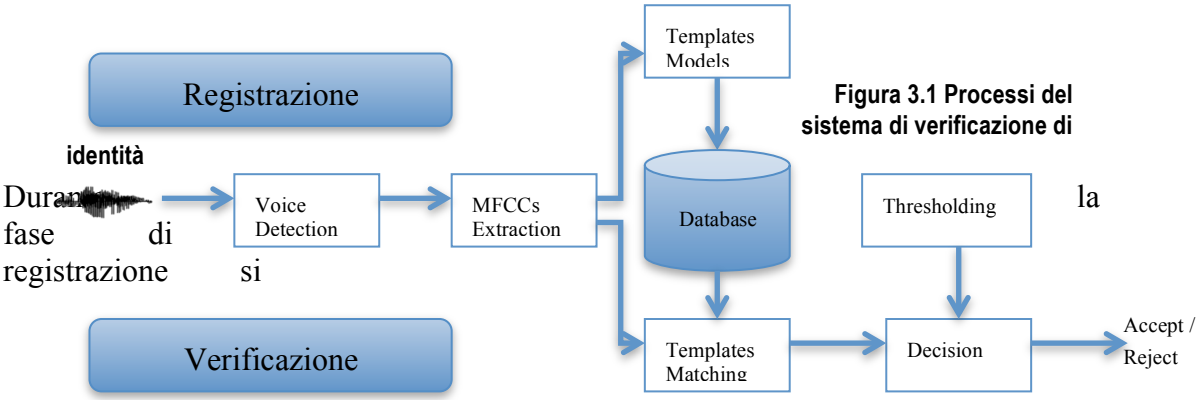

Figura 3.1 Processi del sistema di verificazione di

chiama il processo "TemplatesModels" di cui scopo, è di creare il file "username.bvv" corrispondente al template dell'utente username. I dati presenti verrano utilizzati dal processo "TemplateMatching" durante la verificazione per confrontare due vettori di coefficienti MFCCs tramite l'algoritmo Dynamic Time Warping (DTW); questo processo restituisce la misura di somiglianza tra i due vettori al processo "decision" che decide di accettare o di rifiutare l'utente in base alla soglia definita dal processo "Thresholding".

3.1 Acquisizione e Pre-elaborazione del segnale accoustico

L'estensione del file più comune per i file audio è "wav". MATLAB in grado di leggere tali file wave tramite il comando "wavread". È inoltre possibile utilizzare il comando MATLAB "wavrecord" per leggere i segnali audio dal microfono direttamente. Il formato del comando è:

```
y = wavrecord (n, fs);
```

dove "n" è il numero totale di campioni da registrare, e "fs" è la frequenza di campionamento. MATLAB offre anche un altro comando "Audiorecorder" per offrire un controllo di precisione sulla registrazione. Un esempio di codice di registrazione del segnale è il seguente:

```
    recObj = audiorecorder(fs,nbits,nchannel)
    recordblocking(recObj, duration);
 % Save the recording
 y = getaudiodata(recObj);
```

Nel progetto abbiamo scelto è la frequenza di campionamento "fs" pari a 16 kHz. Questo indica che ci sono 16000 campioni al secondo. La risoluzione in bit "nbits" del segnale audio è di 16 bits. Per quanto riguardo il numero di canale di registrazione "nchannel", uno è sufficiente. Il vettore "y" è un vettore colonna contenente i campioni dei segnali vocali registrati. Possiamo usare il comando "suond (y, fs)" per riprodurre i segnali audio letti.

In MATLAB, tutti i segnali audio vengono normalizzati al numero in virgola mobile nell'intervallo [-1, 1] per una facile manipolazione. Se si desidera ripristinare i valori originali interi, è necessario moltiplicare i valori per $2^{nbits-1}$, dove nbits è la risoluzione in bit.

Il segnale audio registrato da MATLAB oltre a contenere il contenuto informativo necessario per l'estrazione delle caratteristiche pertinenti per identificare il locutore (il codice vocale); è soprattutto carico di pezzi di cui non si può estrarre nessun'informazione caratterizzante il locutore (silenzio e rumore ambientale). Per esempio nella figura 3.2, è rappresentata una registrazione di una durata di tre secondi, dove io pronuncio il mio codice vocale: "Mi chiamo Christian". Infatti, possiamo osservare che i valori dei primi campioni del segnale (dal punto O al punto A) e degli ultimi (dal punto B al punto C) sono quasi nulli corrispondenti quindi al rumore ambientale visto che per il primo pezzo, il soggetto non ha ancora iniziato a dettare il codice vocale e nell'ultimo il soggetto ha realmente finito però sta aspettando la fine della registrazione.

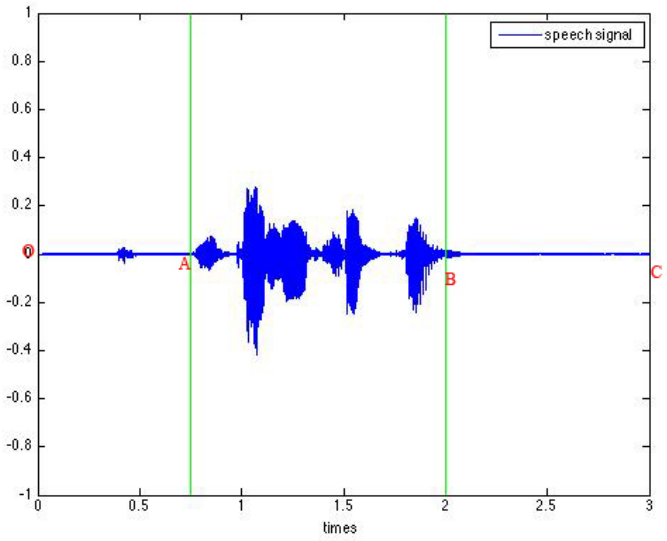

Figura 3.2: Segnale audio.

Il processo di pre-elaborazione "*voiceDetection*" del segnale prende in ingresso il segnale acquisito dal microfono e idealmente, ci deve restituire un segnale di inizio e di fine parlato (segmento A-B). Visto che abbiamo appena detto che i segmenti O-A e B-C sono privi di informazioni sul locutore, allora sarebbe uno spreco temporale andare ad cercare di estrarre una qualche informazioni caratteristiche sul quei segmenti. Però questi segmenti possono risultare molto interessanti per capire quali sono le caratteristiche del segnale di disturbo al momento della registrazione e si può cercare quindi di compensare il suo effetto sul segnale vocalizzato attraverso algoritmi e metodi di compensazione del rumore.

Per individuare automaticamente il momento di inizio parlato (punto A) di fine parlato (punto B) abbiamo sfruttato le proprietà di variazione di energia e di tasso di passaggio al punto zero del segnale vocale.

3.1.1 Energia

L'energia di segnali audio, sono le caratteristiche più importanti secondo la percezione uditiva umana. In generale, ci sono diversi termini intercambiabili che vengono comunemente utilizzati per descrivere l'intensità dei segnali audio, tra cui volume, energia e intensità. Fondamentalmente, l'energia è una caratteristica acustica che è correlato alle ampiezze dei campioni.

Per definire quantitativamente l'energia del segnale audio, è opportuno segmentare il segnale in frame dell'ordine di 10ms, si può impiegare la formula seguente per calcolare l'energia di un dato frame k:

$$E(k) = \sum_{n=1}^{N} |x[n]|^2$$

dove N è il numero totale dei campioni nei frame.

Per rispettare la scala di percezione, si può esprimere l'energia in decibel:

$$E_{db}(k) = 10\log\ E(k)$$

Alcune sue caratteristiche sono riassunte successivamente:

- Il volume è fortemente influenzato dalle impostazioni del microfono (il guadagno del microfono), e la più piccola variazione di distanza tra la sorgente e il microfono è sufficiente per creare perturbazioni di energia. Per eliminare questa variabilità dovuto principalmente dalle condizioni di registrazioni differenti, l'energia può essere normalizzata facendo il rapporto con il suo valore massimo osservato sul segnale globale.
- L'energia dei suoni vocalizzati è solitamente più grande di quella di suoni non vocalizzati, e l'energia dei suoni non vocalizzati è generalmente maggiore di quello del rumore ambientale. In un segnale vocale, l'energia varia molto, questo ci consente non soltanto di distinguere il silenzio ma anche le parole.

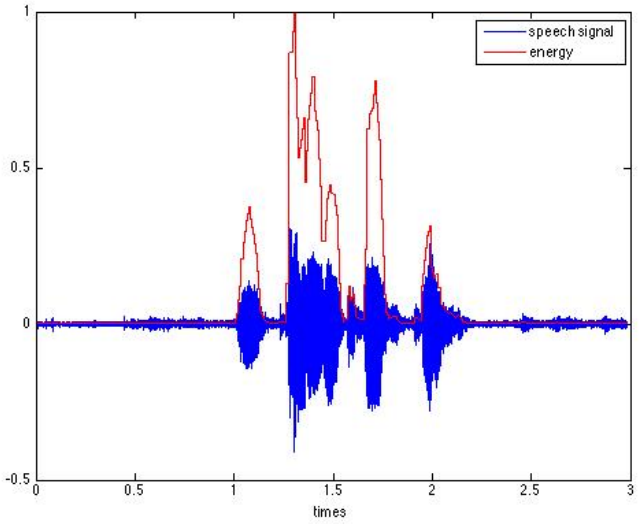

Figura 3.3: Energia del segnale.

3.1.2 Tasso di passaggio a Zero (ZCR)

Dall'inglese Zero-crossing rate (ZCR), il tasso di passaggio a zero è un'altra caratteristica acustica di base che può essere facilmente calcolato. È pari al numero di volte che la forma d'onda cambia di segno (attraversa il punto zero) all'interno di un dato frame k.

$$ZCR(k) = \frac{1}{2N}\sum_{n=1}^{N}|sgn(x(n)) - sgn(x(n-1))|$$

dove N è il numero totale dei campioni nei frame

Il ZCR ha le seguenti caratteristiche:

- il suo valore è molto più basso per i segnali vocalizzati che per i segnali non vocalizzati e i rumori ambientali.
- È difficile distinguere suoni non vocalizzati e i rumori ambientali tramite l'impiego della sola ZCR poiché hanno valori simili di ZCR.

Alcune persone usano la ZCR per la stima approssimativa della frequenza fondamentale, ma è molto inaffidabile.

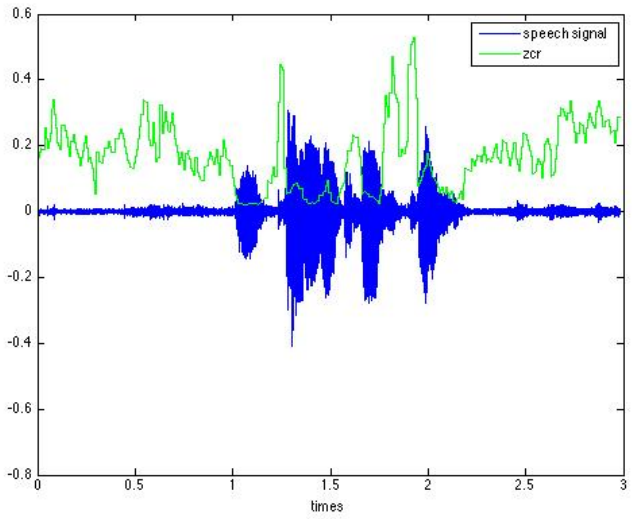

Figura 3.4: ZCR del segnale

Anche in questo caso le variazioni di ZCR è più significativi che il suo valore reali e hanno una caratteristica complementari all'energia, e quindi accoppiando energia e ZCR possiamo discriminare meglio le parte vocalizzate e non vocalizzate del segnale vocale.

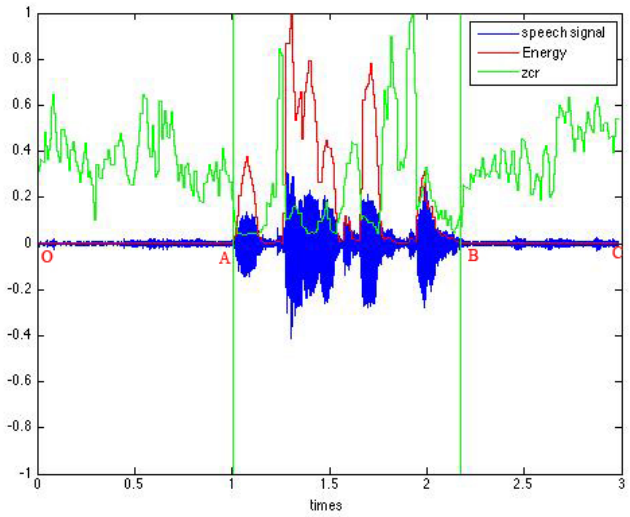

Figura 3.5: Voice Detection

3.2 Mel Frequency Ceptral Coeficient (MFCC)

Il primo passo in qualsiasi sistema automatico di riconoscimento vocale è di estrarre caratteristiche pertinenti cioè identificare i componenti del segnale che sono buoni per identificare il contenuto linguistico e scartando tutte le altre cose che trasporta informazioni come rumore di fondo, emozione, ecc.

Il punto principale per capire la voce è che i suoni generati da un essere umano sono filtrati dalla forma del tratto vocale incluso lingua, denti ecc. Questa forma determina il timbro del suono che esce. Dato che la forma del tratto vocale è unica per ogni individuo, risulta che anche il timbro della voce lo è e quindi se potessimo determinare con precisione, la forma, questo dovrebbe dare una rappresentazione accurata del soggetto che l'ha prodotta. La forma del tratto vocale si manifesta nell'inviluppo dello spettro di potenza a breve tempo, e l'obiettivo di MFCCs è di rappresentare accuratamente questa busta.

La figura 3.6 illustra l'algoritmo per l'estrazione dei MFCCs, esso è composto dei seguenti passi:

1. Frame Blocking
2. Windowing
3. FFT (Fast Fourier Transform)
4. Mel-Frequency Wrapping
5. Cepstrum (Mel Frequency Cepstral Coefficients)

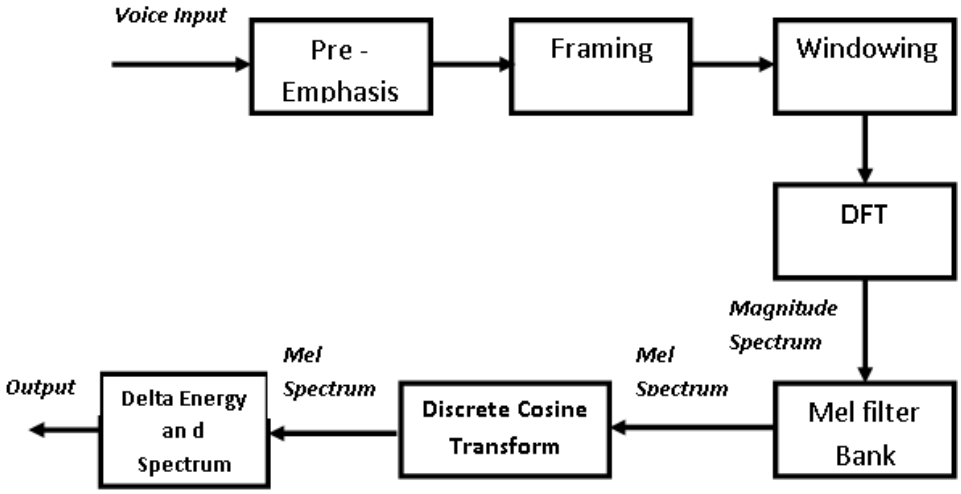

Figura 3.6: Algoritmo Estrazione dei MFCCs

Frame Blocking

Le indagini indicano che le caratteristiche del segnale vocale restano stazionarie(o quasi stazionario) in un intervallo di tempo sufficientemente corto. Per questo motivo, i segnali vocali acquisiti sono elaborati negli intervalli di tempo ridotto. Il segnale è suddiviso generalmente in blocchi chiamati *frame* di dimensioni (windowsize) fra 20 e 30 millisecondi. Ogni frame si sovrappone con il frame precedente di una dimensione (hopsize) predefinita fra 10 e 20 ms. Lo scopo dello schema di sovrapposizione è di lisciare la transizione da frame a frame. A questo punto il MFCC si applica su ogni frame.

Windowing

Il secondo passo serve a eliminare le discontinuità alle estremità del frame. Il singolo frame da processare viene finestrata con le finestra di hamming w(n).

$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi n}{N}\right); 0 < n < N\text{-}1$$

dove N è il numero totale dei campioni in ogni frame

Se x(n) rappresenta il segnale audio contenuto nel frame k, il segnale risultante sarà;

$$y(n) = x(n)w(n)$$

FFT

Il prossimo passo è di prendere la trasformata di fourier di ogni frame. Questa trasformazione è fatta mediante la trasformata veloce discreta di fourier (FDFT) e ci consente di transitare dal dominio temporale al dominio frequenziale.
Ovviamente si deve restituire il logaritmo del modulo della trasformata.

Mel Filter Bank

L'orecchio umano percepisce le frequenze in modo non lineare. Gli studi psicofisici hanno dimostrato che la scala di percezione umana delle frequenze per i suoni è

lineare per le frequenze inferiori a 1kHz e logaritmico per frequenze superiore a 1kHz. Questa scala è chiamata “mel” e di conseguenza possiamo usare la seguente formula approssimativa per convertire in mel una data frequenza f in Hz;

$$Mel(f_{Hz}) = 2595 * \log_{10}\left(1 + \frac{f_{Hz}}{700}\right)$$

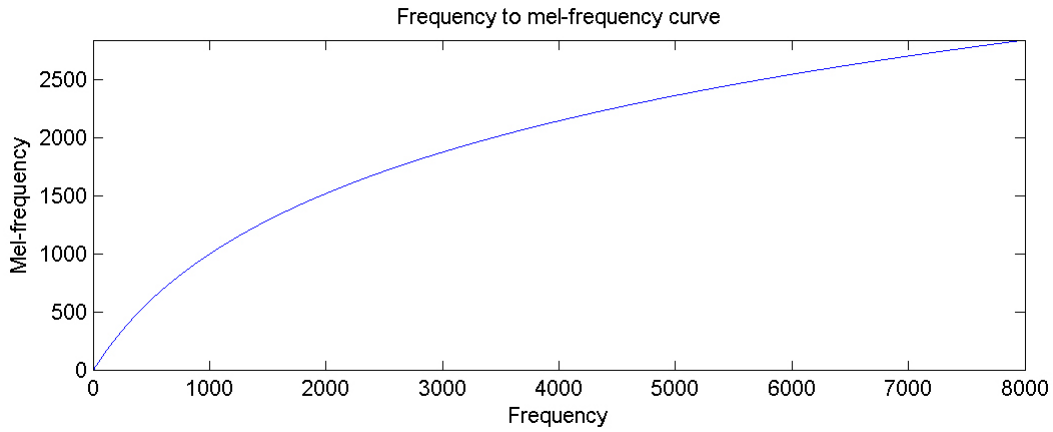

Figura 3.8: Grafico di conversione da frequenza Hz in Mel.

Per simulare la percezione uditiva dell’orecchio umana, si filtra il modulo della risposta frequenziale del segnale acustico con un banco di 20 filtri triangolari. Le posizioni di questi filtri sono equi spaziate lungo la scala Mel come illustrato nella figura 3.9.

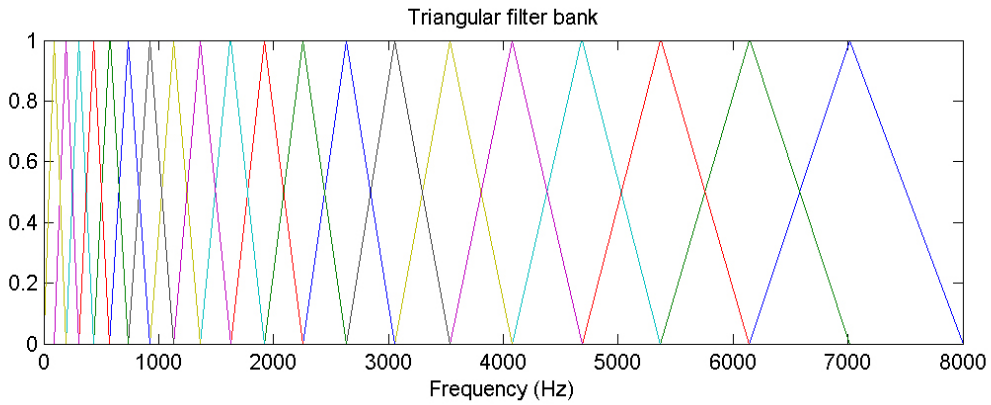

Figura 3.9: Banco di 20 filtri triangolari equi spaziati lungo la scala Mel

Si calcola l’energia a ogni filtro ottenendo alla fine di questa operazione un vettore di 20 elementi.

Una volta che abbiamo le energie dei filtri, prendiamo i loro logaritmi. Questo è anche motivata dalla percezione umana: non si sente il volume sulla scala lineare. Generalmente per raddoppiare il volume di un suono percepito, abbiamo bisogno di aggiungere otto volte l'energia in esso. Ciò significa che le variazioni di energia di grandi dimensioni non può sembrare molto diverso se il suono è forte per

cominciare. Questa operazione di compressione rende le nostre caratteristiche più in linea con ciò che gli esseri umani realmente sentono.

Discret Cosine Trasform

In questo passaggio si applica la DCT sui logaritmi delle energie dei filtri triangolari. La formula per il DCT è la seguente:

$$c_m = \sum_{n=1}^{N} E_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)m\right] \quad k = 1,2,\dots L$$

dove N è il numero di filtri passabanda triangolari, L è il numero di coefficienti da conservare. Di solito N = 20 e L = 12 quindi solo 12 dei 26 coefficienti DCT sono conservati, questo perché i coefficienti più elevati DCT rappresentano veloci cambiamenti nelle energie ai banchi di filtri e si scopre che questi cambiamenti rapidi in realtà degradano le prestazioni del sistema.

La ragione principale di questa operazione è che il nostro banco di filtri triangolari sono tutti sovrapposti, e quindi le loro energie sono abbastanza correlate fra loro. La DCT serve a decorrelare le energie.

Dal momento che abbiamo svolto la FFT, la DCT trasforma il dominio frequenziale in un dominio temporale come chiamato dominio quefrency. Le caratteristiche ottenute sono quindi a cepstrale e sono chiamati in inglese Mel Frequency Ceptrum Coefficient (coeficienti ceptrale a scala mel) o semplicemente MFCC. La MFCC da sola può essere utilizzata come caratteristica di riconoscimento vocale. Per ottenere prestazioni migliori, ci sono un paio di cose più comunemente fatto, a volte l'energia del frame viene aggiunto a ciascun vettore di MFCC. si può anche eseguire operazioni delta, come spiegato nei passaggi successivi.

Delta Ceptrum and Delta-Delta

Noto come coefficienti di velocità e coefficienti di accelerazione. Il vettore di caratteristica MFCC descrive solo l'inviluppo spettrale di potenza di un singolo frame, ma la voce ha anche informazioni nella dinamica e quindi vogliamo determinare quali sono le traiettorie dei coefficienti MFCC nel tempo. Si scopre che il calcolo delle traiettorie MFCC e aggiungendoli al vettore originale aumenta le prestazioni del sistema di biometria.

Per calcolare i coefficienti delta, la formula è il seguente:

$$d(t) = \frac{\sum_{n=1}^{M} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{M} n^2}$$

Il valore di M è di solito 2.

I coefficienti di accelerazione Delta-Delta sono calcolati nello stesso modo, ma vengono calcolati dai coefficienti delta e non i coefficienti MFCC. Se si aggiunge la velocità, la dimensione del vettore di caratteristica è 26. Se si aggiunge sia la velocità e l'accelerazione, la dimensione cresce a 39.

3.3 Dynamic Time Warping (DTW)

Il DTW è un algoritmo per la misurazione di similarità fra due sequenze che possono variare nel tempo o in velocità; è un algoritmo che permette l'allineamento tra due sequenze, e che può portare a una misura di distanza tra le due sequenze allineate. Tale algoritmo è particolarmente utile per trattare sequenze in cui singole componenti hanno caratteristiche che variano nel tempo, e per le quali la semplice espansione o compressione lineare delle due sequenze non porta risultati soddisfacenti.

In generale, DTW è un metodo che permette di trovare una corrispondenza ottima tra due sequenze, attraverso una distorsione non lineare rispetto alla variabile indipendente (tempo). Alcune restrizioni per il calcolo della corrispondenza sono generalmente utilizzate: deve essere garantita la monotonicità nelle corrispondenze, ed il limite massimo di possibili corrispondenze tra elementi contigui della sequenza.

L'algoritmo del DTW si basa sul principio di ricercare similarità tra segnali relativi a due istanze della stessa attività. Esso viene usato per riconoscere pattern in sequenze di dati anche non perfettamente allineate che si riferiscono sostanzialmente alla stessa "forma" di segnale; un tipico esempio è quello di uno stesso movimento o di una stessa parola effettuato/pronunciata a differente velocità.

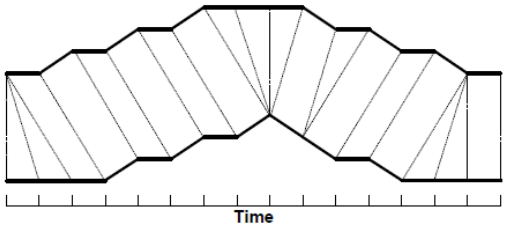

Figura 3.10: Allineamento di vettori tramite DTW.

Nella figura 3.10, ogni riga verticale connette un punto in una serie temporale al suo corrispondente punto simile in un'altra serie temporale.

Il buon fine del riconoscimento è legato ai possibili criteri di "distanza" tra due sequenze e, più in particolare, a tutte le difficoltà derivanti da imperfetti allineamenti temporali tra le due sequenze e/o effetti di distorsione.

L'algoritmo in questione si propone di risolvere il problema considerando le possibili distorsioni nel tempo del segnale d'ingresso e confrontandolo con il segnale di riferimento. Il confronto avviene considerando i due segnali come se fossero due vettori.

Consideriamo i due segnali temporali *Q* e *C,* di lunghezza n e m respectivamente:

$Q = q1, q2,..., qi,...,qn$

$C = c1, c2,..., cj,...,cm$

Il DTW "allinea" le due sequenze mediante una variazione temporale non lineare. Tale variazione temporale avviene secondo quanto illustrato nella Figura 3.10.

Per allineare le due sequenze temporali, DTW costruisce prima una matrice d di dimensione n per m in cui ogni elemento di posizione (i,j) rappresenta la distanza tra

l'elemento i-esimo della sequenza *Q* e l'j-esimo elemento della *C*. La distanza scelta è quella euclidea.

In secondo luogo, DTW costruisce in modo ricorsivo una nuova matrice D sempre di dimensione n per m di cui ogni elemento (i,j) della matrice corrisponde all'allineamento fra i punti *qi* e *cj* definito nel modo seguente:

$$D(i,j) = \min[D(i-1,j-1), D(i-1,j), D(i,j-1)] + d(i,j)$$

Dalla matrice D, si può calcolare il percorso di minimizzazione (Warp-path) muovendosi dal punto (1,1) al punto (n,m) (o viceversa). Il percorso è costituito da un insieme W di k coppie di punti che vengono presi nella matrice D compiendo un percorso secondo le seguenti regole :

• non è possibile un'inversione di direzione nel tempo ;

• l'andamento nel tempo deve essere monotono non decrescente;

• il segnale d'ingresso deve essere continuo nel tempo;

• al fine di ottenere una distanza globale, ogni nuovo elemento del percorso deve essere frutto della somma di distanze locali.

Infine notiamo che per il calcolo di D(1,1) sono necessari valori di D corrispondenti ad indici negativi, e così anche per tutti gli elementi della prima riga e della prima colonna. Per risolvere il problema si pongono precise condizioni al contorno:

• D(1,1) = d(1,1)

• D(1,i) = d(1,i) + D(1,i)

• D(j,1) = d(j,1) + D(j-1,1)

Il percorso sarebbe la diagonale di una matrice di dimensione [n x m] se i due segnali fossero perfettamente allineati. Il termine D(N,M) rappresenta il punteggio associato al percorso minimo tra l'elemento *qn* e l'elemento *cn* e dunque la distanza tra le due forme d'onda.

3.4 Fase di registrazione del sistema

Nella fase di registrazione (enrollment) del sistema così realizzato vedi figura 3.11, viene effettuato due registrazione del codice vocale (per esempio: "*mi chiamo rossi"*) della persona che si vuole registrare nel sistema. Ogni registrazione ha una durata di tre secondi. Poi vengono estratti le caratteristiche MFCC dalle registrazioni 1 e 2 ottenendo quindi due matrici: *mfcc1* e *mfcc2*; tramite l'algoritmo DTW si calcola il valore dtw_ecorrispondente alla misura di similarità tra le due matrici di mfcc. Le tre variabili *mfcc1*, *mfcc2* e dtw_e rappresentano il template del locutore e vengono salvati nel sistema come il file ***nome_utente.bvv***, questo file verrà utilizzato come template di riferimento nelle verifiche successivi per l'estrazione delle variabili *mfcc1*, *mfcc2* e dtw_e. Occorre notare che i file audio dove vengono estratti le caratteristiche non sono salvati nel sistema.

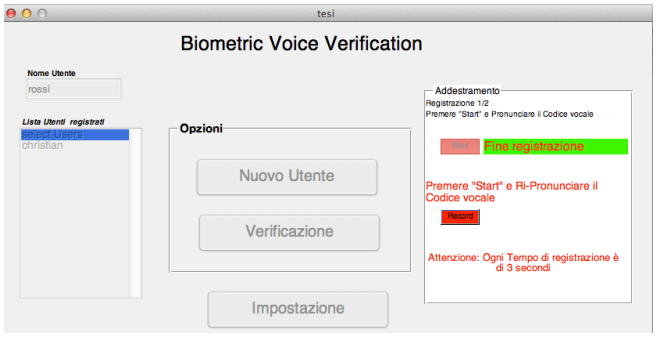

Figura 3.11: Fase di registrazione dell'utente.

3.5 Fase di Verificazione

Nella fase di verificazione (testing), l'utente dichiara la sua identità selezionando il proprio username nella lista degli utenti (figura 3.12), dopodiché viene richiesto al soggetto di pronunciare il suo codice vocale (deve corrispondere a quello registrato nella fase di enrollment). Di seguito si registra il segnale vocale e il modulo di estrazione delle caratteristiche estrae le caratteristiche MFCC, rappresentato tramite la variabile *mfcc*; il prossimo passo consiste a caricare le caratteristiche di riferimento salvati nella fase di registrazione (*mfcc1*, *mfcc2* e dtw_e) e di confrontarli con la nuova appena prelevata: *mfcc*.

Per realizzare il confronto si valutano i valori $dtw1$ e $dtw2$ corrispondenti alle misure di similarità tra *mfcc* e *mfcc1* e tra *mfcc* e *mfcc2* rispettivamente; sia dtw_v il minimo fra $dtw1$ e $dtw2$, la decisione di accettare o di rifiutare il soggetto viene presa in base alla condizione che la misura di somiglianza dtw_v supera o no la soglia prefissata $treshold$.

In generale il valore di $treshold$ è correlata alla sensibilità del sistema di riconoscimento e viene valutata secondo l'algoritmo threshoding.

Figura 3.12: Fase verificazione dell'utente.

L'utente sarà accettato se il valore di dtw_v è minore di $treshold$ nel caso contrario lui verrà negato l'accesso.

3.6 Thresholding

Il decisore prende la decisione in base a una soglia $treshold$. Il sistema offre la possibilità di fare variare la soglia ξ in base al grado di tolleranza t del sistema perciò possiamo dedurre che il valore della soglia è strettamente correlata a quella della tolleranza del sistema.

In realtà, come l'abbiamo implementato noi, la soglia $treshold$ non è statico, se lo fosse, sarebbe difficile trovare un suo valore ideale che non tenga in considerazione la lunghezza del codice vocale perché una misura di somiglianza può risultare bassa semplicemente dovuto al fatto che il codice vocale è costituito di una unica parola. Utilizzeremo il valori dtw_e misura di somiglianza rilevato durante la fase di registrazione e presente nell'archivio del sistema insieme al valore della tolleranza t per determinare il margine di errore ξ il sistema consente di violare durante la fase di verificazione. la soglia $treshold$ è allora pari ad: $treshold = dtw_e + \xi$

La tolleranza del sistema rappresenta la bontà del sistema in termini di sicurezza; un basso grado di tolleranza corrisponde a un alto livello di sicurezza e un alto valore di falsi rifiuti mentre un alto grado di tolleranza implica un basso livello di sicurezza e un alto valore di false accettazioni.

Da questa analisi, si può concludere che i valori ξ e t crescono allo stesso modo: tolleranza t bassa implica errore ξ basso e quindi la probabilità di rifiutare un utente autorizzato è alto; al crescere della tolleranza t crescerà anche l'errore ξ facendo diminuire il tasso di falsi rifiuti e alzare il tasso di false accettazioni.

Nel nostro progetto sopporremo che la tolleranza t rappresenta la percentuale di errore che si deve applicare sul valore dtw_e: $\xi = (dtw_e * t)/100$;

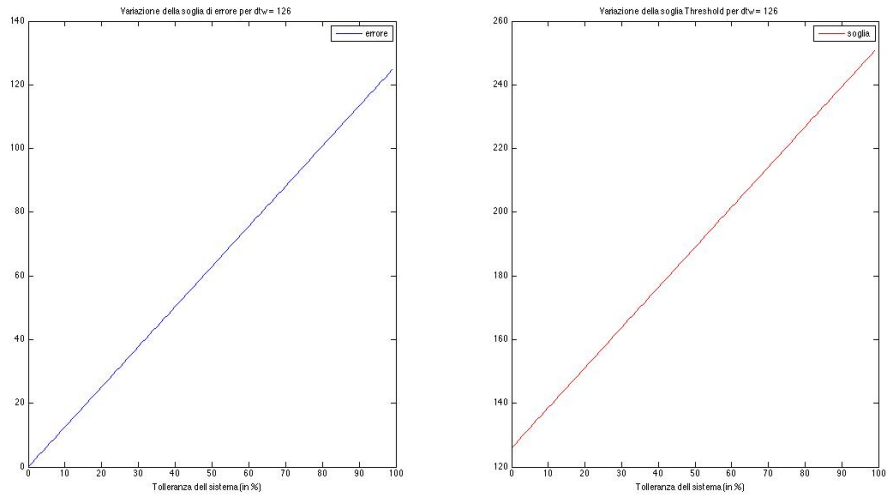

Figura 3.13: Variazione della soglia in funzione della tolleranza percentuale del sistema ($dtw_e = 126$)

Supponiamo la misura di somiglianza $dtw_e = 126$, nel primo grafico della figura 3.13, è rappresentato la variazione di errore consentito quando si fa crescere la tolleranza; e nel secondo grafico, si vede che anche la soglia $treshold$ è monotona crescente.

Il metodo `threshold = threshold(tollerance,dtw_e)` presente nel file matlab threshold.m consente di calcolare la soglia $treshold$ in base alla tolleranza t del sistema e alla misura di somiglianza dtw_e.

3.7 Direttive per l'installazione dell'applicazione

L'installazione dell'applicazione è molto semplice, l'unico requisito per il suo funzionamento è avere sul proprio dispositivo una versione di MATLAB preferibilmente la versione R_2011b.

- Si deve copiare la cartella BVV_Project nella cartella matlab.
- Aprire l'applicazione Matlab.
- Aggiungere nel patch Matlab la cartella BVV_Project insieme a tutte le sue sotto cartelle.
- Mettersi nella cartella BVV_Project
- Lanciare l'applicazione in riga di comando digitando il comando: BVV_Project.
- I template degli utenti sono salvati nella cartella "templates_db" con estensione ".bvv".
- Alcune variabili dell'applicazione come la tolleranza del sistema possono essere modificate tramite il file "setParameter".

4 Sperimentazioni e Analisi dei Risultati

4.1 Esperimenti

Lo scopo degli sperimenti realizzati di seguito è di valutare le prestazioni del sistema realizzato determinando il suo punto di equilibro e il valore di tolleranza ottima; si possono utilizzare questi valori per verifica i miglioramenti che verranno eseguiti sui processi del sistema nel futuro.

Per analizzare il comportamento del sistema di riconoscimento della voce realizzato sono stati effettuati degli esperimenti impostando il sistema su due scenari diversi. Tutti queste prove si sono incentrate in modo tale da valutare le prestazioni in situazioni reali del programma, infatti, nessuno di questi esperimenti è stato realizzato in laboratorio. I sistemi biometrici di riconoscimento del parlatore sono molto sensibili principalmente ai rumori ambientali ad esempio tv accesso, musica in sottofondo, rumori dell'auto, …, infatti, le condizioni ambientali degradano la qualità del segnale vocale perfino a rendere impercettibile la voce, però non possiamo prevedere che l'utilizzazione di questo strumento in condizioni reali di applicazione sia priva di rumore (il segnale vocale sarebbe pulito). E quindi andremmo a eseguire gli esperimenti per valutare le prestazioni del sistema in un ambiente non controllato.

Il primo esperimento che abbiamo eseguito consiste a condividere lo stesso codice vocale "mi chiamo Christian" con tutti gli utenti del sistema, e quindi cerchiamo di calcolare principalmente i valori della funzione $FAR_{user}(t)$, cioè il tasso di false accettazioni relativa all'utente *user* quando la tolleranza del sistema assume il valore t. $FAR_{user}(t)$ è pari al rapporto tra il numero di volte che l'esito della verificazione del campione di voce dell'utente *user* con tutti quelli degli altri utenti è positivo sul numero totale di confronto effettuato.

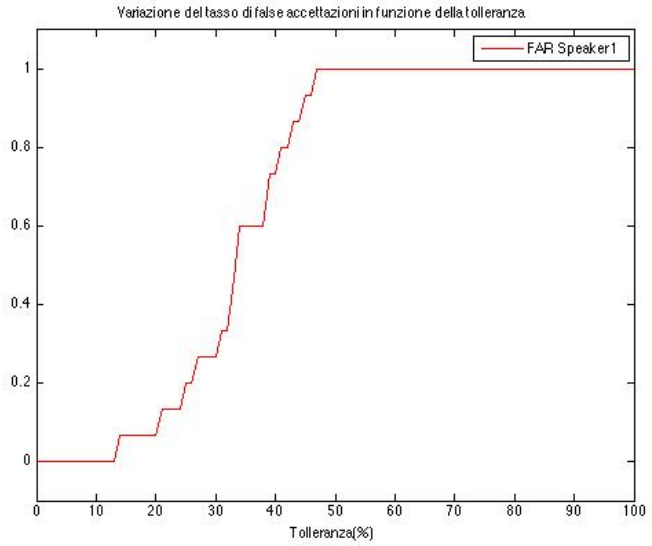

Figura 4.1: Funzione del tasso di false accettazioni che si riferisce all'utente speaker1.

L'obiettivo della seconda sperimentazione realizzata è di determinare i tassi di falsi rifiuti riguardanti gli utenti del sistema. Durante il secondo esperimento, si è lasciata la libertà a ogni utente di scegliere il proprio codice vocale, subito dopo la fase registrazione, l'utente esegue un numero indeterminato di prove di verificazione. Per ogni utenti verrà calcolati i valori della funzione $FRR_{user}(t)$: tasso di falsi rifuiti relativa all'utente *user* quando la tolleranza del sistema assume il valore *t*, ed è pari al rapporto tra il numero di verifiche non superate dall'utente *user* sul il numero totale di prove effettuato dal soggetto.

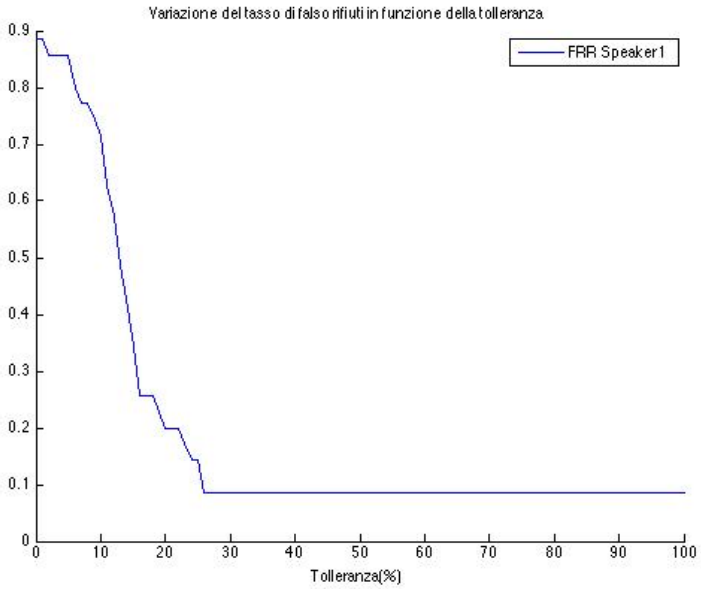

Figura 4.2: Funzione del tasso di falsi rifiuti che si riferisce all'utente speaker1.

4.2 Interpretazione dei risultati

Dai risultati ottenuti durante la sessione delle due sperimentazioni realizzate, abbiamo rilevato che non tutti gli utenti hanno le stesse prestazioni, questo è principalmente dovuto al fatto di parlare o di leggere male una frase stabilita; consideriamo le prestazioni del sistema, il valore medio delle prestazioni (rappresentato sulla figura 4.3.) degli utenti del sistema.

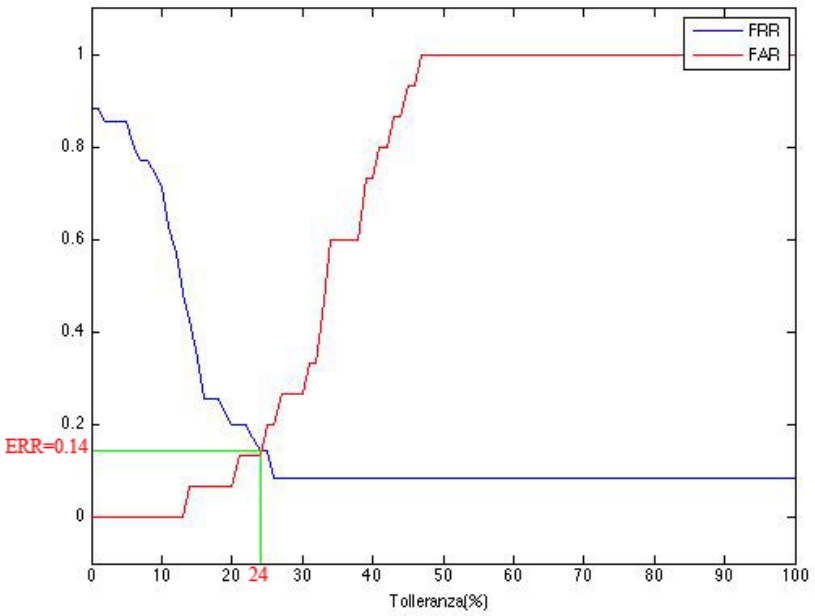

Figura 4.3: Percentuali di errori FRR e FAR del sistema.

Graficamente possiamo vedere che il valore di tolleranza per cui c'è parità perfetta tra i tassi di errori è 24% e quindi l'errore intrinseco del nuovo sistema di verificazione ERR vale 14%. Dal valore di ERR ottenuto, si potrebbe pensare che qualità del sistema è migliore visto che di solito, i valori ERR per i sistemi biometrici di riconoscimento della voce sono compressi tra il 10% e il 20%. In ogni caso sarebbe anticipato trarre una conclusione perché:

- I dati raccolti sono abbastanza pochi,
- Non abbiamo considerato lo stato d'animo dell'utente durante i test e quindi le prestazioni potrebbero essere peggiori se ogni verifica fossa effettuato in momenti differenti rispetto alla fase di registrazione.
- Il primo esperimento, si è svolto con livello di cooperazione minimale perché in realtà l'utente impostore non cerca di ingannare il sistema, pronuncia liberamente il codice vocale che lui stato richiesto, certamente un impostore cambia suo tono di voce per assumere colui di un altro.

4.3 Campi di Applicazione

I principali vantaggi che possiede un sistema biometrico di riconoscimento del locutore rispetto ad altri sistemi biometrici sono:

- È l'unico sistema biometrico che si può anche utilizzare tramite il telefono.
- Il costo della tecnologia è relativamente basso: basta un microfono.
- È considerato poco intrusivo rispetto ad altri.
- Non c'è nessun contatto tra il sensore e l'utente.

Per questi motivi, suo grado di apprezzamento da parte della comunità è il più elevato degli altri sistemi.

Data la potenzialità della biometria, e vista la diffusione e violabilità dei sistemi di autenticazione tradizionali (Pin, Badge, Password), questa tecnologia trova applicazioni in numerosi ambiti: commerciali, web-shops, giudiziari, trasporti, telecomunicazioni al fine di assicurare maggiore sicurezza durante le transazioni.

La voce umana può essere utilizzata come meccanismo per identificare in modo univoco. Le password, i pin, o i badge utilizzati per l'autenticazione sono metodi di verificazione "innaturale"! Non possono attestare con sicurezza l'identità della persona, ma semplicemente garantire che l'utente sia a conoscenza di qualcosa o la possegga. La voce è migliore rispetto a una password, perché essa non può essere persa, prestata, rubata o dimenticata e quindi, una voce, come un'impronta digitale, può essere usata per verificare l'identità di un utente invece di digitare una password. Si tratta di un metodo di autenticazione "naturale", l'utente deve "semplicemente" presentarsi di persona, il sistema di verifica di identità garantisce la presenza della persona.

Il sistema di verifica di identità basato sul parlato può essere utilizzato soprattutto

- Per controllare l'accesso ai sistemi, login su computer, commercio elettronico, …
- Per controllare l'accesso a locali protetti, controllo presenze, sorveglianza ambientale.

Questo sistema può essere utilizzato in alternativa o in complemento con altri meccanismi come il controllo delle impronte digitali, della retina dell'occhio o di un semplice PIN code.

4.4 Svantaggi

Il sistema non garantisce un'accuratezza del 100%, esistono molti problemi durante l'uso di una tecnologia di riconoscimento della voce come la nostra, infatti, basta pensare a una semplice variazione di distanza tra il locutore e il microfono o ancora pronunciare in modo non coretto la parola o la frase creano elementi di perturbazioni.

La qualità dei risultati è influenzata dalle condizioni di registrazione e generalmente degrada quando le condizioni nella fase di verificazione non coincidono con quelle della fase di registrazione. In questo contesto, le condizioni includono le condizioni ambientali (rumore, musica in sottofondo, ...), il comportamento dell'utente

(differente cadenza, stato d'animo, ...), ma anche le condizioni del canale trasmissivo (cambio del microfono utilizzato, ...).

Il livello dei rumori d'ambiente può essere tale da impedire la registrazione dei campioni. Algoritmi di riduzione del rumore possono essere utilizzati per migliorare l'accuratezza, ma l'applicazione scorretta di tali algoritmi può avere l'effetto contrario.

Le caratteristiche possono mutare nel tempo, perciò il normale cambiamento della voce dovuto all'età può inficiare il buon funzionamento del sistema, pertanto alcuni sistemi di riconoscimento della voce aggiornano i modelli dei parlatori dopo ogni verifica completata con successo.

Conclusioni

Nel corso di questo progetto, ci siamo arrivati all'obiettivo prefissato di simulare un processo di riconoscimento della voce. È evidente che il segnale acquisito dal microfono è una somma due segnale: un segnale vocalizzato carica d'informazioni e un segnale di disturbo (rumore ambientale, segnale non vocalizzato) privo d'informazioni, per questo motivo abbiamo trovato opportuno di individuare le regioni di disturbi ed elaborare solo sui segnali vocali vocalizzati. un segnale vocale vocalizzato contiene diverse informazioni specifiche sull'oratore quali il suo sesso, la sua età oppure la sua provenienza. Queste informazioni sono determinate dalle dimensioni del tratto vocale, la lunghezza delle corde vocali. Se la voce cambia con l'età è proprio perché queste caratteristiche del tipo fisiologiche cambiano anche loro con l'età.

Durante la fase di estrazione delle caratteristiche pertinenti, ci siamo limitati all'estrazione delle caratteristiche fisiologiche del tratto vacale ma questa fase può essere migliorata utilizzando delle tecniche di estrazione delle caratteristiche per la sorgente di eccitazione e per tratti comportamentali, tuttavia, il principale limite per entrambi è la non disponibilità di strumenti adeguati per l'estrazione di queste caratteristiche.

L'algoritmo DTW, ci ha consentito di definire una misura di somiglianza tra due impronte vocali.

La realizzazione del progetto ha portato all'esecuzione di una dimostrazione che permette di verificarne le sue potenzialità, utilizzando dati reali tenendo conto soprattutto delle condizioni ambientali, il funzionamento del sistema non ha fatto rilevare particolari problemi.

Sviluppi Futuri

Come possiamo migliorare il sistema?

Osservando struttura del sistema di verifica del locutore si può notare che la tecnica utilizzata per la realizzazione di ogni modulo non dipende dalle altre, quindi

possiamo cercare di migliore in modo indipendente ogni tecniche o ancora quelli che possono essere messi in discussioni.

Durante la fase di acquisizione del segnale vocale per esempio si può provare di migliorare la qualità dei dati acquisiti facendo l'uso degli algoritmi di riduzione del rumore.

Valutare le prestazioni del sistema.

Bibliografia

[1] T. Thrasyvoulou and S. Benton, Speech Parameterization using the Mel scale,

[2] Debnath Bhattacharyya, Rahul Ranjan, Farkhod Alisherov A., and Minkyu Choi, Biometric Authentication: A Review, International Journal of u- and e- Service, Science and Technology Vol. 2, No. 3, September, 2009

[3] Joseph P. Cammpbell, JR., Speaker Recognition: A Tutorial, ieee, vol. 85, no. 9, September 1997.

[4] S. Furui, "Recent Advances in Speaker Recognition", Pattern Recognition Letters, Vol. 18, No. 9, 1997.

[5] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, A. W. Senior, Guide to Biometrics.Springer, 2003.

[6] Bin Ma and Haizhou Li text-independent speaker recognition

[7] W. Andrews, M. Kohler, J.P. Campbell, and J. Godfrey. Phonetic, idiolectal, and acoustic speaker recognition. In Proceedings of the IEEE workshop on speaker and language recognition (Odyssey), 2001. 4.2.3

[8] W. Andrews, M. Kohler, J.P. Campbell, J. Godfrey, and J. Hernandez-Cordero. Gender- dependent phonetic refraction for speaker recognition. In Proc. ICASSP, 2002. 4.2.3, 4.3

[9] B.S. Atal. Effectiveness of linear prediction characteristics of the speech wave for au- tomatic speaker identification and verification. Journal of the Acoustical Society of America, 55:1304–1312, 1974.

[10] B.S. Atal and S.L. Hanauer. Speech analysis and synthesis by linear prediction. Journal of the Acoustical Society of America, 50:637–655, 1971.

[11] Shi-Huang Chen and Yu-Ren Luo Speaker Verification Using MFCC and Support Vector Machine IMECS 2009, March 18 - 20, 2009.

[12] Minho Jin, Frank K. Soong, and Chang D. Yoo, “A Syllable Lattice Approach to Speaker Verification,” IEEE Trans. Audio, Speech, and Language Processing, Vol. 15, No. 8, pp. 2476-2484, 2007.

[13] A.E. Rosenberg, “Automatic speaker verification: A review,” IEEE Proceedings, Vol. 64, pp. 475-487, 1976.

[14] Guiwen Ou and Dengfeng Ke, “Text-independent speaker verification based on relation of MFCC components,” 2004 International Symposium on Chinese Spoken Language Processing, pp. 57-60, Dec. 2004.

[15] A. Mezghani and D. O'Shaughnessy, “Speaker verification using a new representation based on a combination of MFCC and formants,” 2005 Canadian Conference on Electrical and Computer Engineering, pp. 1461-1464, May 2005.

[16] M.M Homayounpour and I. Rezaian, “Robust Speaker Verification Based on Multi Stage Vector Quantization of MFCC Parameters on Narrow Bandwidth Channels,” ICACT 2008, vol 1, pp.336-340, Feb. 2008

[17] C.C. Lin, S.H. Chen, T. K. Truong, and Yukon Chang, “Audio Classification and Categorization Based on Wavelets and Support Vector Machine,” IEEE Trans. on Speech and Audio Processing, Vol. 13, No. 5, pp. 644-651, Sept. 2005.

[18] L. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.

Webiografia

[1] http://www.animations.physics.unsw.edu.au/jw/dB.htm [What is a decibel?].

[2] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/ (Mel Frequency Cepstral Coefficient (MFCC) tutorial).

[2] http://www.firenetltd.it/index.php?option=com_content&view=article&id=190&Itemid=190 (cos è la biometria).

[3] http://www.aooi.it/contents/attached/c6/aerod.pdf (l'elettroglottografia e gli indici aerodinamici)