



UNIVERSITÀ POLITECNICA DELLE MARCHE  
FACOLTÀ DI INGEGNERIA

*Corso di Laurea triennale in Ingegneria Biomedica*

**Studio di parametri oggettivi per la  
valutazione della qualità sonora di  
apparecchi acustici**

Study of objective parameters for the evaluation of hearing  
aids

*Tesi di laurea di:*

**SANTORO FRANCESCO**

*Relatore:*

Prof. **CECCHI STEFANIA**

*Correlatori:*

Ing. **TERENZI ALESSANDRO**

Ing. **SPINSANTE SUSANNA**

---

*Anno Accademico 2019-2020*

# Abstract

Nell'ambito della progettazione di dispositivi di ausilio per individui ipoacusici, assicurare la qualità sonora è un fattore importante. Infatti, la maggior parte dei soggetti con ipoacusie utilizza degli ausili acustici per migliorare la capacità di ascolto. Ma questi dispositivi soffrono di varie problematiche come l'introduzione di un rumore di fondo o di una distorsione causati dall'amplificazione del segnale sonoro, e questo provoca un degrado della fedeltà e della qualità sonora dell'audio.

Al fine di risolvere le tematiche sopra citate, si stanno eseguendo vari tentativi con lo scopo di ridurre tali disagi e per migliorarne anche il confort acustico dei pazienti, con la ricerca di approcci migliori per stimare in maniera oggettiva le performance di tali dispositivi.

A tale scopo, il seguente elaborato riassume la letteratura già esistente riguardante la stima della qualità del parlato percepita, in particolare discutendo di un algoritmo per il calcolo del Mean Score Opinion (MOS). Sono stati eseguiti dei test sperimentali su ausili uditivi con differenti caratteristiche elettroacustiche, utilizzando un manichino binaurale.

# Indice

<b>INTRODUZIONE</b>	<b>4</b>
<b>CAPITOLO 1: APPROCCI PER LA VALUTAZIONE SOGGETTIVA E OGGETTIVA DELLA QUALITA' DEL PARLATO</b>	<b>6</b>
1.1 VALUTAZIONE SOGGETTIVA	6
1.2 VALUTAZIONE OGGETTIVA	7
<b>CAPITOLO 2: Normativa [ITU-T P.862]</b>	<b>9</b>
2.1 PESQ	9
2.2 PANORAMICA DEL PESQ	10
2.3 CONFRONTO TRA RISULTATI OGGETTIVI E SOGGETTIVI	11
<b>CAPITOLO 3: DESCRIZIONE DELL'ALGORITMO PESQ</b>	<b>12</b>
3.1 LIVELLO E PRE-ELABORAZIONE DELL'ALINEAMENTO TEMPORALE	15
3.2 MODELLO PERCETTIVO	16
3.3 MOS SCORE	16
3.4 VALORI MOS SOGGETTIVI	17
3.5 INTERPRETAZIONE VALORI MOS	18
3.6 VALORI OGGETTIVI DEL MOS	20
3.7 MOS RELATIVO ALLE SITUAZIONI DI ASCOLTO	20
3.7.1 MOS-LQS	20
3.7.2 MOS-LQO	20
<b>CAPITOLO 4: TEST SPERIMENTALI</b>	<b>22</b>
5.1 DESCRIZIONE DEL SOFTWARE UTILIZZATO	22
5.1 DESCRIZIONE SETUP DI MISURA	24
5.2 DISCUSSIONE DEI RISULTATI	29
<b>CONCLUSIONI</b>	<b>43</b>
<b>BIBLIOGRAFIA</b>	<b>44</b>

# Introduzione

L'udito è un senso molto importante per l'essere umano. Infatti, ci permette di poter avere delle relazioni sociali con il mondo, ascoltare musica e persino mantenere l'equilibrio. Tutto ciò rappresenta un problema nei soggetti audiolesi, in quanto presentano dei disagi psicologici vista l'assenza della possibilità di poter avere delle relazioni sociali con il mondo. Per risolvere questo problema sono state introdotte le protesi acustiche che permettono a questi soggetti di poter interfacciarsi con il mondo esterno.

La protesi acustica è composta da un microfono che capta le onde sonore incidenti sulla membrana acustica, un amplificatore che amplifica in modo selettivo le onde sonore e un ricevitore che invia queste onde amplificate al timpano del soggetto in modo da poterle percepire.

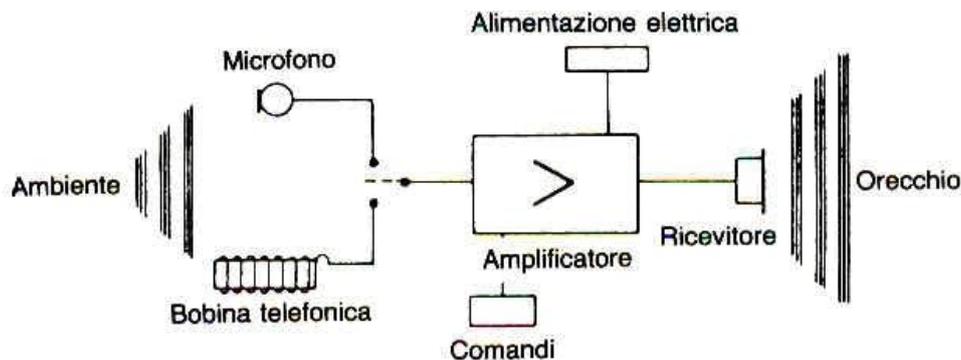


Figura 1 Schema apparecchio acustico

Grazie al progresso tecnologico che ha portato a numerosi sviluppi in campo audio-protesico, si sono realizzati apparecchi acustici di dimensioni sempre più ridotte diventando anche più discrete ed invisibili. Inoltre, il progresso tecnologico ha permesso lo sviluppo di apparecchi digitali che ha introdotto molti vantaggi rispetto agli apparecchi acustici analogici. Tuttavia, nonostante questi dispositivi nascano con lo scopo di poter migliorare la qualità della comunicazione dei soggetti audiolesi, possono presentare dei fattori che vanno a degradare le performance andando a introdurre distorsioni e a produrre degli effetti negativi.

Il presente studio discute un approccio che permette di stimare oggettivamente la qualità sonora degli apparecchi in modo da poter stabilire in anticipo la qualità.

Il lavoro è così suddiviso.

Il capitolo 1 fornisce una visione degli approcci utilizzati nella valutazione della qualità oggettiva e soggettiva del parlato.

Il capitolo 2 fornisce una panoramica sullo scopo della normativa [ITU-T P.865], fornisce un'introduzione dell'algoritmo PESQ (Perceptual Evaluation Speech Quality) ed espone la differenza tra le misure oggettive e soggettive.

Il capitolo 3 fornisce una descrizione più accurata dell'algoritmo PESQ. Inoltre, fornisce una descrizione e comprensione dei valori del MOS, generati in uscita dall'algoritmo precedentemente descritto, per comprendere la qualità del parlato nella misura percettiva.

Infine, il capitolo 4, riguarda i test sperimentali ed i risultati ottenuti.

# Capitolo 1

## Approcci per la valutazione soggettiva e oggettiva della qualità del parlato

Per tanti anni la valutazione del parlato si è basata sulle misure soggettive, di conseguenza era affidata all'orecchio umano. Nel corso degli anni, con l'evoluzione della tecnologia, si è introdotta la valutazione oggettiva tramite l'utilizzo di algoritmi.

In questo capitolo verrà fornito un confronto tra i due metodi di valutazione.

### 1.1 Valutazione soggettiva<sup>[11]</sup>

Per poter valutare la qualità del parlato si usano dei soggetti come “giudici”. Essi devono testare i campioni vocali all'interno di una camera silenziosa con un volume compreso tra i 30 e i 120  $m^3$  e un tempo di riverbero inferiore ai 500 ms, di norma tra i 200-300 ms. Inoltre, il livello del rumore della camera dev'essere al di sotto di 30 dBA senza picchi dominanti nello spettro. I soggetti devono fornire una valutazione numerica della qualità con un valore che varia tra 1 e 5, dove 1 rappresenta una pessima qualità e 5 una qualità eccellente. Queste valutazioni sono il punto di riferimento per i metodi oggettivi, ma sono lenti, infatti richiedono tempo e costi.

L'affidabilità dei risultati è data dal fatto che vi sono dei soggetti reali per la valutazione della qualità. Elenchiamone ora i vantaggi:

- Fornisce una valutazione numerica della qualità del parlato
- Si concentra sul discorso reale
- Usa persone reali per giudicare la buona qualità dalla cattiva
- Standardizzato dall'ITU-T P.830

Ma in tutto ciò vi sono presenti degli svantaggi:

- Tempo tra processi di preparazione, pianificazione e lavorazione
- Costi alti

- Istantanea nel tempo rispetto alla preferenza del giudice
- Inesperienza
- Errore umano
- Interpretazione

Alla luce di questi svantaggi, l'ITU (International Telecommunication Union) ha standardizzato un algoritmo per la stima oggettiva.

## 2.2 Valutazione oggettiva<sup>[4]</sup>

Per ovviare ai problemi delle valutazioni soggettive, nel 2001, l'ITU finalizza un algoritmo, chiamato PSQM, che permette di valutare in maniera oggettiva la qualità del parlato seguendo le linee guida della normativa P.861. Questo algoritmo rilascia un valore MOS (Mean Opinion Score, utilizzato anche nelle misure soggettive, che determina la qualità del parlato.

Nel 1996, l'ITU standardizzò il PSQM<sup>[16]</sup> (Perceptual Speech Quality Measure) come [ITU-T P.861] ed era utilizzato principalmente per la trasmissione mobile, come il GSM. Successivamente, con l'avanzare degli anni, i requisiti richiesti dalle apparecchiature hanno spinto l'ITU alla standardizzazione del PESQ che si basa sulla previsione della qualità della voce di esseri umani. Infine, nel 2011, l'ITU standardizza il POLQA<sup>[15]</sup> (Perceptual Objective Listening Quality Analysis) che permette di prevedere la qualità del parlato mediante l'analisi del segnale vocale digitale. Questo metodo di valutazione è considerato intrusivo, infatti è più accurato, ma non sono adatti per il monitoraggio del traffico dei dati in tempo reale a causa della necessità di dati di riferimento e per l'utilizzo della rete.

Invece, i metodi non intrusivi, sono una misura basata sull'osservazione di parametri che consentono l'individuazione della qualità del segnale vocale<sup>[17]</sup>.

Nella stima della qualità acustica vi sono due componenti essenziali: trasformazioni percettive e misure di distanza. La trasformazione percettiva modifica una rappresentazione di un segnale audio in un modo che è, in maniera approssimativa, equivalente al processo dell'udito umano. Invece, una misura di distanza riflette l'ampiezza di una distanza percepita tra due segnali percettivamente trasformati. Questo approccio utilizza una trasformazione percettiva e una misura di distanza che consiste in una gerarchia di misurazione dei blocchi di normalizzazione. Ogni blocco integra due segnali trasformati percettivamente in un intervallo di tempo o di frequenza per determinare la differenza media in tale intervallo. Questa differenza viene normalizzata da un segnale e viene elaborata per generare una o più misurazioni<sup>[18]</sup>.

I vantaggi rispetto alle valutazioni soggettive sono dati dall'abbattimento dei costi, del tempo e l'abolizione della necessità di convocare una giuria per la valutazione della qualità. Invece l'unico svantaggio è che questo approccio non ci fornisce indicazioni su quale sia la caratteristica che riduce la qualità del suono, cosa che in una misura soggettiva è possibile in quanto è il soggetto stesso a spiegarlo secondo la sua valutazione.

# Capitolo 2

## Normativa [ITU-T P.862]

Questa normativa fornisce delle informazioni importanti e fondamentali per poter ricavare informazioni affidabili, stabili e, nella pratica, risultati significativi delle misurazioni oggettive. La funzione principale della normativa è quella di testare la rete dal vivo, ove si valuta il sistema in condizioni di reali.

### 2.1 PESQ

Il PSQM è un algoritmo computazionale che valuta oggettivamente e quantifica la qualità in un range di frequenza di 300-3400 Hz, proprio perché rappresenta il range di frequenza della voce umana. Nel corso degli anni i requisiti richiesti dalle apparecchiature di misurazione sono cambiati radicalmente. Di conseguenza, l'ITU ha istituito un nuovo gruppo di lavoro per revisionare lo standard P.861 affinché potesse fronteggiare le nuove esigenze derivante dalle moderne reti, ad esempio il VoIP (Voice Over IP). Queste reti moderne mettono l'algoritmo alla prova con delle distorsioni molto più elevate a differenza dei codec GSM, ma il fattore più importante riguarda il ritardo tra il segnale di riferimento e il segnale di test che non è più un segnale costante.

Con il nuovo standard [ITU P.862] (PESQ) questo problema viene eliminato. Infatti, l'algoritmo PESQ combina il modello psicoacustico e cognitivo del PSQM con un algoritmo di allineamento temporale che gestisce alla perfezione i ritardi variabili.

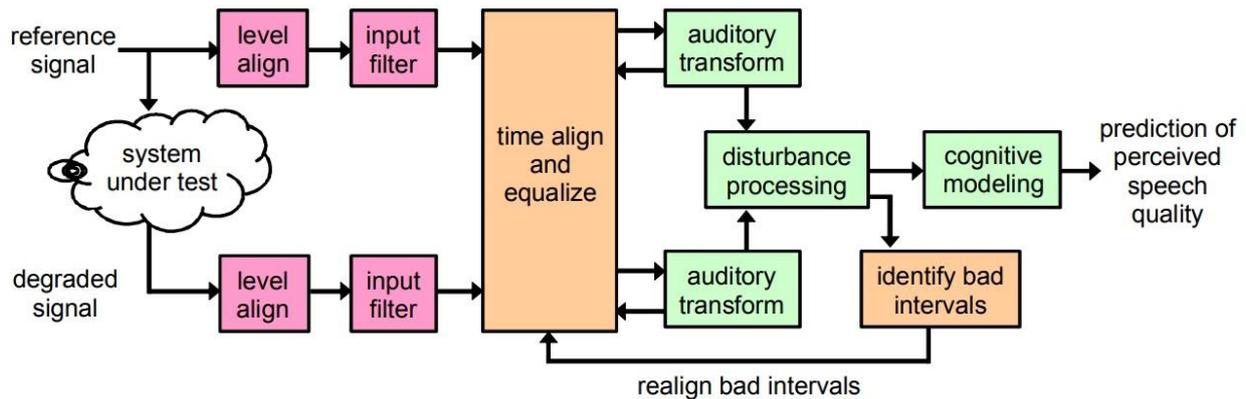


Figura 2 Struttura dell'algoritmo PESQ

## 2.2 Panoramica del PESQ

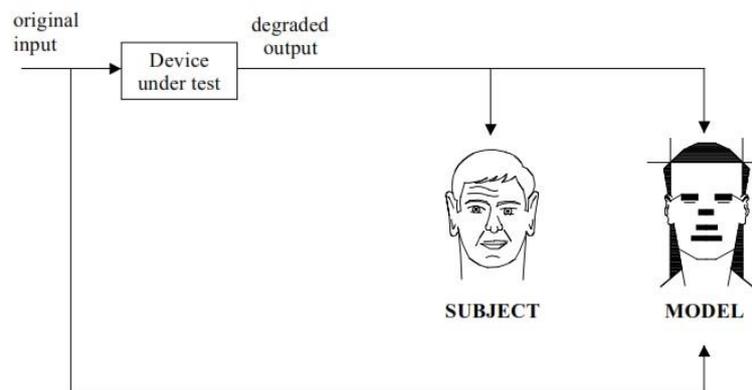
L'algoritmo PESQ confronta un segnale originale  $X(t)$  con un segnale degradato  $Y(t)$  che rappresenta il risultato dato dal passaggio del segnale  $X(t)$  attraverso il sistema che dev'essere testato. La risposta in uscita è una previsione della qualità percepita che viene data a  $Y(t)$  dai soggetti in un test di ascolto soggettivo (figura 2).

Nel primo step del PESQ si calcolano una serie di ritardi tra il segnale in input originale e degradato per ogni intervallo di tempo ove il ritardo è significativamente diverso dal precedente intervallo di tempo. Per ogni singolo intervallo viene calcolato un punto iniziale ed uno finale.

L'algoritmo, sulla base della serie di ritardi rilevati, confronta il segnale in ingresso (originale) con la sua versione degradata e allineate nel tempo, in uscita dal dispositivo che si usa in prova utilizzando un modello percettivo. La chiave di questo processo sta nella trasformazione dei segnali originali e degradati in una rappresentazione interna, conforme alla rappresentazione psicofisica dei segnali audio nel sistema uditivo dell'uomo, tenendo in considerazione della frequenza percettiva, misurata in Bark<sup>[13]</sup>, e del volume, misurato in Sone<sup>[14]</sup>. La scala Bark va da 1 a 24 Bark corrispondenti alle prime 24 bande critiche dell'udito, invece la scala del Sone rappresenta la misura del volume "soggettivo". Questo risultato è raggiunto in diversi step:

- Allineamento temporale
- Allineamento di livello ad un livello di ascolto calibrato
- Mappatura in tempo-frequenza

- Deformazione della frequenza
- Scalatura del volume compressivo



*Figura 3 panoramica della filosofia base utilizzata nel PESQ*

## 2.3 Confronto tra risultati oggettivi e soggettivi

Nei test soggettivi, i risultati della giuria sono influenzati da alcuni fattori quale le preferenze dei singoli soggetti e il contesto del test stesso. Difatti, è necessario effettuare un processo di regressione prima di effettuare un confronto con i soggetti stessi. Questa regressione dev'essere di tipo monotona per far sì che le informazioni siano preservate, ed è utilizzato per poter mappare il risultato della misura oggettiva sul punteggio soggettivo.

Infatti, un metodo preferito per poter calcolare la correlazione tra il punteggio oggettivo e quello soggettivo consiste nell'utilizzo di un polinomio di terzo ordine, vincolato ad essere monotono. Quindi, per poter effettuare un confronto tra i punteggi oggettivi e soggettivi, i punteggi oggettivi del MOS devono derivare da un test di ascolto eseguito secondo le regole presenti nella normativa [ITU-T P.830].

# Capitolo 3

## Descrizione dell' algoritmo PESQ

Dato che molti passaggi nel PESQ sono complessi da esporre dal punto di vista algoritmico, una descrizione a livello matematico non è facile. Per ottenere una descrizione dettagliata dell' algoritmo, il lettore si basa su un codice sorgente in C. Le figure 4, 5a e 5b forniscono una panoramica dell' algoritmo in forma di diagramma a blocchi: nella figura 4 è descritto il processo riguardante l' allineamento, nella figura 5a è descritto il nucleo del modello percettivo e, infine, nella figura 5b è descritta la determinazione finale del punteggio PESQ.

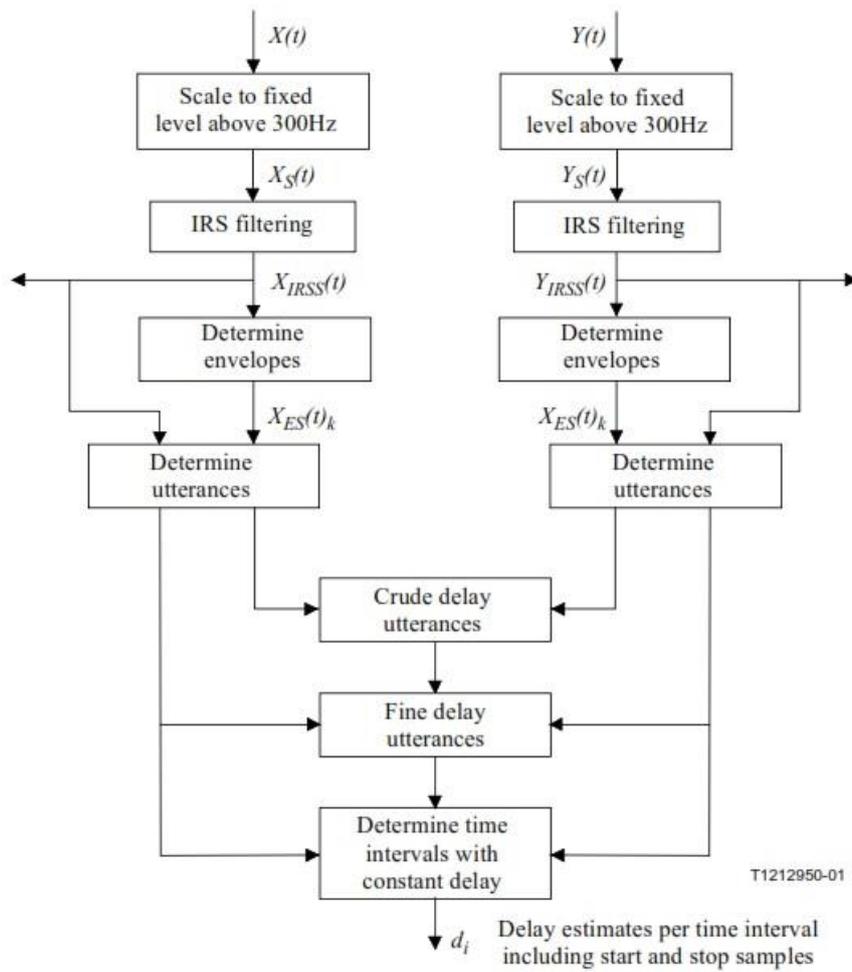


Figura 4: Panoramica della routine di allineamento utilizzata in PESQ per determinare il ritardo per intervallo di tempo  $d_i$

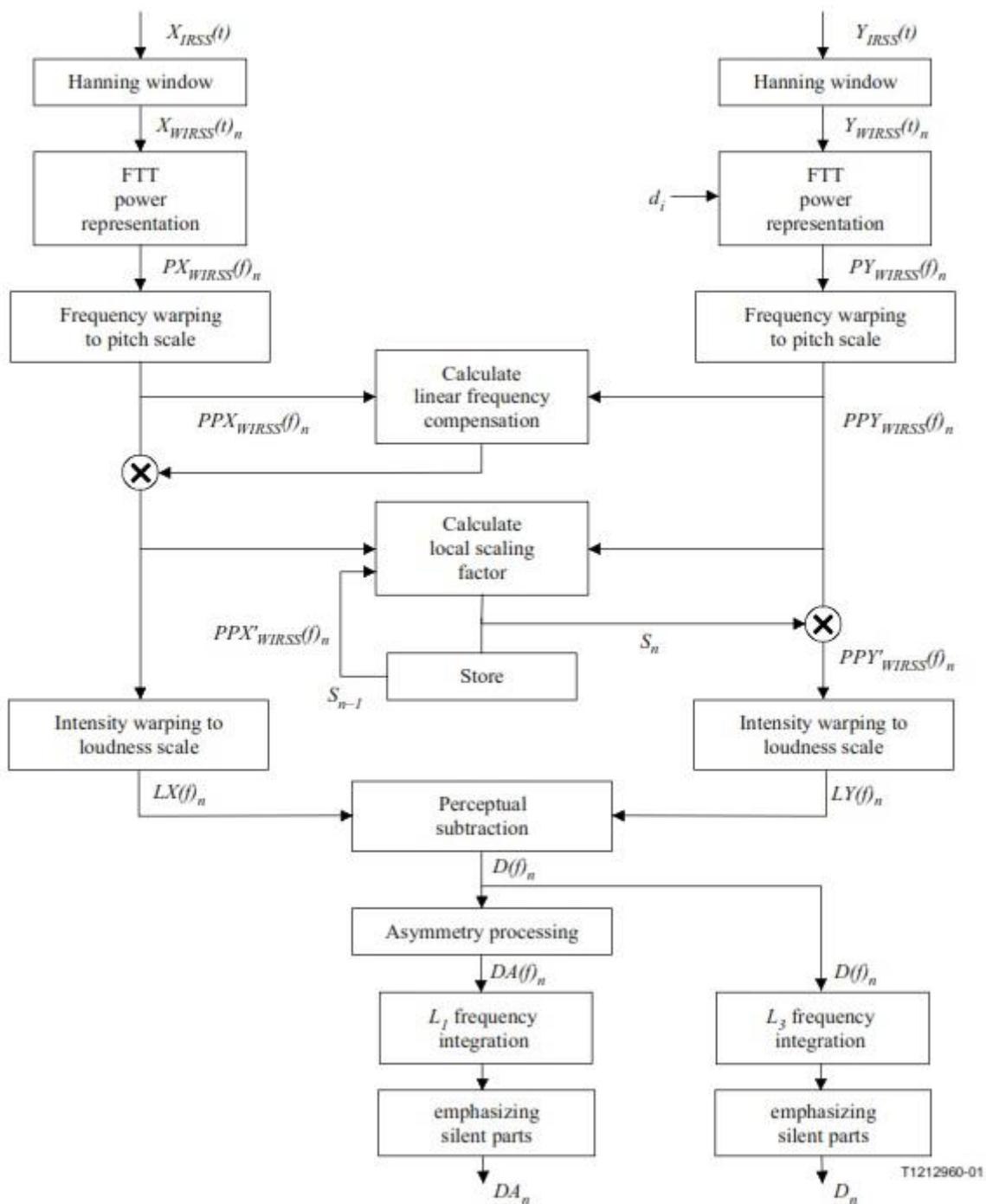


Figura 5a: Nucleo del modello percettivo

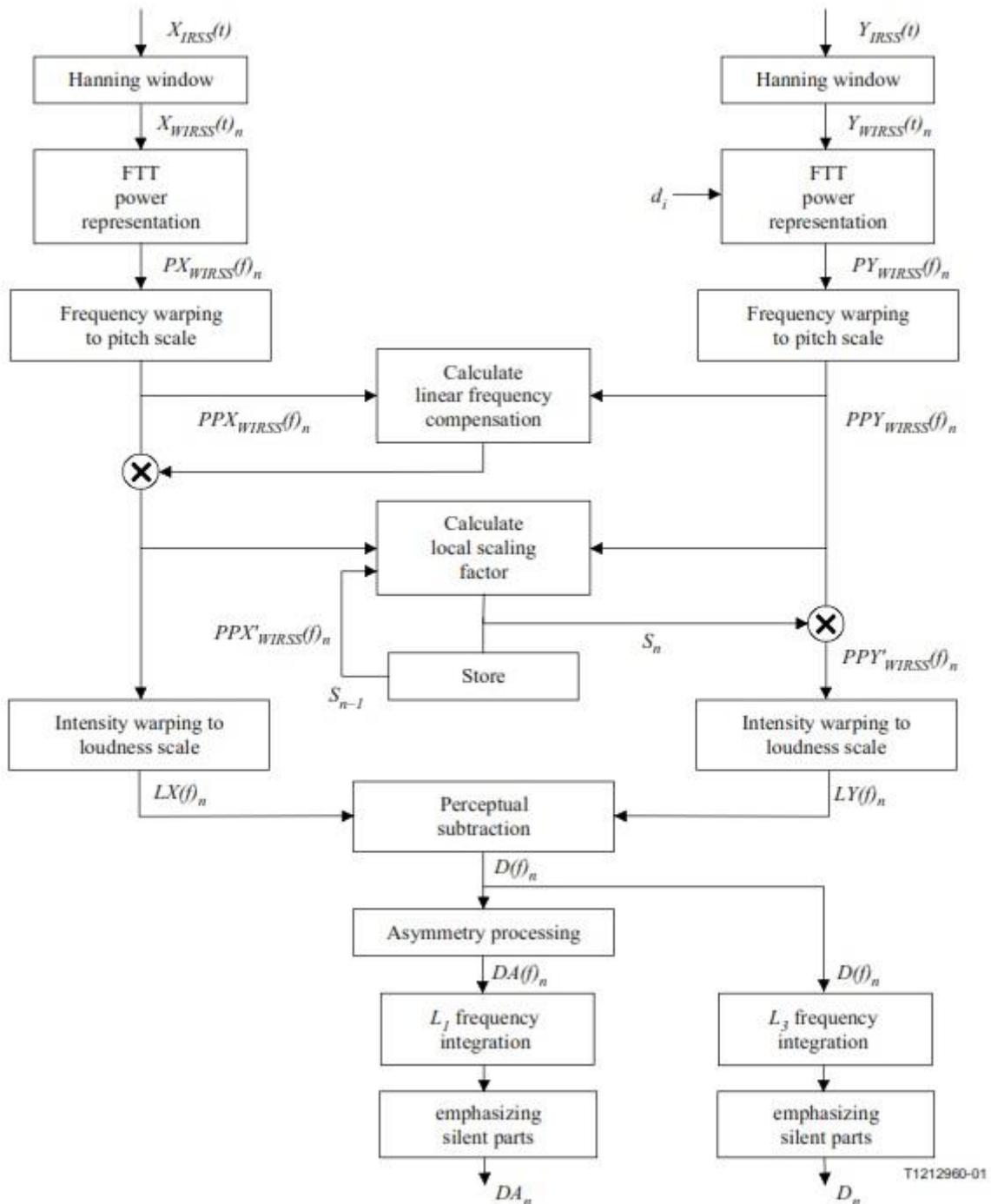


Figura 5b: Determinazione punteggio PESQ

### 3.1 Livello e preelaborazione dell'allineamento temporale

Come descritto nella figura 4 possiamo notare che le componenti principale di questo schema a blocchi sono:

- Calcolo del guadagno complessivo del sistema: nel sistema in prova non è noto e può variare notevolmente. Per questo è necessario allineare a livello il segnale originale e degradato con un livello di potenza che PESQ assume essere un SPL costante di 79 dB.
- Filtraggio IRS (Intermediate Reference System): in quanto si presume che i test di ascolto siano stati effettuati con l'utilizzo di una ricezione IRS per la definizione dei livelli di rumorosità. Infatti, l'IRS riceve sia le versioni filtrate del segnale vocale originale e di quello degradato.
- Allineamento temporale: fornisce i valori di ritardo, per consentire al modello percettivo di confrontare le parti corrispondenti di segnale del file originale e degradato.

### 3.2 Modello percettivo

Questo modello del PESQ è utilizzato per il calcolo della distanza tra il segnale vocale originale e degradato. Esso fornisce una previsione del MOS soggettivo per un dato test soggettivo. Da qui il punteggio del PESQ viene mappato su una scala MOS con un intervallo di uscita compreso tra 1 e 4,5 e che rappresenta l'intervallo normale di valori MOS ricavati in un esperimento ACR.

### 3.3 MOS score

La qualità audio e video sono quantità intrinsecamente soggettive. Ciò significa che la linea di base per la qualità dell'audio e del video sono opinione dell'utente. Tuttavia, l'opinione di una persona su ciò che può considerare "buono" può essere molto diversa dall'opinione espressa da un'altra persona: nessuna delle due opinioni è da considerare corretta o sbagliata.

Prima di implementare una nuova tecnologia di trasmissione audio o video è buona norma valutare la qualità usando uno o più esperimenti soggettivi. Lo scopo di un esperimento soggettivo consiste nel raccogliere le opinioni di più persone (soggetti) sulle prestazioni del sistema di un numero di casi d'uso ben definiti (condizioni). Il MOS per una data condizione è semplicemente la media delle opinioni (voti) raccolte per quel caso d'uso.

Gli algoritmi di misurazione della qualità oggettiva mirano a prevedere il valore MOS che un dato segnale di ingresso produrrebbe in un esperimento soggettivo. Quindi, quando si interpreta un valore MOS derivato oggettivamente, è essenziale comprendere il progetto di base dell'esperimento previsto.

### 3.4 Valori MOS soggettivi

Esiste un malinteso comune secondo cui i valori MOS riguardino solo i servizi vocali, ma il processo di chiedere ai soggetti di fornire la loro valutazione della qualità può essere applicato anche al video. È anche possibile chiedere ai soggetti di valutare la qualità audiovisiva di un servizio. L'ITU ha prodotto vari standard che descrivono diversi aspetti di test soggettivi per applicazioni vocali e sono elencati nelle normative appartenenti alla serie P.800.

Gli esperimenti soggettivi possono essere sostanzialmente di due tipi: passivi e interattivi. In un esperimento soggettivo passivo, ai soggetti vengono presentati campioni di test preregistrati che rappresentano le condizioni di interesse. Ai soggetti viene chiesto di ascoltare passivamente e/o guardare il materiale di prova fornendo la propria opinione utilizzando la scala di giudizio fornita. In un esperimento interattivo, due o più soggetti partecipano attivamente alla conversazione utilizzando apparecchiature con lo scopo di emulare i casi d'uso di interesse. Ai soggetti vengono assegnati dei compiti per stimolare la conversazione e l'interazione. Generalmente, la maggior parte degli esperimenti tendono ad essere di natura passiva. Tuttavia, ci sono alcuni aspetti dell'esperienza utente, per esempio, gli effetti del ritardo e dell'eco, che diventano evidenti solo negli scenari di conversazione.

Di solito, in un esperimento soggettivo, ai soggetti viene chiesto di fornire le proprie opinioni utilizzando una scala di valutazione. Lo scopo di tale scala è quello di tradurre la valutazione della qualità fornita da un soggetto in un valore numerico che può essere la media tra soggetti e altri fattori sperimentali.

Esistono diverse scale di valutazione di uso comune e i relativi vantaggi delle scale diverse sono al di fuori dal campo di applicazione della seguente normativa. La scala più usata è a cinque punti su una scala ACR:

1. Pessimo
2. Scarso
3. Accettabile
4. Buono
5. Eccellente

La scala ACR è una scala discreta, il che significa che la risposta del soggetto è limitata a uno dei cinque valori sopra elencati. Tuttavia, il processo di calcolo della media utilizzato per combinare i risultati di diversi soggetti significa che i valori MOS non sono limitati a valori interi. Alcune scale di valutazione hanno più di cinque etichette discrete, mentre altre consentono al soggetto di fornire risposte intermedie nei punti tra le etichette.

La parte assoluta dell'ACR si riferisce al fatto che ai soggetti viene chiesto di valutare ciascuno in modo indipendente un campione. Alcune scale di valutazione, come la Degradation Category Rating (DCR), richiedono un'opinione soggettiva tra un campione elaborato attraverso la condizione di interesse e una versione non elaborata

dello stesso campione. Il valore MOS prodotto in un tale esperimento è chiamato MOS di degradazione o DMOS.

Nella maggior parte dei progetti sperimentali, ai soggetti viene chiesto di valutare la qualità di brevi campioni audio o video. Tali campioni hanno una durata solitamente compresa tra 6 e 10 secondi, poiché fornisce al soggetto il tempo sufficiente per formarsi un'opinione senza introdurre alcun pregiudizio verso la fine del campione. È difficile per un singolo campione di questa durata rappresentare un'intera condizione e quindi ai soggetti viene generalmente chiesto di valutare più campioni derivati dallo stesso caso d'uso. Ad esempio, in un esperimento vocale, ogni condizione di rete sotto test potrebbe essere rappresentata con campioni vocali da tre oratori maschi e tre femmine. Ciò significa che i valori MOS possono essere prodotti per l'intera condizione facendo la media tra soggetti e oratori, o per una sotto-condizione, come un particolare oratore o genere di chi parla.

I metodi di test, come la valutazione continua della qualità a stimolo singolo (SSCQE), utilizzano campioni di prova molto più lunghi, e richiedono al soggetto di aggiornare continuamente la propria opinione sulla qualità così come lo si aggiorna il campione in prova in esecuzione. Ciò si traduce in una sequenza temporale di valutazioni di qualità da ciascun soggetto, piuttosto che da un singolo valore di opinione.

Alcune metodologie di test, invece, richiedono che il soggetto risponda a più domande. Questo non solo fornisce più informazioni sulle condizioni in prova, ma può essere una parte necessaria del progetto del test. Ad esempio, il metodo di test ITU-T P.835 richiede che il soggetto fornisca opinioni separate sulla qualità del parlato e sulla qualità del rumore di un campione prima di fornire un punteggio di qualità complessivo. È stato riscontrato che questo processo produce risultati più stabili con sistemi di soppressione del rumore rispetto al metodo di prova ACR a domanda singola. Va notato che alcune domande potrebbero non essere direttamente correlate alla qualità, ma potrebbero riguardare un aspetto diverso della comunicazione, ad esempio, [b-ITU-T P.800] definisce una scala dello sforzo di ascolto per gli esperimenti vocali. Allo stesso modo, alcuni esperimenti di conversazione chiedono al soggetto la sua esperienza quando parla, piuttosto che durante l'ascolto.

### **3.5 Interpretazione valori MOS**

La discussione si concentra inizialmente sui valori vocali del MOS; tuttavia, molti dei punti evidenziati nelle sottosezioni si applicano ugualmente ai valori MOS video, audio e audio-video.

L'idea che un particolare codec vocale abbia un particolare punteggio MOS è un altro comune malinteso. Una fonte di questo malinteso è l'uso diffuso di modelli oggettivi di valutazione della qualità, che producono risultati molto ripetibili. Tali modelli sono progettati per prevedere o stimare l'output di esperimenti soggettivi; tuttavia, per un dato codec a una data velocità di trasmissione, il valore MOS ottenuto in un

esperimento soggettivo può variare sostanzialmente da esperimento ad esperimento. Ci sono molte ragioni per questo.

In primo luogo, i valori MOS esatti ottenuti per una particolare condizione in un esperimento soggettivo possono essere influenzati da un gran numero di fattori, inclusi ma non limitati a:

- Le istruzioni impartite ai soggetti e la formulazione della scala di opinione;
- L’attrezzatura utilizzata per presentare il materiale;
- Presentazione monofonica, diotica binaurale o stereo;
- Livello di presentazione
- Ambiente acustico;
- Profilo del soggetto;
- Differenze nell’interpretazione e nell’uso di scale di valutazione tra le culture;
- Materiale vocale
- Linguaggio;

In secondo luogo, il valore MOS esatto ottenuto per una data condizione in un esperimento soggettivo dipende dalla qualità delle altre condizioni dell’esperimento. Ad esempio, una condizione del codec vocale ITU-T G.729 può ottenere un punteggio superiore a 3,9 in un esperimento ACR se la maggior parte delle altre condizioni sono di qualità peggiore rispetto all’ITU-T G.729; al contrario la condizione ITU-T G.729 può ottenere un punteggio significativo inferiore a 3,9 se la maggior parte delle condizioni mostra una qualità migliore.

In terzo luogo, se un esperimento viene eseguito con codec che operano a larghezze di banda audio diverse, la presenza di condizioni di larghezza di banda più elevate ridurrà il MOS prodotto per condizioni con un audio inferiore alla larghezza di banda. La massima larghezza di banda audio presente in un esperimento vocale è spesso chiamata il “contesto” dell’esperimento. Ad esempio, una condizione di codec vocale ITU-T G.711 spesso produrrà un punteggio superiore a 4,0 in un esperimento a banda stretta (300-3700 Hz), considerando che è più probabile che produca un punteggio tra 3,5 e 3,7 in un esperimento ACR a banda larga (50-7000 Hz), a causa della presenza di campioni a banda larga di qualità superiore.

Questi due ultimi punti riflettono il fatto che i soggetti negli esperimenti tendono ad adattare il loro uso della scala di valutazione al contenuto dell’esperimento. In effetti, esperimenti ben progettati includono un periodo di pratica all’inizio dell’esperimento quando i soggetti ascoltano esempi di una serie di condizioni, comprese le migliori e le peggiori.

Una delle conseguenze più importanti delle considerazioni sopra descritte è che non è significativo confrontare direttamente i valori MOS prodotti da esperimenti separati a meno che quegli esperimenti siano stati progettati esplicitamente per essere confrontati, e anche in questo caso i dati dovrebbero essere analizzati staticamente per garantire che tale confronto sia valido.

## 3.6 Valori oggettivi del MOS

Lo scopo di un modello di qualità oggettiva è quello di prevedere il valore MOS che un segnale audio o video otterrebbe in un esperimento soggettivo. Come discusso in precedenza, il valore MOS prodotto in un dato esperimento per un particolare codec o catena di trasmissione dipende da molti aspetti diversi della progettazione e dell'esecuzione dell'esperimento. I progettisti di modelli oggettivi devono prevedere un esperimento idealizzato. Questo è in genere un esperimento condotto in base a una metodologia di prova specifica, solitamente in ACR, e include un campione equilibrato delle distorsioni che si incontreranno nell'area di applicazione di interesse.

Uno dei vantaggi di un modello oggettivo è che i risultati sono ripetibili e quindi le misurazioni realizzati in tempi e luoghi diversi possono essere confrontati direttamente. Tuttavia, si dovrebbe comunque prestare attenzione poiché alcuni fattori, come la scelta del materiale di prova e qualsiasi pre o post-elaborazione, possono ancora introdurre una distorsione nei risultati.

Per ragioni che ora dovrebbero essere evidenti, diversi modelli oggettivi possono produrre previsioni diverse dei valori MOS per le stesse condizioni. Ad esempio, i modelli ITU-T P.862.1 e ITU-T P.863 non producono esattamente gli stessi valori MOS per la voce modificata previsti dall'ITU-T G.729, anche se questo codec rientra nell'ambito di entrambi i modelli. Ciò è in parte dovuto al fatto che i due modelli sono stati addestrati e ottimizzati utilizzando diversi esperimenti soggettivi. Per questo motivo, quando si confrontano previsioni MOS oggettive con soglie.

## 3.7 MOS relativo alle situazioni di solo ascolto

Questi punteggi MOS sono applicabili a una situazione di solo ascolto. Vi sono due casi diversi e distinti:

- MOS-LQS: MOS Listening Quality Subjective
- MOS-LQO: MOS Listening Quality Objective

### 3.7.1 MOS-LQS

Questo punteggio è stato raccolto in un test di laboratorio calcolando il valore medio aritmetico dei giudizi soggettivi in una scala di qualità ACR a 5 punti come definito in [ITU-T P.800]. I test soggettivi eseguiti secondo [ITU-T P.830], [ITU-T P.835] e [ITU-T P.840] forniscono i risultati in termini di MOS-LQS.

### 3.7.2 MOS-LQO

Tale punteggio è calcolato mediante un modello oggettivo che mira a prevedere la qualità di una situazione di prova di solo ascolto. Le misurazioni oggettive effettuate

utilizzando i modelli forniti in [ITU-T P.862.1], [ITU-T P.862.2] e [ITU-T P.862] forniscono i risultati in termini di MOS-LQO.

In questo contesto, va notato che è necessaria una distinzione tra misurazioni end-to-end intrusive della qualità dell'ascolto finalizzate a:

- MOS-LQO (elettrico): Questo tipo di misurazione viene eseguita solo nelle interfacce elettriche. Per prevedere la qualità dell'ascolto percepita dall'utente, vengono fatte delle ipotesi per i terminali in termini di IRS o risposta in frequenza IRS corretta: questo implicitamente, include l'ipotesi di una condizione sigillata tra il ricevitore e l'orecchio dell'utente. [ITU-T P.862] rientra in questa categoria.
- MOS-LQO (acustico): Questo tipo di misurazione viene eseguita sulle interfacce acustiche. Per prevedere la qualità dell'ascolto percepita dall'utente, questa misurazione include il telefono vero e proprio impostare i prodotti forniti dal produttore o venditore. In combinazione con la scelta del ricevitore acustico nel test di laboratorio, ci sarà una condizione più o meno permeabile tra il ricevitore del portatile e l'orecchio artificiale. Di conseguenza, per scenari di test più realistici potrebbe esserci una degradazione del valore MOS misurato, mentre per scenari di test più artificiali potrebbe esserci una differenza trascurabile.

# Capitolo 4

## Test sperimentali

Per lo studio dei parametri oggettivi, le misure sono state eseguite su un manichino Brüel & Kjaer 4128C<sup>[12]</sup> costituito da una testa e busto strutturati con l'obiettivo di riprodurre al meglio le proporzioni reali della fisionomia dell'uomo. Sulla testa di questo manichino sono presenti delle orecchie realizzate con calco di orecchio umano per simulare in maniera più accurata possibile le caratteristiche del padiglione auricolare umano. Su di esso sono stati testati tre tipologie di apparecchi acustici con caratteristiche elettroacustiche differenti. Le risposte registrate dall'uscita degli apparecchi acustici sono state successivamente processate con l'algoritmo in questione.

Si procede, quindi, con la descrizione del software utilizzato, per poi seguire con la descrizione del setup sperimentale e, infine, con le stime dei valori MOS ricavati dai test e con annessa valutazione.

## 5.1 Descrizione del software utilizzato

In questa sezione viene descritto tutto il software che è stato utilizzato durante i test sperimentali.

In primis si pone l'attenzione sull'algoritmo in C. La normativa fornisce un'implementazione software così composta:

- File dsp.c: routine DSP di base
- Libreria dsp.h: libreria per il file dsp.c
- Libreria pesq.h: libreria generale
- File pesqdsp.c: routine PESQ DSP
- File pesqio.c: file input/output
- File pesqmain.c: programma principale
- File pesqmod.c: modello ad alto livello del PESQ
- Libreria pesqpar.h: definizione del modello percettivo del PESQ

Per poter eseguire l'algoritmo è stato creato un progetto, tramite Visual Studio, contenente tutti i cinque file C e le tre librerie. Subito dopo, viene compilato il pesqmain per poter ottenere il file eseguibile che ci permette di eseguire il codice fuori dall'ambiente di sviluppo. Una volta ottenuto il file eseguibile, si apre il prompt dei comandi, in questo studio è stato usato il prompt di Windows, si dichiara la cartella di destinazione del file eseguibile e si scrive la seguente riga di codice:

```
pesqmain +8000/16000 ref.wav deg.wav
```

dove:

- Pesqmain rappresenta il richiamo al file eseguibile
- +8000/16000 rappresentano le frequenze di campionamento dei file vocali
- Ref indica il file wav di riferimento, cioè il file vocale originale
- Deg indica il file wav di confronto, cioè il file vocale degradato

Per esempio, se si vuole confrontare il file originale, di nome or105, e il suo degradato, di nome dg105, entrambi campionati ad una frequenza di 8000 Hz, dovremmo scrivere la funzione in questo modo:

**pesqmain +8000 or105.wav dg105.wav.**

Nello step successivo si va ad utilizzare un codice in Matlab con lo scopo di richiamare la funzione C in modo da verificare che l'algoritmo funzioni correttamente. Per poter verificare ciò si usa la seguente riga di comando per richiamare l'eseguibile C:

```
[Return,strout]=system(['pesqmain ',rate,' ',CleanSignal,' ',DegradedSignal])
c=strfind(strout,'(Raw MOS, MOS-LQO):');
```

La prima riga richiama l'eseguibile del pesq, la frequenza, il segnale originale e degradato, e stabilisce anche l'ordine con cui dovrebbero essere scritte nella finestra di comando affinché esso compili con successo. Nella seconda riga invece è definito cosa stamperà in output.

Per ottenere i risultati per il confronto con quelli ottenuti dall'algoritmo in C, nella finestra di comando di Matlab si scrive la seguente riga di codice:

**pesqmain +8000/16000 ref.wav deg.wav**

Infine, per la registrazione dei file audio è stato utilizzato il software Nu – tech<sup>[10]</sup>. Durante il settaggio delle impostazioni del software è stato utilizzato un campionamento a 44100 Hz. La scelta di impostare la frequenza a 44100 Hz è dovuta ai settaggi della scheda audio che accetta dei sample rate ben definiti e limitati tra cui non è presente la frequenza a 8000 Hz. Per poter registrare i file degradati tramite tale piattaforma è stato necessario ricampionare tutti i file audio originali a 44100 Hz e successivamente ricampionare i file audio degradati a 8000 Hz. Il codice utilizzato per il ricampionamento è così strutturato:

```
[old_rec, old_fs]=audioread('Cartella/audio.wav');
[P, Q] = rat(8000/old_fs);
new_rec = resample(old_rec, P, Q);
audiowrite('Nuovacartella/nuovoaudio.wav', new_rec, 8000);
```

Nella prima riga del codice si richiama la posizione del file audio a 44100 Hz, successivamente nella seconda riga si dichiara la frequenza a cui lo si vuole ricampionare. Nella terza riga è dichiarata la funzione del ricampionamento, e nell'ultima riga andiamo a dichiarare il percorso in cui vogliamo salvare il nuovo file

audio a 8000 Hz. La stessa procedura va eseguita per ricampionare i file da 8000 a 44100 Hz, andando a sostituire il valore della frequenza nella seconda e quarta riga.

## 5.2 Descrizione setup di misura

Per stimare la qualità sonora degli apparecchi acustici è stato utilizzato uno spazio di prova semianecoico con lo scopo di poter effettuare i test in condizioni controllate. La camera semianecoica è un locale appositamente progettato sia per annullare le riflessioni sonore interne, sia per evitare le interferenze provenienti dall'esterno. La camera in questione utilizzata in questi studi è una camera semianecoica, in quanto presenta un pavimento riflettente mentre tutte le altre pareti sono assorbenti. Per poter creare l'assenza di eco nella camera, si ricorre all'uso di pareti e soffitto realizzati con materiali altamente fonoassorbenti, modellati secondo specifiche forme appuntite cuneiformi e sporgenti, come possiamo vedere in figura 6, in modo che il suono venga assorbito e non venga riflesso dalle superfici.



*Figura 6: Fotografia della struttura di una tipica camera anecoica*

In figura 7 possiamo osservare una fotografia del setup messo a punto per realizzare i test sperimentali, ed è così costituito:

- manichino “Head and Torso Simulator” Bruel & Kjaer 4128C
- apparecchio acustico da testare
- cuffie auricolari
- scheda audio USB collegata ad un computer

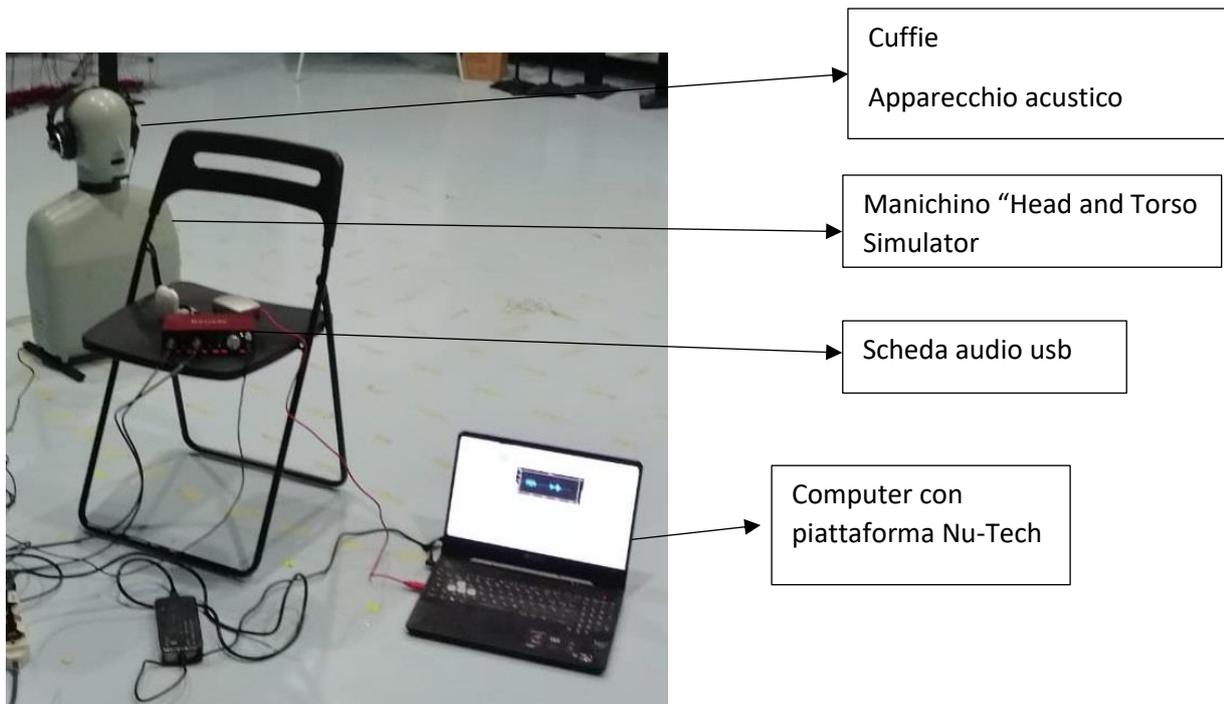


Figura 7: Fotografia rappresentante la catena di misura e i suoi elementi utilizzati per eseguire i test sperimentali

Il manichino in questione, come possiamo vedere in figura 8, il quale rappresenta una simulazione abbastanza realistica del sistema testa-torso di un soggetto adulto di medie dimensioni. Entrambe le orecchie sono composte da un padiglione di gomma siliconica, realizzato su impronta di orecchie umane, e da un canale auricolare che termina in un simulatore delle prestazioni acustiche da  $\frac{1}{2}$  pollice con un apposito adattatore per il preamplificatore. Con tale sistema si può misurare la pressione sonora in corrispondenza della membrana timpanica artificiale del manichino, andando a considerare l'effetto dovuto alla presenza del torso e della testa, oltre che al padiglione e condotto uditivo.

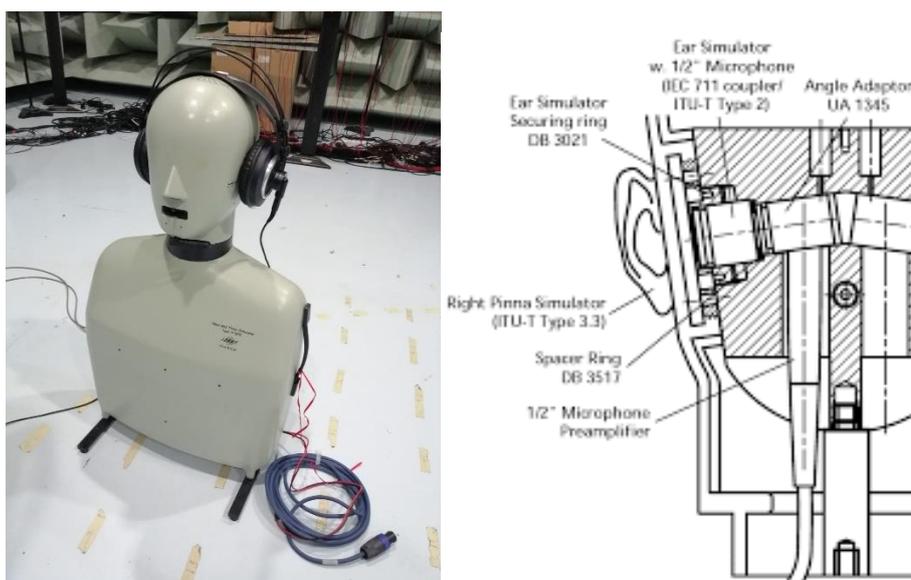


Figura 8: A sinistra il manichino Brüel Kjaer 4128C, a destra la sua struttura interna

Riportiamo ora i dati tecnici che lo caratterizzano:

- Risposta in frequenza dell'ascoltatore: Conforme a ITU-T Rec. P.58 fino a 16 kHz per le misurazioni sui dispositivi di telecomunicazione e secondo IEC60318-7 e ANSIS.3.36-1985 per le misurazioni su apparecchi acustici a conduzione d'aria
- Simulatore dell'orecchio: IEC 60318-4/ITU-T Rec. P.57 Simulatore di orecchio calibrato su tipo 3.3 conforme a ITU-T Rec. P.57, IEC 60318-4 e Ansi S3.21 standard. L'uscita dal simulatore dell'orecchio avviene tramite un cavo a 7 conduttori da 3 metri terminando con una spina Lemo.
- Sensibilità tipica:  $11.6 \text{ mV/Pa} = -38.7 \text{ dB}$
- Distribuzione pressione sonora: Conforme a ITU-T Rec. P.58
- Sensibilità tipica a 1 kHz:  $80 \text{ dB SPL } 2 \text{ V/500 mm}$
- Distorsione a 94 dB SPL:
  - <14% da 100 Hz fino a 200 Hz
  - <2% da 200Hz fino a 250 Hz
  - <1% oltre i 250 Hz
- Massima potenza di ingresso media: 10 W
- Massima potenza in ingresso pulsata: 50 W per 2 secondi

Per ogni sessione di test, un apparecchio acustico da testare è stato collocato a livello dell'orecchio destro. Le prove per la stima della qualità sonora degli apparecchi acustici retroauricolari sono effettuate su tre diverse tipologie di apparecchi.

Nella figura 9 si può notare il padiglione auricolare in gomma silconica di un orecchio del manichino prima (sinistra) e dopo (destra) l'applicazione di un apparecchio da utilizzare in fase di misura.

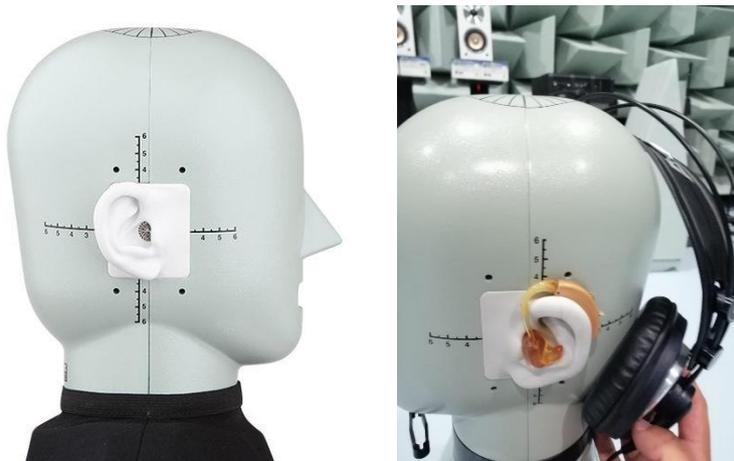


Figura 9

Il processo di misura consiste nella riproduzione di un file vocale originale, fornito insieme alla normativa, e inviato tramite cuffie al padiglione auricolare e percepito dall'apparecchio acustico. Tramite l'utilizzo della piattaforma Nu-tech è possibile

selezionare quale file vocale originale mandare al padiglione auricolare, il quale lo elaborerà tramite l'apparecchio acustico e lo registrerà in file wav che rappresenterà il segnale vocale degradato che verrà utilizzato come file di confronto dall'algoritmo per poter stimare il valore MOS. Tutte le misure sono state effettuate ad una frequenza di 44100 kHz.

La piattaforma Nu-tech è una piattaforma DSP, che sta per Digital Signal Processing, è utilizzata per la convalida e il debug di algoritmi complessi in tempo reale, sfruttando un comune PC. Essa è un processore dedicato e ottimizzato per poter eseguire in maniera efficiente sequenze di istruzioni ricorrenti. Nella figura 10 possiamo vedere come appare la nostra configurazione utilizzata durante la fase dei test sulla piattaforma Nu-tech.

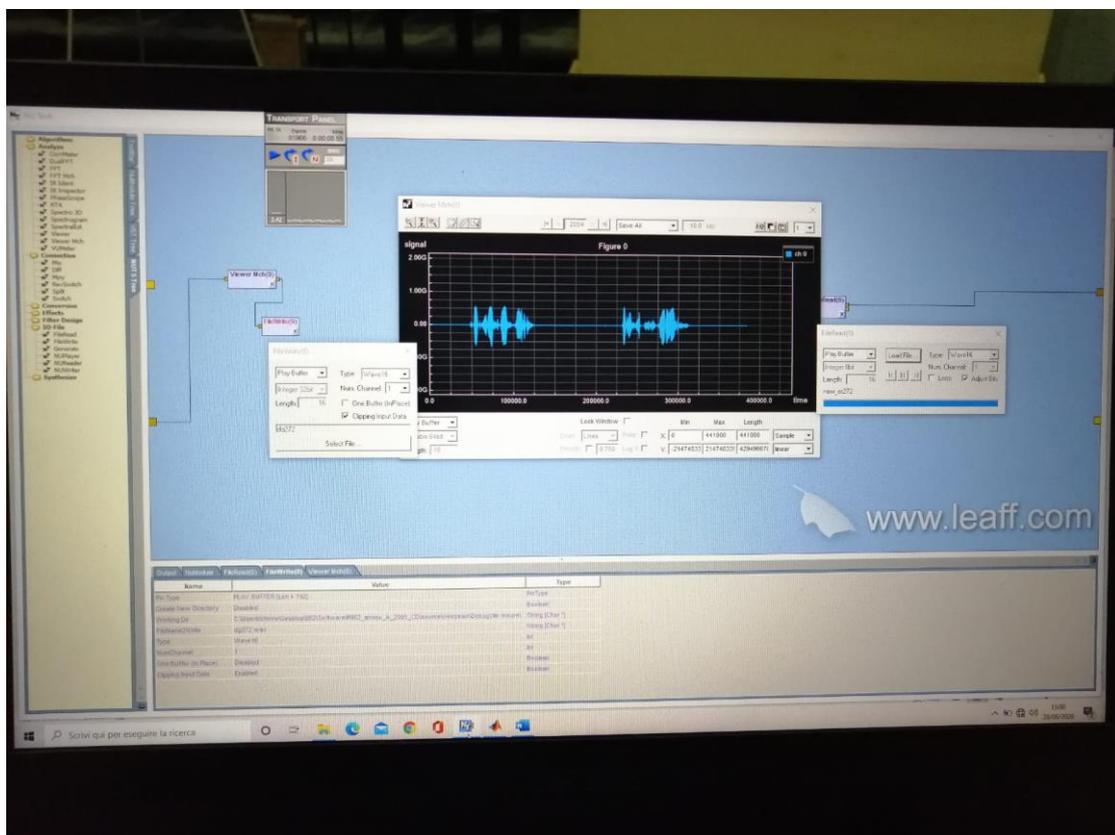


Figura 10: Schema della piattaforma Nu-tech utilizzato durante i test

Ovviamente, il collegamento tra il computer e il manichino per l'invio dei segnali è possibile per mezzo di una scheda audio usb, alla quale verranno collegati il manichino, le cuffie e il computer.

La scheda audio in questione è un a Focusrite Scarlett 2i2, ed è una scheda di espansione di un computer che si occupa dell'elaborazione di un flusso audio, in questo caso ottenuto tramite cuffie auricolare, per essere riprodotto in un segnale sonoro a favore dell'utente. Nella figura 11 viene mostrata la scheda audio utilizzata nei test sperimentali.



Figura11: Scheda audio utilizzata come interfaccia tra manichino e computer

Riportiamo ora i dati tecnici che la caratterizzano:

- Frequenze di campionamento supportate: 44,1 kHz, 48 kHz, 88,2 kHz, 96 kHz, 176,4 kHz, 192 kHz
- Ingresso del microfono con: -Risposta in frequenza: 20 Hz – 20 kHz  $\pm 0.1$  dB

-Gamma dinamica: 111 dB

-THD+ N <0.0012%

-Impedenza: 3 k $\Omega$

-Gamma del guadagno: 56 dB

-Livello di ingresso massimo: 9 dBu

Ingresso di linea con:

-Risposta in frequenza: 20 Hz – 20 kHz  $\pm 0.1$  dB

-Gamma dinamica: 110.5 dB

-TDH+ N <0.002%

-Impedenza < 60  $\Omega$

-Gamma del guadagno: 56 dB

-Livello di ingresso massimo: 22 dBu

Uscita delle cuffie con:

-Gamma dinamica: 104 dB

-THD+ N < 0.002%

-Impedenza: 4390  $\Omega$

-Livello di uscita massimo: 15 dBu

## 5.2 Discussione dei risultati

Messo a punto il setup sperimentale, si procede andando ad illustrare le prove effettuate sui tre diversi apparecchi acustici con l'obiettivo di stimare i parametri oggettivi per valutare la qualità sonora. L'algoritmo ci rilascerà due valori: RAW MOS e MOS\_LQO. Il RAW MOS rappresentano i valori grezzi, cioè si basano su una scala che va da -0,5 a 4,5, invece il MOS\_LQO rappresenta il valore che prenderemo in considerazione in quanto si basa su una scala che va da 1 a 5. Da qui verranno calcolati per ogni tabella la deviazione standard sia per il RAW MOS, che per il MOS\_LQO. Essa ci permetterà di vedere di quando i valori individuali possano differire dalla media, e la formula utilizzata è la seguente:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

ove  $n$  rappresenta la dimensione del campione e  $\bar{x}$  la media aritmetica del campione.

Prima di iniziare ad eseguire i test sugli apparecchi acustici è stato eseguito un test di misura a "vuoto", cioè senza apparecchi acustici sul manichino. Ciò ci permetterà di capire come risponde ai segnali vocali che manderemo in ingresso e per poter confrontarne la differenza con la percezione sonora in presenza degli apparecchi acustici.

Nella tabella 1 possiamo andare a vedere tutti i risultati MOS ottenuti tramite l'elaborazione dell'algoritmo dai file vocali ottenuti da tale misurazione.

File audio	RAW MOS	MOS-LQO
105	3,616	3,710
109	2,937	2,730
114	3,974	4,128
129	3,989	4,143
137	3,922	4,074
145	4,256	4,381
149	4,190	4,327
152	4,270	4,392
154	3,861	4,007
155	4,137	4,283
161	4,170	4,311
164	4,219	4,352
166	4,166	4,308

170	4,109	4,258
221	3,869	4,045
229	4,171	4,312
246	4,126	4,273
272	4,163	4,305

Dai dati ottenuti, abbiamo una media del RAW MOS pari a 4,008 e una deviazione standard pari a 0,316. Invece abbiamo una media del MOS\_LQO pari a 4,129 e una deviazione standard pari a 0,388 e possiamo stabilire che presenta una qualità oggettiva molto buona nonostante un valore basso nel file vocale 109.

Da qui andiamo a vedere i grafici dell'onda sonora, mostrati in figura 12 e 13, prendendo in considerazione il file vocale 109, che rappresenta una qualità bassa tendente al mediocre, e il file vocale 152, che rappresenta la qualità migliore tra tutti gli altri file vocali.

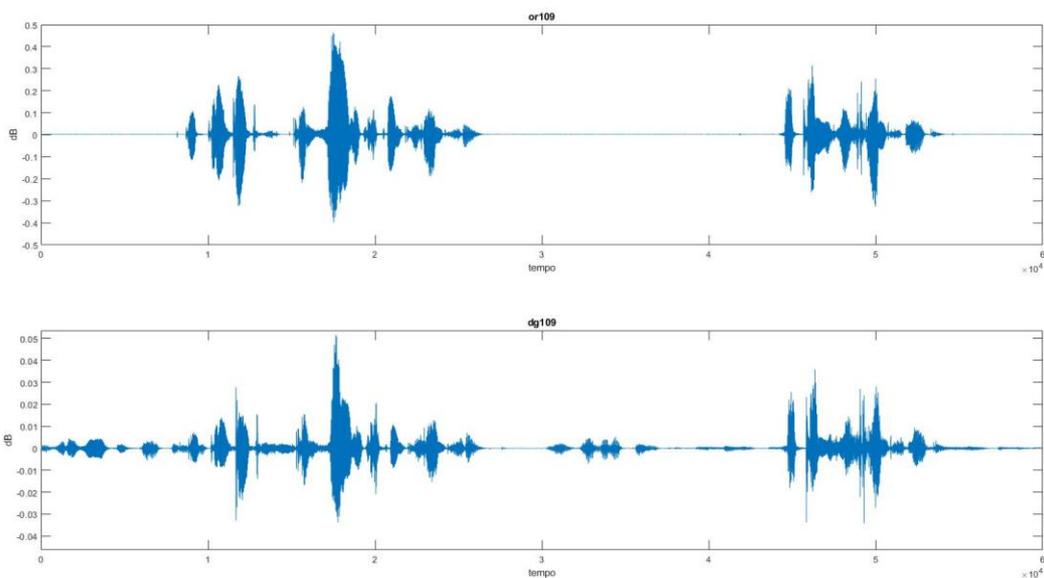


Figura 12: Grafico forma d'onda file vocale 109 originale(sopra) e degradato(sotto)

Questi due grafici rappresentano il confronto tra il file originale e il file “degradato” ottenuto dalla misura a vuoto con il manichino. Dal secondo grafico possiamo notare come l'onda del parlato attivo sia rimasta, con un'analisi qualitativa, simile a quella originale, ma si nota anche che l'onda del silenzio sia stata recepita come un disturbo che ha abbassato la qualità sonora del file audio che ha fatto sì che il valore MOS finale risultasse di 2,730.

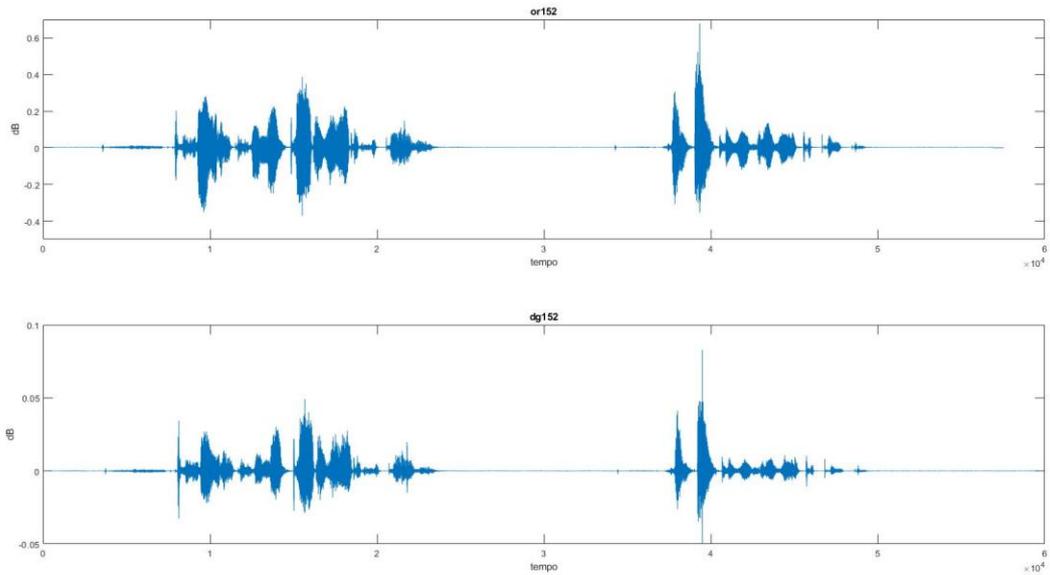


Figura 13: Grafico forma d'onda file vocale 152 originale(sopra) e degradato(sotto)

Da questi ultimi due grafici in confronto possiamo notare come la forma d'onda risulti, qualitativamente, simile a quella originale, tranne per qualche piccola riduzione della sua ampiezza in qualche piccola parte del parlato attivo, ma non ha influito sulla qualità oggettiva in quanto il suo risultato MOS è stato di 4,329.

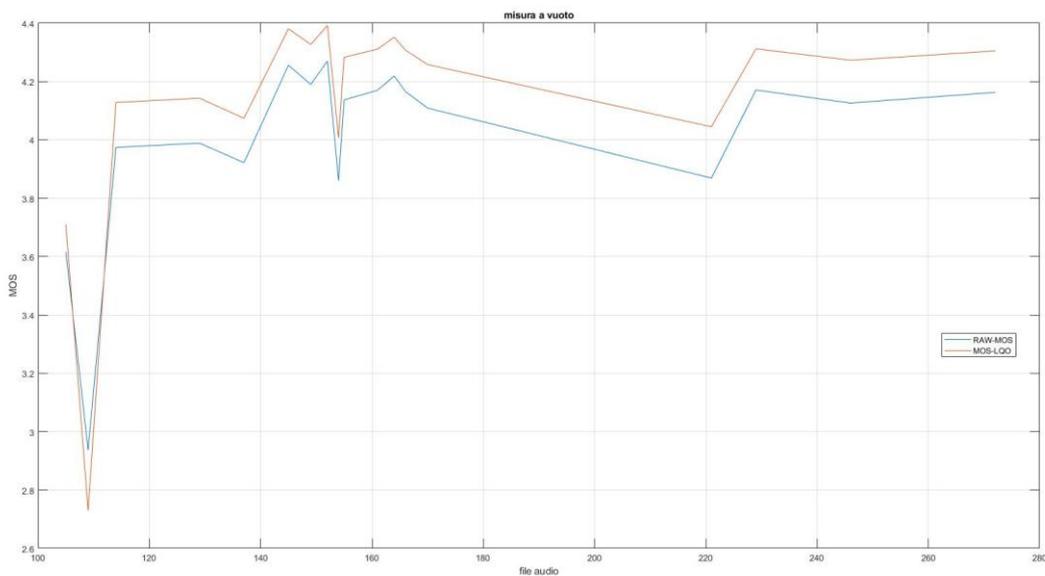


Figura14: Grafico valori MOS per file audio

Nella figura 14 possiamo notare l'andamento dei due grafici MOS per ogni file audio ottenuti nelle misure a vuoto. Possiamo confermare che la media dei valori MOS si attesta sul 4, rappresentando una buona qualità del parlato.

Dopo questo test si procede al test con un apparecchio acustico modello Beurer 641,06, illustrato in figura 15.

Riportiamo le sue caratteristiche elettroacustiche:

- Volume massimo: 128 dB
- Intervallo di frequenze: 200-5000 Hz
- Amplificazione massima: 40 dB



Figura 15

Tale dispositivo è considerato estremamente economico e le sue prestazioni possono essere considerate modeste in confronto con il prezzo.

Il dispositivo testato fornisce i seguenti valori MOS che riportiamo nella tabella 2.

File audio	RAW MOS	MOS-LQO
105	3,528	3,593
109	3,351	3,343
114	3,717	3,838
129	3,722	3,844
137	3,744	3,871
145	3,819	3,960
149	3,768	3,900
152	3,828	3,970
154	3,894	4,044
155	3,985	4,139
161	3,892	4,041
164	3,914	4,065
166	3,889	4,038
170	3,856	4,002
221	3,955	4,108
229	3,982	4,136
246	3,899	4,049

272	4,023	4,177
-----	-------	-------

Da questa tabella possiamo notare come le qualità sonore con l'apparecchio acustico siano leggermente diminuite rispetto alle misure a vuoto e questo è dovuto al fatto che l'apparecchio acustico non è altro che un amplificatore sonoro e ciò comporta una distorsione che influisce in ciò. Infatti, la media dei valori RAW MOS risulta 3,820 con una deviazione standard pari a 0,167. Invece, per i valori MOS\_LQO risulta 3,951 con una deviazione standard pari a 0,206.

Nella figura 16 e 17 possiamo vedere le forme d'onda dei file audio 109 e 272 che rappresentano rispettivamente i file con il valore MOS più basso e il più alto.

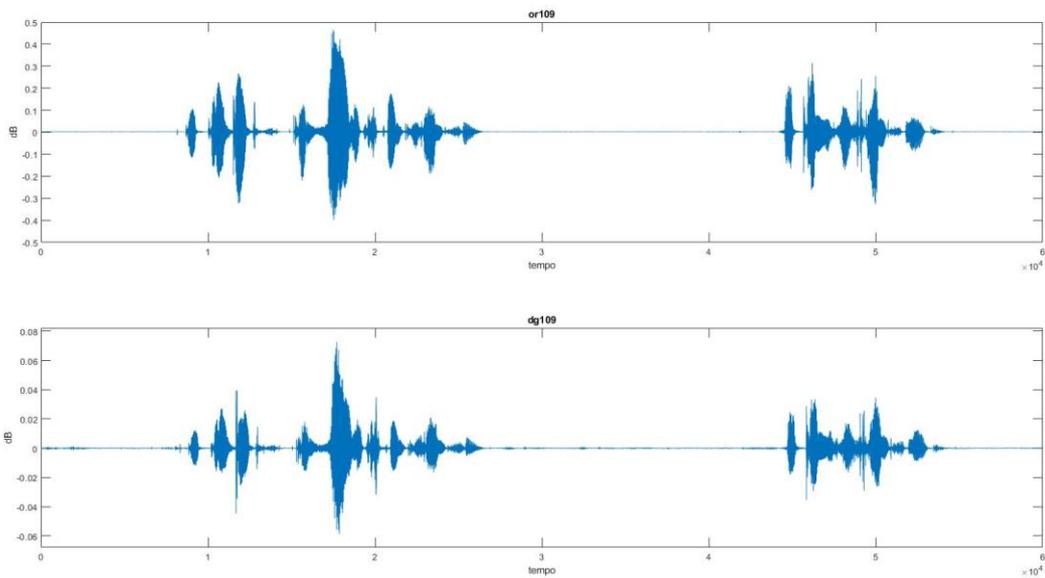


Figura 16: Forma d'onda del file vocale 109 originale(sopra) e degradato(sotto)

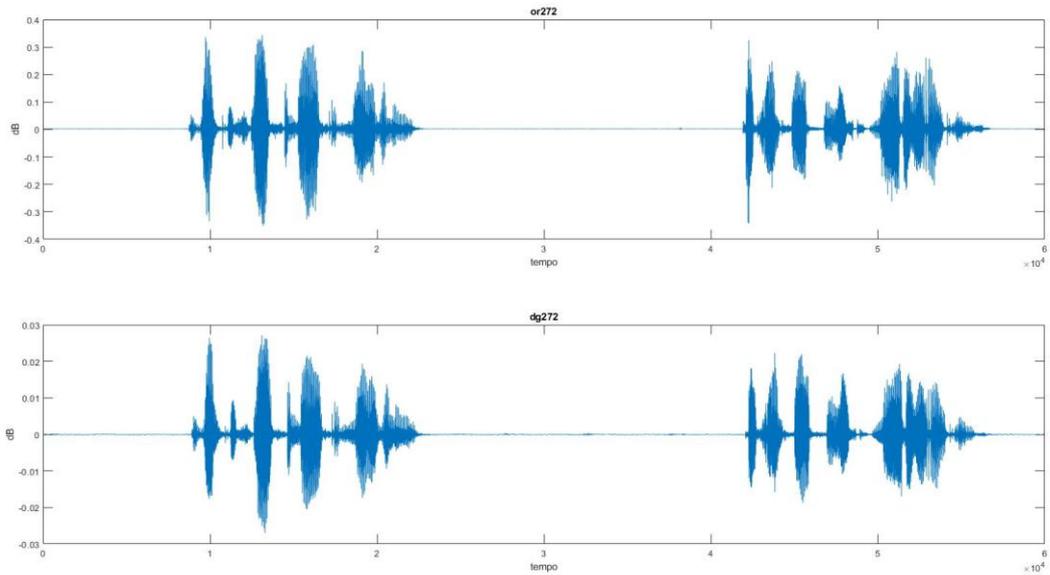


Figura 17: Forma d'onda del file vocale 272 originale(sopra) e degradato(sotto).

Nelle figure 16 e 17 possiamo notare come l'ausilio acustico abbia amplificato l'ampiezza della forma d'onda andando a mantenere una qualità sonora molto simile a quella ottenuta rispetto alle misure a vuoto.

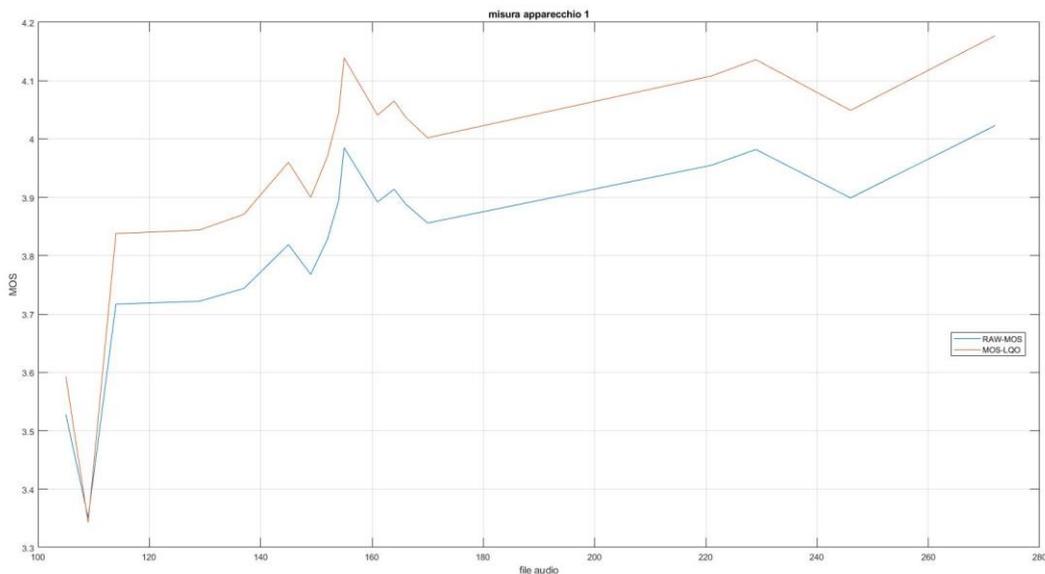


Figura 18: Grafico valori MOS per file audio del primo apparecchio acustico testato

Nella figura 18, possiamo notare l'andamento dei due grafici MOS per ogni file audio ottenuti nelle misure con il primo apparecchio acustico testato. Possiamo confermare che la media dei valori MOS si attesta sul 3,9, rappresentando una qualità del parlato molto vicina al buono.

Successivamente passiamo al test con il secondo apparecchio mostrato in figura 19. Si tratta dell'apparecchio acustico Acustika Melody. Anch'esso è un apparecchio estremamente economico come quello utilizzato nella misura precedente.



Figura 19

Elenchiamone ora le sue caratteristiche:

- Volume massimo: 128 dB
- Intervallo di frequenze: 200-5000 Hz
- Amplificazione massima 40 dB

Dopo averne ricavato i file audio e i relativi risultati MOS, li andiamo ad elencare nella tabella 3.

File audio	RAW MOS	MOS-LQO
105	3,188	3,103
109	2,847	2,599
114	3,401	3,425
129	3,392	3,402
137	3,394	3,511
145	3,498	3,511
149	3,347	3,337
152	3,058	2,908
154	3,325	3,305
155	3,316	3,293
161	3,174	3,082
164	3,264	3,216

166	3,357	3,351
170	3,526	3,590
221	3.358	3,354
229	3,585	3,670
246	3,095	2,964
272	3,449	3,348

In questa tabella, se confrontata con le altre due tabelle, possiamo notare come la qualità sonora si basa su una media del valore MOS di 3,2 che rappresenta una qualità mediocre. Questo è dovuto soprattutto alle distorsioni che produce un apparecchio acustico economico. Da questi risultati ottenuti, abbiamo una media dei valori RAW MOS pari a 3,300 e una deviazione standard 0,182. Invece, i per i valori MOS\_LQO abbiamo una media di 3,276 e una deviazione standard pari a 0,264.

Di seguito, nelle figure 20 e 21 possiamo andare a confrontare i grafici dei file audio originali e degradati prendendo in considerazione un file audio con il valore MOS più basso ottenuto e un file audio con un valore MOS più alto ottenuto.

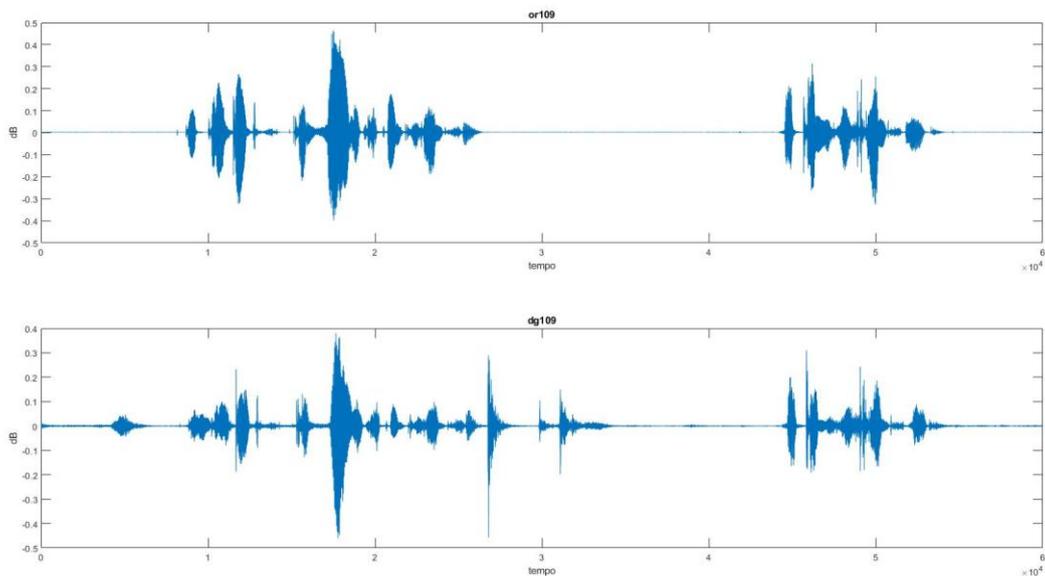


Figura 20: Forma d'onda file audio 109 originale(sopra) e degradato(sotto)

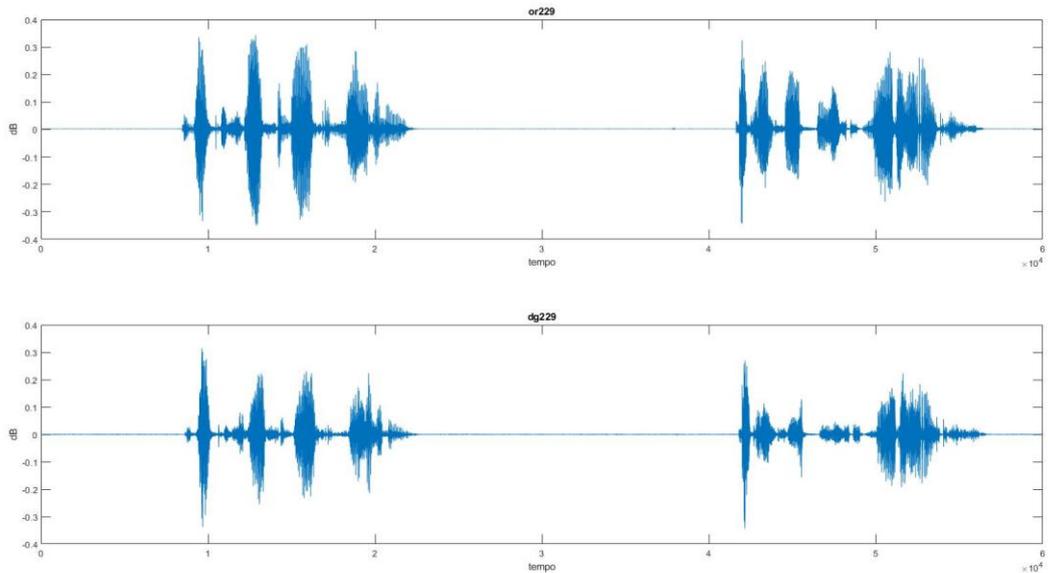


Figura 21: Forma d'onda file audio 229 originale(sopra) e degradato(sotto)

In figura 20 possiamo notare come le forme d'onda del file degradato siano più influenzate dal disturbo durante la fase di silenzio andando ad abbassare di 0,131 punti la qualità sonora rispetto alla qualità ottenuta nella misura a vuoto. A differenza, nella figura 21 possiamo notare come le forme d'onda subiscono delle compressioni di ampiezza durante la fase del parlato attivo andando ad influenzare la qualità complessiva del file audio ottenendo un valore MOS pari a 3,670 che comincia ad essere considerato accettabile.

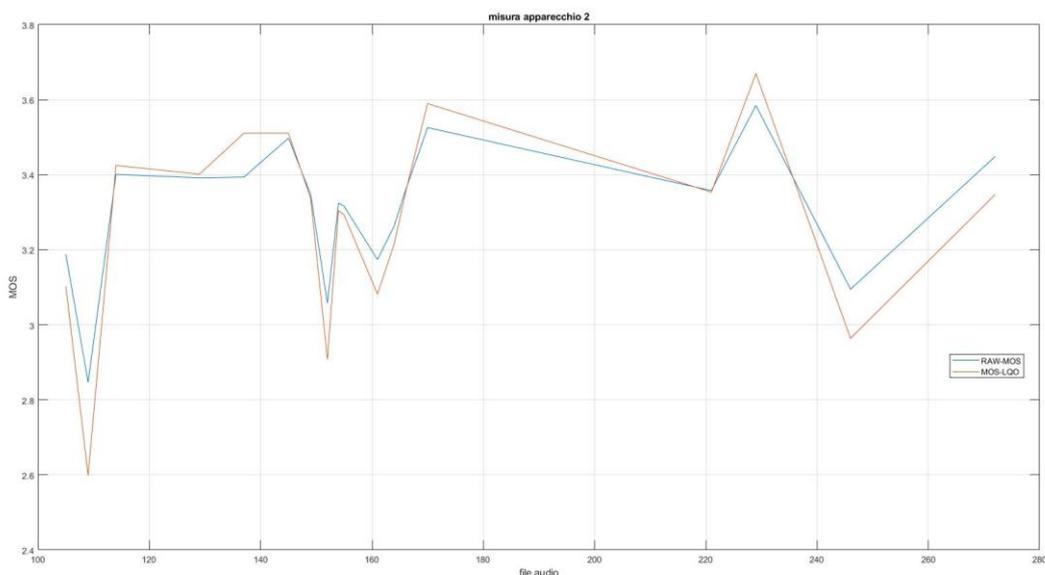


Figura 22: Grafico valori MOS per file audio del secondo apparecchio testato

Nella figura 22 possiamo notare l'andamento dei due grafici MOS per ogni file audio ottenuti nelle misure del secondo apparecchio testato. Possiamo confermare che la media dei valori MOS si attesta sul 3,2, rappresentando una mediocre qualità del parlato.

Per concludere andiamo ad effettuare i test di misura tramite l'ultimo apparecchio acustico. Si tratta di un modello Hansaton 42 C-WR (figura 23), che è considerato un apparecchio professionale ed è usato in casi di ipoacusia molto gravi. Rispetto ai modelli precedenti, questo apparecchio presenta dei valori di distorsione complessivamente inferiori.



Figura 23: Apparecchio Hansaton 42 C-WR

Nella tabella 3 vengono elencati tutti i valori MOS ottenuti dall'elaborazione, tramite algoritmo, dei file vocali ottenuti tramite esso.

File audio	RAW MOS	MOS-LQO
105	3,654	3,759
109	3,851	3,996
114	3,676	3,787
129	3,615	3,709
137	3,610	3,702
145	3,796	3,933
149	3,874	4,021
152	4,018	4,172

154	3,651	3,756
155	3,739	3,865
161	3,921	4,073
164	4,222	4,354
166	3,800	3,938
170	3,759	3,889
221	3,961	4,114
229	3,704	3,822
246	3,863	4,009
272	3,780	3,914

Da questa tabella, in confronto con le altre tabelle, la prima cosa che notiamo è che i valori MOS quasi tutti intorno al valore 4. La seconda cosa che possiamo notare riguarda appunto il file audio 109, che a confronto, notiamo come la qualità sia migliorata tantissimo, arrivano a un valore MOS di 3,996, grazie ai valori di distorsione molto più bassi. Infine, da questi risultati, abbiamo una media per i valori RAW MOS pari a 3,800 e una deviazione standard di 0,157. Infine, per i valori MOS\_LQO abbiamo una media di 3,935 e una sua deviazione standard pari a 0,171.

Nelle figure 24 e 25 possiamo vedere i grafici delle forme d'onda del file audio, originale e degradato, 137 con un valore MOS di 3,702 e del file audio, originale e degradato, 152 con un valore MOS di 4,172.

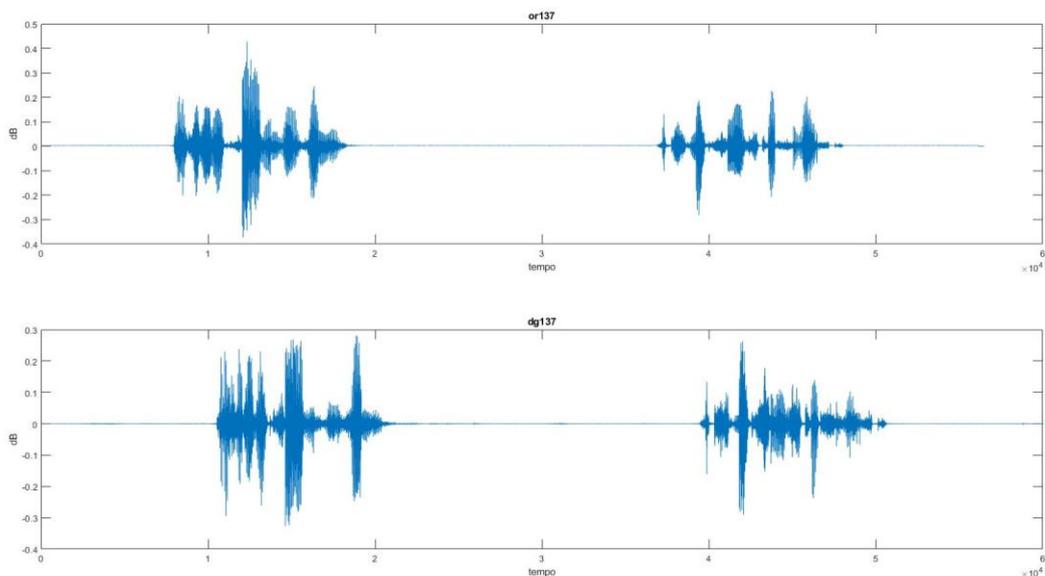


Figura 24: Forma d'onda file audio 137 originale(sopra) e degradato(sotto)

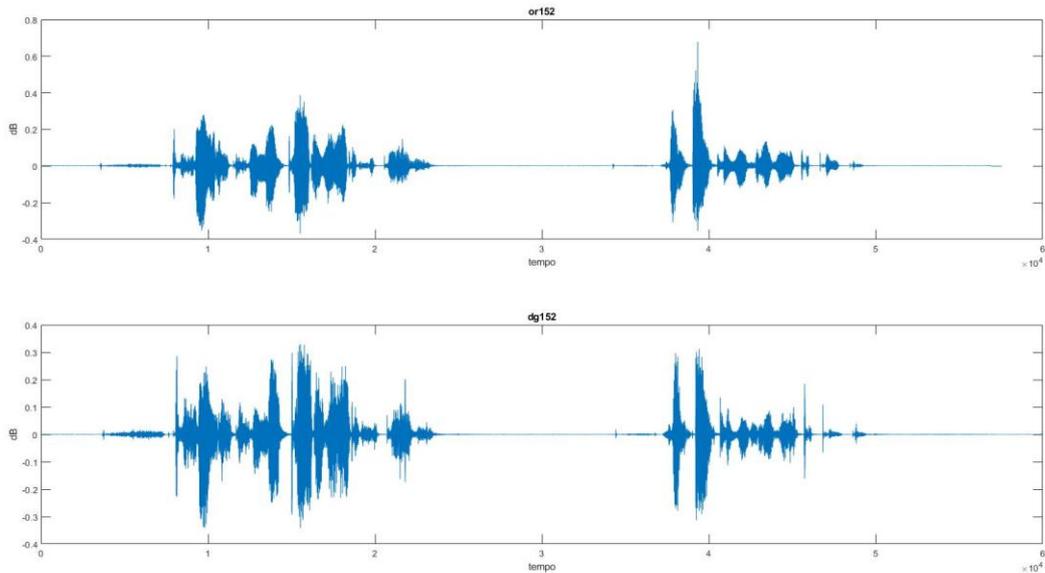


Figura 25: Forma d'onda file audio 152 originale(sopra) e degradato(sotto)

Nelle due figure possiamo notare come le forme d'onda dei file audio degradati siano amplificate di ampiezza, ma grazie ai valori di distorsione molto bassi, non abbiamo disturbi che ci permettono di abbassare la qualità sonora di questa protesi acustica.

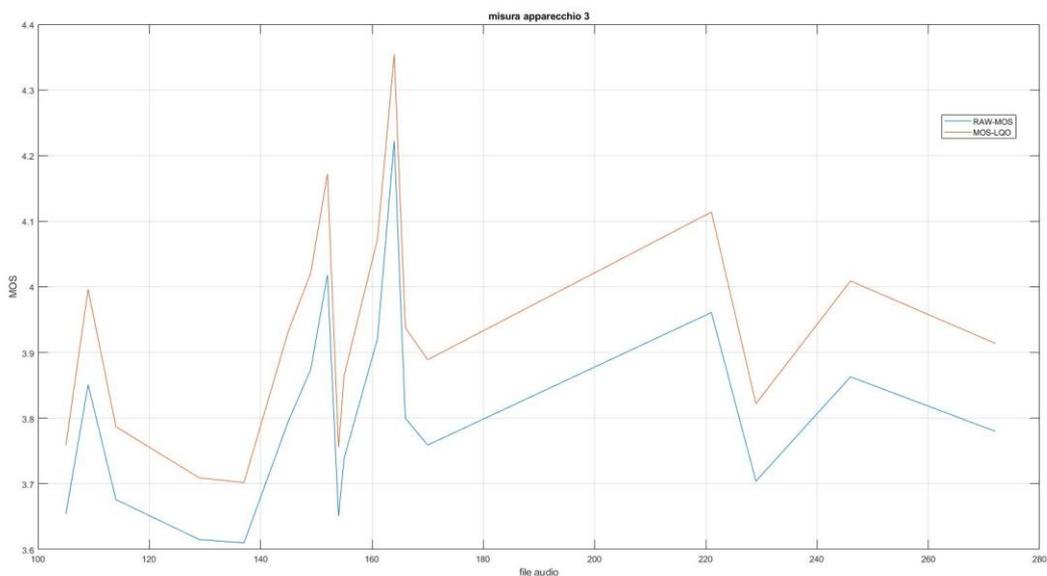


Figura26: Grafico valori MOS del terzo apparecchio acustico testato

Nella figura 26 possiamo notare l'andamento dei due grafici MOS per ogni file audio ottenuti nelle misure del terzo apparecchio acustico testato. Possiamo confermare che la media dei valori MOS si attesta sul 3,9, rappresentando una qualità del parlato molto vicina, anch'esso come il primo apparecchio testato, al buono.

Nella tabella 4 confrontiamo tutte le medie dei valori ottenuti con la loro annessa deviazione standard.

Misura	RAW MOS	Deviazione standard	MOS_LQO	Deviazione standard
Misure a vuoto	4,008	0,316	4,129	0,388
Misure apparecchio 1	3,820	0,167	3,951	0,206
Misure apparecchio 2	3,300	0,182	3,276	0,264
Misure apparecchio 2	3,800	0,157	3,935	0,171

Per concludere andiamo a mettere a confronto due file audio con tutte le risposte ottenute sia dalle misure a vuoto che con gli apparecchi acustici. Andiamo a confrontare il file audio 109, che nelle misure a vuoto è risultato essere quello con una scarsa qualità audio, e il file audio 152, che, a differenza del file audio 109, è risultato essere quello con la migliore qualità (figure 27 e 28).

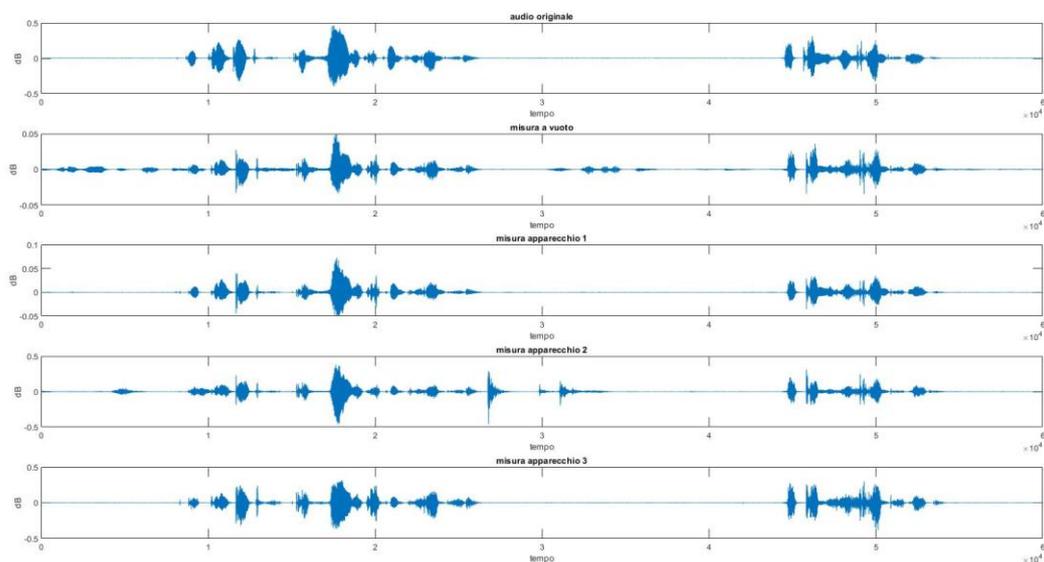


Figura 27: Confronto del file audio 109 con tutti i test di misura effettuati

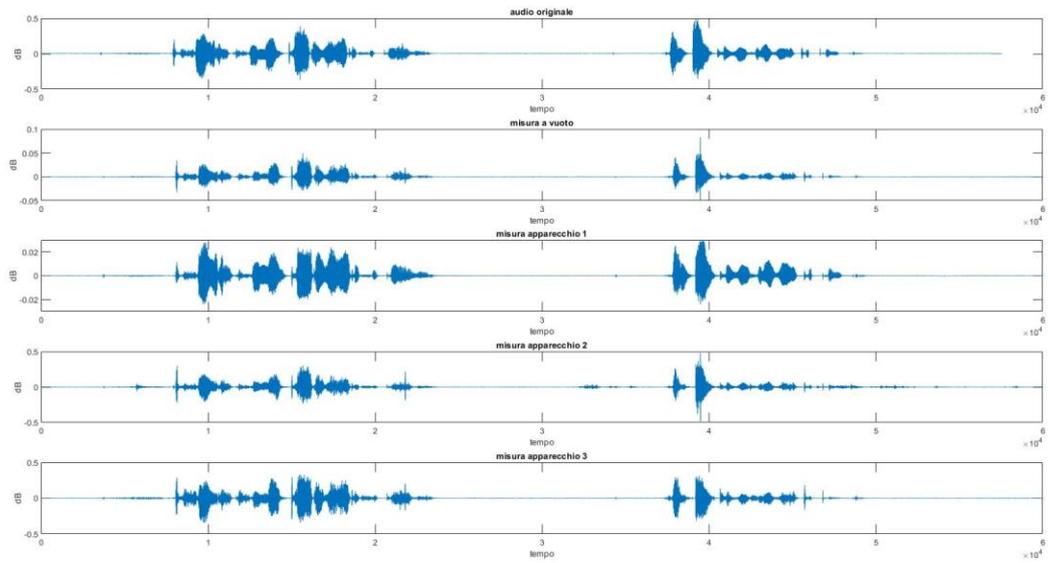


Figura 28: Confronto file audio 152 con tutti i test di misura effettuati

Nella figura 27 possiamo vedere a confronto le forme d'onda dell'audio originale, della misura a vuoto e delle misure con i tre apparecchi e si possono notare i tre livelli di amplificazione delle ampiezze d'onda ottenute da tali ausili e come hanno influito sul valore della qualità sonora di essi. La stessa cosa possiamo notarla nella figura 28.

# Conclusioni

Questo studio ha cercato di valutare i parametri oggettivi per la valutazione della qualità oggettiva degli apparecchi acustici, i quali potranno essere utili per stimare la qualità sonora dei segnali vocali con essi. A tale scopo, è stato condotto uno studio delle tecniche per la stima delle valutazioni oggettivi presenti in letteratura diffuse per la valutazione dei codec vocali in situazioni normali.

L'obiettivo principale di questo elaborato è stato lo studio di tecniche reali e simulate della previsione della qualità acustica tramite gli ausili acustici. A tal scopo sono state presentate delle misure innovative che utilizzano l'uso di un manichino "Head and Torso Simulator", che costituisce una realistica simulazione di un sistema testa-torso di un soggetto adulto di medie dimensioni, anche grazie all'orecchio artificiale realizzato in calco umano. Rispetto alle misure che sono state proposte nella normativa ITU-T P.862, il setup di misura attuato permette, tramite l'uso del manichino, di riprodurre in maniera più confidenziale le percezioni acustiche e le caratteristiche del padiglione auricolare umano.

Per poter stimare tali qualità sonore, sono stati eseguiti dei test sperimentali su quattro ausili auditivi con caratteristiche elettroacustiche differenti. Tramite questi test, realizzati utilizzando l'algoritmo PESQ per ottenere il valore della qualità acustica, ci ha permesso di stimare la qualità sonora per ogni apparecchio acustico, andando ad osservare come rispondono in base alle loro caratteristiche.

# Bibliografia

- [1] International Telecommunication Union, ITU-T P.800.1, 07/2016
- [2] International Telecommunication Union, ITU-T P.800.2, 07/2016
- [3] International Telecommunication Union, ITU-T P.830, 02/1996
- [4] International Telecommunication Union, ITU-T P.862, 02/2001
- [5][https://www.youtube.com/watch?v=BEEViqD44iA&t=39s&ab\\_channel=Audiope dia](https://www.youtube.com/watch?v=BEEViqD44iA&t=39s&ab_channel=Audiope dia)
- [6][https://www.youtube.com/watch?v=pj3DBj2FSlw&ab\\_channel=HEADacousticsI nternational](https://www.youtube.com/watch?v=pj3DBj2FSlw&ab_channel=HEADacousticsI nternational)
- [7] Psychoacoustic Audio Quality Measurements Using R&S UPV audio analyzer, Rohde&Schwarz
- [8]<https://focusrite.com/en/usb-audio-interface/scarlett/scarlett-2i2>
- [9]<https://it.mathworks.com/products/matlab.html>
- [10]<https://www.leaff.it/content.php?id=31>
- [11] International Telecommunication Union, ITU-T P.800, 08/1996
- [12] <https://www.bksv.com/-/media/literature/Product-Data/bp0521.ashx>
- [13] [https://ccrma.stanford.edu/~jos/bbt/Bark\\_Frequency\\_Scale.html](https://ccrma.stanford.edu/~jos/bbt/Bark_Frequency_Scale.html)
- [14] <https://www.britannica.com/science/son e>
- [15] <http://www.polqa.info/>
- [16] International Telecommunication Union, ITU-T P.861, 08/1996
- [17] Overview on VoIP: Subjective and Objective Measurement Methods, IJCSNS, VOL.6 No.1B, 01/2006
- [18] Objective estimation of perceived speech quality-part 1: Development of measuring normalizing block technique, IEEE transaction on speech and audio processing, Vol.7, No.4, 07/1999