

# **Appunti per il Progetto “SPEECON”**

Unità Operativa di Pisa

Lorenzo Cioni

Laboratorio di Linguistica, Scuola Normale Superiore, Pisa

## ***Introduzione***

Il presente documento rappresenta una sintesi dei documenti “SPEECON Deliverable D212”, “SPEECON Deliverable D33”, “Recording\_Instructions.add v1.1” e “SONY Speech Recording Studio” ed è stato concepito come un ausilio per le persone che effettueranno le registrazioni nelle località di Firenze e Milano nell’ambito del Progetto “SPEECON” e che saranno coordinate e supervisionate dall’autore. Date queste premesse il presente documento non è autocontenuto e per ogni carenza/chiarimento il lettore dovrà fare riferimento ai testi citati in questa introduzione.

## ***1 “SPEECON Deliverable D212”***

Il documento contiene indicazioni circa gli scenari delle registrazioni da effettuare nell’ambito del progetto, indicazioni basate su considerazioni circa le applicazioni di interesse e sugli ambienti di tali applicazioni. Ciò si traduce in condizioni ambientali che influenzano sia l’esecuzione delle registrazioni sia la scelta degli strumenti.

Il documento parte dalla definizione di “applicazioni di interesse” e di “ambienti acustici”.

Le prime non presentano interesse per chi dovrà effettuare materialmente le acquisizioni e, pertanto, per una loro descrizione si rimanda al documento. I secondi saranno ampiamente descritti nel resto della sezione.

Gli ambienti acustici sono suddivisi in due categorie denominate “Home” e “Mobile”.

All’interno della prima categoria troviamo gli ambienti “Ufficio” “Intrattenimento, bambini” e “Casa” (household) mentre all’interno della seconda troviamo “Auto”, “Luoghi pubblici all’aperto”, “Luoghi pubblici all’interno” e “Trasporti pubblici”.

Il passo successivo è quello di tradurre gli ambienti acustici in ambienti detti di registrazione, in pratica nei posizionamenti dei microfoni con le relative tipologie.

E’ infatti pensabile raggruppare gli ambienti acustici in base alle posizioni attese dei microfoni nelle varie applicazioni in modo da definire due soli settaggi: home e mobile.

### ***1.1 Home***

L’ambiente o settaggio “home” è concepito per coprire le applicazioni in ambito ufficio e domestico (household).

Gli utenti sono sia adulti sia bambini e il rumore di fondo (dovuto sia ai bambini stessi sia ad apparecchiature come una TV) è relativamente basso.

Per gestire l’ambiente “home” si individuano tre posizioni microfoniche:

- 1) microfono in posizione “close distance” ovvero a fianco della bocca (headset) o subito sotto il mento (lavalier o da cravatta);
- 2) microfono in posizione “medium distance” ovvero su un tavolo ad una distanza compresa fra mezzo metro e un metro dal parlante e
- 3) microfono in posizione “far distance” ovvero a una distanza dal parlante compresa fra 2 e 3 metri.

In merito alle condizioni di registrazione e al livello di rumore previsto si possono individuare per l’ambiente “home” tre sottocasi:

- ufficio,
- intrattenimento e
- bambini.

L’ufficio è un ambiente più o meno tranquillo caratterizzato da un rumore di fondo di tipo stazionario e con le caratteristiche del rumore bianco. Per i microfoni in posizione 2) e 3) è probabile la presenza di riverbero, dovuto alle caratteristiche fisiche dell’ambiente di registrazione, mentre le registrazioni fatte con il microfono in posizione 1) dovrebbero essere pulite (rapporto segnale/rumore molto buono e riverbero molto basso).

L’ambiente intrattenimento è più rumoroso (e il rumore può non essere stazionario ed essere “colorato” ovvero può essere di tipo impulsivo e avere un contenuto in frequenza non diffuso su tutto lo spettro) del precedente mentre il riverbero è dello stesso ordine di grandezza.

Nel caso dei bambini (che comunque non verranno presi in considerazione nel seguito se non per brevi cenni) si hanno difficoltà di tipo logistico, legale e pratico per le quali si rimanda al documento.

## *1.2 Mobile*

L’ambiente “mobile” nelle intenzioni copre applicazioni sia in auto (che comunque non verranno prese in considerazione nel seguito se non per brevi cenni) sia in luoghi pubblici (sia all’aperto sia al chiuso).

Gli utenti sono adulti o bambini che agiscono in presenza di un livello atteso di rumore di fondo relativamente alto (parlottio di persone, rumore di traffico o rumore aerodinamico nel caso dell’auto, ovviamente in movimento).

In questo ambiente il microfono in posizione 3) è inutile a causa del cattivo rapporto segnale/rumore atteso in tale ambiente per cui si mantengono solo le due posizioni microfoniche seguenti:

- 1) microfono in posizione “close distance” ovvero a fianco della bocca o subito sotto il mento e
- 2) microfono in posizione “medium distance” ovvero su un tavolo ad una distanza compresa fra mezzo metro e un metro dal parlante.

In merito alle condizioni di registrazione e al livello di rumore previsto si possono individuare per l’ambiente “mobile” due sottocasi:

- auto e
- luoghi pubblici.

Nel caso dell'auto si assume un riverbero trascurabile ma un livello di rumore da medio ad alto di tipo sia stazionario (rumore aerodinamico) sia impulsivo (clacson, tergicristallo) sia di tipo bianco sia colorato.

Nel caso delle registrazioni in luoghi pubblici (sia al chiuso sia all'aperto) è praticamente impossibile fare predizioni sulle caratteristiche acustiche dell'ambiente. L'unica caratteristica predicibile è il riverbero che può essere molto forte mentre per quanto riguarda il rumore si rimanda (con le dovute cautele) alle considerazioni svolte per il caso dell'auto.

### *1.3 Ambienti di registrazione e numero di parlanti*

Si rimanda al punto 3 del presente documento (per considerazioni pratiche e tabelle) e al punto 2.2.3 del documento SPEECON Deliverable D212 (per considerazioni di tipo teorico).

### *1.4 Microfoni*

Si esaminano brevemente i tipi dei microfoni, le loro posizioni durante le registrazioni e in particolare le misure che devono essere effettuate in modo da garantire che il segnale abbia le caratteristiche richieste

Per i tipi dei microfoni e le figure citate si rimanda al documento "SPEECON Deliverable D212" (in particolare le figure delle pagine 11, 12, 14 e 15).

Si fa notare che si suppone che i parlanti stiano seduti su una sedia durante le sessioni di registrazione per cui i microfoni "medium distance" e "far distance" devono essere montati ad un'altezza media di circa 1.2m con una deviazione consentita di  $\pm 0.5m$ .

La definizione delle aree di montaggio (vedi oltre) e delle altezze di montaggio va considerata operativamente nel senso che non è necessario misurare l'esatta posizione dei microfoni ma è sufficiente fare una stima di tali posizioni in modo che ricadano nei limiti indicati anche in considerazione del fatto che una violazione non eccessiva dei limiti non si riflette in modo significativo sulla qualità dei segnali acquisiti.

#### *Montaggio "close distance".*

In questo caso si usano microfoni di tipo "headset" e di tipo "lavalier" o "da cravatta". Il primo simula con buona approssimazione la posizione del microfono dei cellulari e consente di raccogliere registrazioni "pulite" mentre il secondo simula tutte le situazioni di chiamata con mani libere.

Il microfono di tipo "headset" deve essere montato al livello della bocca a una distanza di 1 o 2 centimetri da un angolo di questa mentre il lato della testa su cui è montato è irrilevante. La posizione è importante in modo da evitare la presenza di rumori impulsivi nei segnali registrati da tale microfono. Allo stesso scopo il microfono dovrebbe essere schermato con un cappuccio di materiale spugnoso.

Il microfono di tipo “lavalier” deve essere montato in modo da rimanere in una posizione appena sotto il mento (fra la spalla e il mento) che consente di classificarlo come microfono “close distance”.

Il microfono di tipo “lavalier” è composto da una pinza per il fissaggio ai vestiti, un microfono vero e proprio e un auricolare che non ha nessuna funzione ma che deve essere messo in uno dei due orecchi del parlante.

#### *Montaggio “medium distance”.*

Si fa uso di microfoni direzionali (tranne che nel caso dei bambini per i quali si usano microfoni omnidirezionali) montati ad una distanza compresa fra 0.5m e 1m dal parlante in modo che non ci siano ostacoli fra il parlante e il microfono (quali lo schermo di un portatile, vedi oltre) ma in modo che il microfono non sia in posizione troppo laterale rispetto al parlante.

Considerando un asse che parte dalla bocca del parlante e che individua la direzione dello speech si individua un’area rettangolare all’interno della quale è possibile montare il microfono “medium distance”.

Tale area è delimitata da due rette perpendicolari alla direzione dello speech a distanze rispettivamente di 0.5m e 1m dal parlante e da due rette parallele alla direzione dello speech distanti da questa 0.5m a destra e 0.5m a sinistra. All’interno di tale area è possibile montare il microfono “medium distance” in una posizione tale che non ci sia nessun ostacolo fra la bocca del parlante e il microfono.

#### *Montaggio “far distance”.*

Si fa uso di microfoni in tale posizione solo negli ambienti “ufficio”, “intrattenimento” e “bambini” e dal momento che, nelle applicazioni che fanno riferimento a tali ambienti, la posizione del parlante rispetto al microfono “far distance” non è predicibile si fa uso di microfoni omnidirezionali.

Come nel caso precedente, considerando un asse che parte dalla bocca del parlante e che individua la direzione dello speech si individua un’area rettangolare all’interno della quale è possibile montare il microfono “far distance”.

Tale area è delimitata da due rette perpendicolari alla direzione dello speech a distanze rispettivamente di 2m e 3m dal parlante e da due rette parallele alla direzione dello speech distanti da questa 0.5m a destra e 0.5m a sinistra. All’interno di tale area è possibile montare il microfono “far distance” in una posizione tale che non ci sia nessun ostacolo fra la bocca del parlante e il microfono.

#### *Montaggio per le misure.*

E’ necessario, per gli scopi del progetto, avere una accurata e completa conoscenza delle reali e possibili condizioni di registrazione per cui è necessario caratterizzare i reali ambienti di registrazione in termini di:

- risposta impulsiva della stanza e

- livello medio di rumore.

Per misurare la risposta impulsiva dell'ambiente si fa uso di una tecnica di "cross-correlation" (correlazione incrociata o mutua).

In base a tale tecnica un piccolo altoparlante a banda larga (che riproduce un ampio spettro di frequenze) viene messo al posto del parlante e gli viene fatto riprodurre un segnale di riferimento caratterizzato da "rumore rosa" (concentrato alle basse frequenze).

Il segnale viene catturato utilizzando due microfoni (identici, di alta qualità e omnidirezionali) simultaneamente: un microfono posizionato proprio davanti all'altoparlante ad una distanza compresa fra 2 e 5 cm e un altro microfono nella posizione reale corrispondente alla "medium distance" o "far distance".

Il calcolo della risposta impulsiva verrà fatto utilizzando i segnali registrati mediante i due microfoni.

In tutti i casi in cui ha significato il montaggio "far distance" si devono fare due misure:

- una con il primo microfono montato in prossimità dell'altoparlante e l'altro montato "medium distance" e
- una con il primo microfono montato in prossimità dell'altoparlante e l'altro montato "far distance".

Nel caso dei luoghi pubblici dato che non si definisce il montaggio "far distance" si ha solo la misura con il primo microfono montato in prossimità dell'altoparlante e l'altro montato "medium distance" mentre nel caso dell'auto non si esegue nessun tipo di misura.

Ulteriori dettagli in merito alla misura della risposta impulsiva saranno dati al punto 1.6 insieme ad indicazioni su come misurare il livello medio di rumore.

### *1.5 Specifica dettagliata delle condizioni e degli scenari di registrazione*

#### ***Specifiche delle etichette delle condizioni***

All'interno di ciascuno degli ambienti di registrazione è possibile immaginare una grande varietà di situazioni reali di registrazione sebbene, da un punto di vista pratico, sarebbe preferibile modificare la disposizione degli strumenti di registrazione il meno possibile. Ciò porterebbe ad avere un database piuttosto omogeneo con poca variabilità relativamente alle caratteristiche di rumore e di riverbero degli ambienti di registrazione. Tali caratteristiche sarebbero in contrasto con gli scopi del progetto che sono quelli di analizzare il riconoscimento del parlato in ambienti rumorosi e riverberanti, scopi che richiedono un data base caratterizzato da una grande varietà di influenze ambientali (rumore e riverbero).

Per soddisfare tali requisiti ciascuno scenario di registrazione è suddiviso in più categorie ciascuna delle quali è individuata da una combinazione delle cosiddette "etichette di condizione". Tali etichette forzano le persone che fanno acquisizione a

registrare un certo numero di parlanti per categoria in modo da garantire una elevata variabilità del materiale registrato per quanto riguarda le caratteristiche di riverbero e di rumore degli ambienti.

Tali etichette oltre a specificare in modo univoco le reali condizioni di registrazione vengono usate anche per scopi di documentazione.

### *Le etichette di luogo.*

Le etichette di luogo servono per caratterizzare il tipo di ambiente di registrazione e per distinguere fra loro ambienti di registrazione della stessa categoria, ad esempio diversi uffici.

Esempi di etichette di luogo per l'ufficio sono:

OFFICE\_01, OFFICE\_02 e simili per l'ufficio,

ENTERTAIN\_01, ENTERTAIN\_02 e simili per l'intrattenimento

PUBHALL\_01, PUBHALL\_02" e simili per i luoghi pubblici al chiuso.

Ad esempio ENTERTAIN individua l'intrattenimento da cui ENTERTAIN\_01 e simili, PUBHALL luoghi pubblici al chiuso caratterizzati come PUBHALL\_01 e simili, PUBOPEN luoghi pubblici all'aperto e così via, per ulteriori dettagli si rimanda alla tabella a pagina 16 del documento più volte citato.

### *Le etichette di posizione.*

In merito alle caratteristiche di riverbero dei diversi luoghi è importante conoscere la posizione del parlante rispetto a superfici riflettenti quali le pareti. Le etichette di posizione sono utilizzate per caratterizzare tale posizione in modo generico in base alla distanza del parlante dal muro più vicino.

Si parte definendo una striscia virtuale che costeggia i muri circostanti la posizione di registrazione (figura 7 pagina 17) in modo che se il parlante è all'interno della striscia la sua posizione è detta CLOSE\_WALL mentre se è all'esterno la si dice "FAR\_WALL". La striscia ha un'ampiezza di 2m nel caso dei luoghi pubblici e di 1m nel caso di ufficio, intrattenimento e bambini mentre non è definita nel caso dell'auto.

L'ampiezza della striscia deve essere tale da consentire al riverbero di esercitare un qualche influsso ma non così ampia da creare problemi nell'individuazione di posizioni "FAR\_WALL" negli ambienti reali da cui la limitazione di 1m nel caso degli ambienti suddetti.

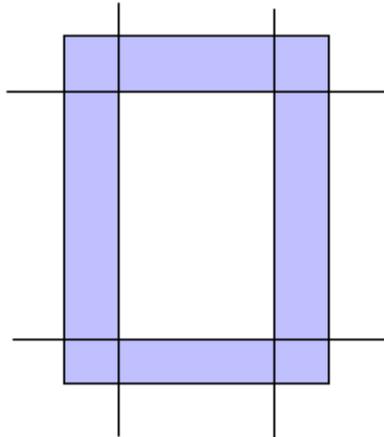
Nel caso di luoghi pubblici (all'aperto) se il muro più vicino dista più di 10m si definisce la posizione NO\_WALL caratterizzata da scarsi se non nulli effetti di riverbero.

Nel caso dell'auto si ha la posizione CO\_DRIVER.

A ciascuna posizione corrisponde un'etichetta composta dalla "posizione" seguita dal carattere "\_" e da un numero d'ordine su due cifre per cui esempi di etichette legali sono i seguenti:

CLOSE\_WALL01 e simili  
FAR\_WALL01 e simili  
NO\_WALL01 e simili.

Nella figura che segue l'area in colore (di ampiezza 1 m o 2m a seconda dei casi) individua le zone CLOSE\_WALL mentre quella bianca individua quelle FAR\_WALL.



Una etichetta di posizione (position label) deve essere sempre usata insieme ad una etichetta di luogo (place label).

In merito alle etichette CLOSE\_WALL e FAR\_WALL si possono avere fino a quattro distinte posizioni per lo stesso luogo (ad esempio un ufficio) per cui in un ufficio si possono avere registrazioni fino ad 8 diverse posizioni (4 in CLOSE\_WALL e 4 in FAR\_WALL).

E' ovvio che distanze di pochi centimetri fra le posizioni di due parlanti non permettano di individuare due posizioni distinte e lo stesso vale per piccole variazioni nel settaggio delle apparecchiature di registrazione e per i movimenti dei parlanti e le differenze delle loro altezze.

Nel caso di ufficio e intrattenimento perché sia necessario usare due etichette di posizione distinte per individuare due posizioni distinte il microfono nella posizione "medium distance" deve essere spostato di almeno 1 m.

Nel caso di luoghi pubblici perché sia necessario usare due etichette di posizione distinte per individuare due posizioni distinte il microfono nella posizione "medium distance" deve essere spostato di almeno 5 m.

In merito al parlante si ha che in tutti gli ambienti (ufficio, intrattenimento e luoghi pubblici) dopo che è stata fissata una posizione di registrazione i parlanti non si deve posizionare ad una distanza maggiore di più di mezzo metro altrimenti è necessario definire una nuova posizione ed usare una nuova etichetta di posizione e quindi è necessario spostare il microfono nella posizione "medium distance" di almeno 1 m.

Ovviamente ogni volta che si cambia la posizione di registrazione è necessario ripetere le misure delle caratteristiche di riverbero.

### *Le etichette di dimensione.*

Le etichette di dimensione sono utili per caratterizzare un ambiente di registrazione in base alla sua superficie stimata in metri quadri. Etichette legali sono del tipo SQM\_##\_## dove le coppie di ## individuano rispettivamente il limite inferiore e quello superiore della stima della superficie in modo che SQM\_10\_20 individui un ambiente la cui superficie è stimata fra 10 e 20 metri quadri. Se il limite superiore non è noto si usa la forma SQM\_##+.

Nel caso di luoghi pubblici è obbligatorio scegliere locali la cui dimensione sia superiore a 100 metri quadri.

### *Le etichette per la guida.*

Si rimanda alla pagina 19 del documento “SPEECON Deliverable D212”, comunque non interessano.

### *Le etichette per l’audio.*

Sono usate solo per il caso del divertimento (o intrattenimento, entertainment) e sono utili per documentare lo stato degli apparati audio (TV, radio, lettori di CD) presenti nell’ambiente. Tali etichette hanno valore OFF se tutti i dispositivi audio presenti nella stanza sono spenti, ON se almeno uno dei dispositivi audio presenti nella stanza è acceso e dà luogo a musica o parlato di sottofondo.

### ***Specifica dettagliata degli scenari di registrazione***

E’ necessario definire le possibili distribuzioni dei parlanti per gli scenari di registrazione ovvero relativamente alle label di condizione (place label, position label, size label e audio label).

#### *Ambiente “ufficio”*

Tale ambiente individua un ufficio ovvero una stanza in cui le persone lavorano alla scrivania di solito o saltuariamente facendo uso di computer. Durante la registrazione l’ufficio non dovrebbe essere sede di discussioni o riunioni.

Il livello di rumore deve essere compreso fra 30 e 60 dBA.

Il valore del SNR come misurato dal “Recording Studio” per il microfono di tipo headset deve essere almeno di 15dB o superiore mentre non sono definiti limiti per gli altri microfoni.

Le etichette da usare sono:

- di luogo;
- di posizione (CLOSE\_WALL e FAR\_WALL da 1 a 4 posizioni per luogo) e
- di dimensione a solo scopo di documentazione.

Ad esempio la situazione di una registrazione nel primo ufficio, nella prima posizione vicino al muro in un ambiente di dimensione stimata di 25 mq si traduce nella combinazione:

OFFICE\_01, CLOSE\_WALL\_01, SQM\_20\_30

Ulteriori dettagli a pagina 20 del documento “SPEECON Deliverable D212”.

### *Ambiente “intrattenimento”*

Tale ambiente individua un soggiorno ovvero una stanza con mobili, posti in cui le persone possono sedere. Possono essere presenti un tavolo, un TV o qualche apparecchio audio. Al posto di un soggiorno è possibile usare una stanza d'albergo.

Il livello di rumore deve essere compreso fra 30 e 65 dBA.

Il valore del SNR come misurato dal “Recording Studio” per il microfono di tipo headset deve essere almeno di 15dB o superiore mentre non sono definiti limiti per gli altri microfoni.

E' necessario che nessun apparecchio audio in funzione sia a meno di 2 metri dal microfono in posizione “far distance”.

Le etichette da usare sono:

- di luogo;
- di posizione (CLOSE\_WALL e FAR\_WALL da 1 a 4 posizioni per luogo) e
- audio (valore ON OFF): il numero di registrazioni con diverse condizioni audio è da mettere in relazione con il numero totale di parlanti ed è indipendente da posizione e luogo.

Ad esempio la situazione di una registrazione nel primo ambiente di tipo intrattenimento, nella prima posizione lontano dal muro in un ambiente di dimensione stimata di 11 mq si traduce nella combinazione:

ENTERTAINMENT\_01, FAR\_WALL\_01, SQM\_10\_20

Ulteriori dettagli a pagina 21 del documento “SPEECON Deliverable D212”.

### *Ambiente “luogo pubblico”*

Tale ambiente individua una stanza molto grande o un salone, un corridoio. Il salone (o corridoio) dovrebbe avere almeno tre pareti e un soffitto, dovrebbero essere presenti persone variamente impegnate ma non troppo tranquille.

Un'area aperta è priva di muri e di soffitto e può essere delimitata dai muri degli edifici circostanti nel qual caso almeno due dei muri devono essere ad una distanza inferiore ai 10 m in modo da consentire l'esecuzione delle registrazioni nell'angolo di due edifici.

In tutti i casi alberi, negozietti caffè all'aperto, traffico e zone pedonali sono ammissibili.

Il livello di rumore deve essere compreso fra 45 e 90 dBA.

Il valore del SNR come misurato dal “Recording Studio” per il microfono di tipo headset deve essere almeno di 15dB o superiore mentre non sono definiti limiti per gli altri microfoni.

Le etichette da usare sono:

- di luogo (PUBHALL al chiuso e PUBOPEN all’aperto);
- di posizione (CLOSE\_WALL e FAR\_WALL da 1 a 4 posizioni per luogo) e
- di dimensione: un luogo pubblico per essere adatto deve avere una superficie di almeno 100 mq. A parte questo le etichette di dimensione (size label) servono solo a solo scopo di documentazione.

Ad esempio la situazione di una registrazione in un luogo pubblico al chiuso, nella prima posizione lontano dal muro in un ambiente di dimensione stimata di 150 mq si traduce nella combinazione:

PUBHALL\_01, FAR\_WALL\_01, SQM\_100\_200

Come ulteriore esempio la situazione di una registrazione in un luogo pubblico all’aperto, nella prima posizione in assenza di muri in una piazza di dimensione stimata di 250 mq si traduce nella combinazione:

PUBOPEN\_01, NO\_WALL\_01, SQM\_200+

Ulteriori dettagli a pagina 22 del documento “SPEECON Deliverable D212”.

### *1.6 Ulteriori dettagli circa la procedura di misura*

#### *La risposta impulsiva.*

Ogni volta che l’attrezzatura di registrazione è messa in una nuova posizione (a maggior ragione in un nuovo luogo) e si rende necessaria una nuova etichetta di posizione se ne devono misurare le caratteristiche di riverberazione. La misura deve essere fatta una volta per posizione e non per ciascun parlante nella stessa posizione.

Per la misura si usano tecniche di cross-correlation utilizzando un segnale con caratteristiche di rumore rosa riprodotto mediante un altoparlante messo nella stessa posizione che occuperanno i parlanti.

Per l’acquisizione del segnale campione si fa uso di due microfoni, uno in stretta prossimità dell’altoparlante e l’altro nelle due posizioni “medium distance” e “far distance” rispettivamente.

Si hanno i seguenti casi:

- ufficio e intrattenimento: si devono effettuare due misure, una con il secondo microfono in posizione “medium distance” e l’altra con il secondo microfono in posizione “far distance”;

- luoghi pubblici: si deve eseguire solo la misura con il secondo microfono in posizione “medium distance”.

Dal momento che i parlanti si muovono durante la sessione di registrazione (che dura un’ora) e che parlanti distinti occupano posizioni che differiscono sia pure di poco (sebbene siano individuate dalla stessa etichetta di posizione) è necessario tenere conto delle variazioni che possono influenzare le reali registrazioni del parlato e ciò lo si fa effettuando ciascuna misura delle caratteristiche di riverbero di una posizione in tre tempi ovvero con l’altoparlante nella stessa posizione della testa del parlante, con l’altoparlante spostato di 50 cm a sinistra sulla linea delle spalle e, infine, con l’altoparlante spostato di 50 cm a destra, sempre sulla linea delle spalle.

Una misura della risposta impulsiva come definita in precedenza si traduce in effetti in tre misure, una per ciascuna posizione dell’altoparlante.

Le tre misure devono essere prese allo stesso livello che sarà occupato dalla testa dei parlanti e con l’altoparlante sempre ortogonale alla linea delle spalle.

In dettaglio si ha (vedi pagine 25 e 26 di SPEECON Deliverable D212):

- 1 si deve settare l’apparecchiatura per lo scenario corrente (microfoni “close distance”, “medium distance” e “far distance” connessi con l’apparecchiatura per la registrazione);
- 2 si mette l’altoparlante allo stesso posto del parlante e si aggiusta il livello di rumore in prossimità dell’altoparlante far 100 e 103 dB(A) in modo da assicurare un elevato livello di rumore (rispetto al rumore di fondo) in prossimità dei microfoni montati in “medium distance” e “far distance”;
- 3 si eseguono tre registrazioni con il secondo microfono in posizione “medium distance” (una con l’altoparlante in linea con il microfono, una con l’altoparlante spostato di 50 cm a sinistra e una con l’altoparlante spostato di 50 cm a destra);
- 4 si eseguono tre registrazioni con il secondo microfono in posizione “far distance” (una con l’altoparlante in linea con il microfono, una con l’altoparlante spostato di 50 cm a sinistra e una con l’altoparlante spostato di 50 cm a destra), si usi lo stesso preamplificatore del caso precedente e lo stesso valore del guadagno;

Nei casi in cui si usano le etichette CLOSE\_WALL e FAR\_WALL (ufficio, intrattenimento e luoghi pubblici) fare riferimento alla figura 12 di pagina 27 di SPEECON Deliverable D212 per le tipiche configurazioni di microfoni e altoparlante.

Ogni singola registrazione deve essere eseguita come di seguito specificato:

- assicurarsi che l’altoparlante stia immobile durante le registrazioni, non sono ammessi né movimenti dell’altoparlante né la presenza di persone in prossimità dell’apparecchiatura di misura;
- assicurarsi che il rumore ambientale sia ridotto o assente durante la misura (da eseguirsi dopo l’ora di punta e in assenza di estranei);
- aggiustare il livello di registrazione che dovrà rimanere lo stesso nei due casi (“medium distance” e “far distance” e in tutte le registrazioni);

- iniziare la registrazione con l'altoparlante nella stessa posizione della testa del parlante;
- riprodurre la sequenza di rumore da CD;
- interrompere la registrazione dopo che la sequenza è terminata;
- ripetere gli ultimi tre passi per le altre due posizioni dell'altoparlante (ovvero quella con l'altoparlante spostato di 50 cm a sinistra e l'altra con l'altoparlante spostato di 50 cm a destra).

### *Il livello medio di rumore.*

Per ciascun parlante è necessario misurare il livello medio di rumore della sessione di registrazione corrente. La misura si compone di due parti ovvero:

- la registrazione di un campione di rumore (la cosiddetta “parola silenziosa”) prima delle registrazioni reali e
- la misura del livello equivalente di pressione sonora ( $L_{Aeq}$ ) durante la registrazione della parola silenziosa.

La misura di  $L_{Aeq}$  consente di accertarsi che il rumore ambientale cade all'interno del range specificato per l'ambiente corrente. Insieme alla registrazione del rumore il valore  $L_{Aeq}$  serve anche come una misura assoluta di riferimento per il rumore ambientale.

In dettaglio la procedura di misura viene eseguita come segue:

- si prepara l'attrezzatura per lo scenario corrente (microfoni “close distance”, “medium distance” e “far distance” connessi con l'apparecchiatura per la registrazione);
- si aggiustano i livelli dei preamplificatori con l'ausilio del parlante presente al momento della misura che dovrà parlare il più a voce alta possibile, livelli che non dovranno più essere toccati per quel parlante;
- si posiziona il misuratore di livello sonoro IdB accanto al microfono “medium distance”;
- si esegue una registrazione del rumore di fondo della durata di 20 secondi, la cosiddetta registrazione della “parola silenziosa”, in presenza del parlante che dovrà mantenere il più assoluto silenzio;
- mentre si esegue la registrazione si setta lo strumento IdB sul modo  $L_{eq}$  e si esegue la misura di  $L_{eq}$  durante la registrazione della “parola silenziosa”.

### *1.7 La piattaforma di registrazione*

La piattaforma di registrazione si basa su un portatile che fa uso di due slot di tipo PCMCIA per interfacciarsi con le apparecchiature audio vere e proprie.

Il formato dei file audio è il seguente:

- frequenza di campionamento: 16kHz;
- quantizzazione lineare;
- ordine dei byte little endian (ovvero i byte sono numerati da destra a sinistra partendo da 0);
- silenzio prima del parlato superiore a 0.5 secondi;
- silenzio dopo il parlato superiore a 0.5 secondi.

Per garantire una elevata qualità dei materiali registrati l'apparecchiatura audio identica per tutte le registrazioni comprende elementi delle stesse marche interfacciati con il portatile mediante schede dello stesso tipo.

La misurazione della risposta impulsiva della stanza e del livello di rumore richiede l'uso di apparecchiature aggiuntive quali:

- un piccolo altoparlante attivo a banda larga;
- un lettore portatile di CD e un misuratore di rumore, entrambi alimentati a batterie.
- 

Per le specifiche del portatile si rimanda alla pagina 29 di SPEECON Deliverable D212.

Le interfacce con il portatile sono costituite da due schede PCMCIA VXpocket ciascuna delle quali ha due canali analogici di ingresso e due canali analogici di uscita in modo che due schede mettano a disposizione i quattro canali necessari.

Le registrazioni devono essere effettuate tutte ad una frequenza di campionamento di 16kHz e con campioni di 16 bit e quantizzazione lineare.

Per tutti i microfoni si fa uso di preamplificatori esterni (dettagli alla pagina 30 di SPEECON Deliverable D212).

I microfoni che saranno usati sono di vari tipi diversi (compresi quelli per l'auto che a noi non interessano) e che rientrano nelle tipologie:

- "close distance" (headset e lavalier);
- "medium distance";
- "far distance".

Per le relative specifiche vedi le pagine da 32 a 37 di SPEECON Deliverable D212.

Altri dispositivi necessari sono:

- l'altoparlante, necessario per misurare la risposta impulsiva della stanza, deve essere in grado di riprodurre "rumore rosa" ad elevati livelli di segnale (100 dBA), deve essere mobile e alimentabile con 12V;

- il lettore di CD, necessario per riprodurre il “rumore rosa” durante la misurazione del la risposta impulsiva della stanza;
- il misuratore del livello di rumore IdB, necessario per monitorare il livello di rumore durante le registrazioni e per controllare il livello di segnale durante la misurazione del la risposta impulsiva della stanza;
- il phantom power supply che fornisce una sorgente di alimentazione e il disaccoppiamento dei segnali, per dettagli vedi a pagina 39 di SPEECON Deliverable D212;
- supporti per microfoni da tavolo e da pavimento;
- una cuffia per il monitoraggio delle registrazioni e
- una scrivania e due sedie.

## 2 “SPEECON Deliverable D33”

Il documento contiene le linee guida e suggerimenti pratici in merito alle fasi più critiche del processo di creazione del data base di parlato.

Il database viene creato in due fasi.

La fase che ci interessa è la seconda nella quale si passa al “reclutamento”, alla registrazione e alla annotazione dei parlanti. Sulla base delle specifiche, parlanti con le caratteristiche che vengono descritte al punto 3 devono essere “reclutati” e devono venire registrati negli opportuni ambienti acustici. Eseguite le registrazioni queste devono essere annotate e su di esse deve essere eseguito del post-processing che completa il processo di creazione del data base.

### 2.1 Istruzioni per la registrazione: settaggio di base

In tutti gli ambienti di registrazione è necessario eseguire un settaggio di base delle attrezzature che comprende l’installazione dei portatili, il montaggio delle schede VXpocket e degli amplificatori oltre ai microfoni di tipo headset e altri di uso generale in tutti gli ambienti.

Gli scenari cui si applicano le istruzioni sono i seguenti:

- ufficio,
- intrattenimento e
- luoghi pubblici.

L’apparecchiatura necessaria comprende i seguenti elementi:

- misuratore IdB (per le misure di risposta impulsiva e di rumore ambientale);
- un certo numero di cavi XLR;
- uno o due portatili oppure (opzionale) al posto del secondo portatile un ulteriore monitor con connessione di rete;
- entrambe le schede VXpocket;
- il microfono lavalier o da cravatta;
- il microfono di tipo headset e
- la valigia di registrazione con 4 amplificatori (uno per ciascun potenziale microfono: due “close distance”, uno “medium distance” e uno “far distance”) e un sistema di alimentazione autonomo (phantom power device o ppd).

Il settaggio dell’apparecchiatura richiede che vengano seguiti i passi in elenco:

- si installa il primo portatile (quello dell’operatore) e lo schermo di visualizzazione usato dal parlante (altro portatile o monitor ausiliario);
- si installano le due schede VXpocket sul primo portatile;
- si fa il bootstrap del portatile di operatore e di quello di prompt;
- si lancia il programma di registrazione e lo si setta per registrare il parlante successivo;

- si connettono con le necessarie prese a jack i canali sinistro e destro delle schede VXpocket con i 4 amplificatori;
- si connette il microfono da cravatta con il jack apposito alla opportuna presa del ppd;
- si accendono tutti gli alimentatori (sia degli amplificatori sia il ppd);
- si accendono o spengono i filtri passa lato degli amplificatori (tale decisione deve essere presa una volta per tutte e mantenuta su tutto un data base);
- si aggiusta il guadagno degli amplificatori in base ai valori indicati nella tabella di pagina 5/16.

## 2.2 *Settaggio per le misurazioni*

Ogni volta che il luogo di registrazione viene cambiato oppure ogni volta che all'interno di un luogo di registrazione cambiano le posizioni del microfono o del parlante è necessario eseguire di nuovo le misurazioni delle caratteristiche acustiche della nuova posizione.

Ciò significa che l'attrezzatura per le misurazioni deve essere montata e devono essere eseguite le misure delle caratteristiche acustiche dell'ambiente di registrazione.

Una nuova misura è necessaria anche per ciascuna nuova posizione di registrazione ovvero per le varie posizioni all'interno di uno stesso ambiente.

Durante l'esecuzione delle registrazioni necessarie per le misure è importante che nessun'altra sorgente sonora sia presente nell'ambiente per cui è necessario prendere alcune precauzioni:

- tutti i presenti devono mantenere il più assoluto silenzio;
- non devono essere prodotti rumori di nessun genere (ad esempio caduta di oggetti, sbattere di porte e simili);
- nessuno deve stare davanti all'altoparlante, sulla linea che va dall'altoparlante al microfono;
- in caso di luoghi rumorosi (se possibile) attendere momenti di relativa tranquillità;
- è consigliabile ripetere la registrazione fino ad ottenerne una di qualità ragionevolmente buona: se si verificano rumori di qualche tipo oppure variazioni nell'ambiente è necessario ripetere la registrazione.

Gli ***scenari di registrazione*** cui si applicano le presenti istruzioni sono tutti quelli di interesse ovvero: “ufficio”, “intrattenimento” (abitazione) e “luoghi pubblici” (sia all'aperto sia al chiuso).

L'attrezzatura necessaria è quella descritta al punto 2.1 con in più:

- 2 microfoni Haun con sostegni a terra;
- un altoparlante con un supporto per microfono e un cavo di alimentazione;
- un lettore di CD connesso all'altoparlante e
- il misuratore I-dB.

Per il settaggio dell'apparecchiatura oltre ai passi elencati al punto 2.1 si richiede che vengano eseguite le seguenti misurazioni:

- misura del rumore ambientale;
  - misura della risposta impulsiva dell'ambiente
- i cui dettagli sono già stati forniti.

### 2.3 *Settaggio dell'ambiente "ufficio"*

Le seguenti istruzioni si applicano agli scenari di registrazione "ufficio" e "intrattenimento".

L'attrezzatura necessaria è quella descritta al punto 2.1 con in più:

- due microfoni, uno per la posizione "medium distance" e l'altro per la posizione "far distance";
- un supporto da tavolo e uno da pavimento e
- i necessari cavi di connessione.

Per il settaggio dell'apparecchiatura oltre ai passi elencati al punto 2.1 si richiede che vengano eseguite le seguenti operazioni:

- sistemazione del microfono "medium distance" e sua connessione con l'amplificatore 3;
- sistemazione del microfono "far distance" e sua connessione con l'amplificatore 4;
- aggiustamento del guadagno degli amplificatori 3 e 4 come visto al punto 2.1;
- far indossare al parlante il microfono di tipo headset e sistemarlo in modo che sia posizionato non di fronte alla bocca ma in prossimità di un angolo di questa;
- registrazione di alcuni elementi del corpus in modalità di test in modo da controllare gli indicatori di saturazione (o clipping);
- in caso di assenza di clipping passare in modalità registrazione ripartendo da capo con gli elementi da registrare;
- in caso di presenza di clipping è necessario modificare i guadagni degli amplificatori e ripetere le registrazioni di prova fino a che non ci si porta nel caso precedente.

Si fa notare come sia importante che il parlante usi nella fase di test la sua altezza di voce naturale e che in caso si modificano i guadagni li si deve riportare alle posizioni originarie prima di proseguire con le registrazioni in altri ambienti.

### 2.4 *Settaggio dell'ambiente "luoghi pubblici"*

Le seguenti istruzioni si applicano agli scenari di registrazione in "luogo pubblico" sia all'aperto sia al chiuso.

L'attrezzatura necessaria è quella descritta al punto 2.1 con in più:

- due microfoni, entrambi per la posizione "medium distance";

- due supporti per microfono e
- i necessari cavi di connessione.

Per il settaggio dell'apparecchiatura oltre ai passi elencati al punto 2.1 si richiede che vengano eseguite le seguenti operazioni:

- sistemazione di uno dei due microfoni “medium distance” e sua connessione con l'amplificatore 3;
- sistemazione dell'altro microfono “medium distance” e sua connessione con l'amplificatore 4;
- aggiustamento del guadagno degli amplificatori 3 e 4 come visto al punto 2.1;
- far indossare al parlante il microfono di tipo headset e sistemarlo in modo che sia posizionato non di fronte alla bocca ma in prossimità di un angolo di questa;
- registrazione di alcuni elementi del corpus in modalità di test in modo da controllare gli indicatori di saturazione (o clipping);
- in caso di assenza di clipping passare in modalità registrazione ripartendo da capo con gli elementi da registrare;
- in caso di presenza di clipping è necessario modificare i guadagni degli amplificatori e ripetere le registrazioni di prova fino a che non ci si porta nel caso precedente.

Di nuovo, in caso si modifichino i guadagni li si deve riportare alle posizioni originarie prima di proseguire con le registrazioni in altri ambienti.

## *2.5 Uso delle schede VXpocket e del software “Recording Studio” sul portatile “di operatore”*

### ***Preliminari***

- è necessario disconnettere sia la valigetta di registrazione sia i portatili dalle sorgenti di alimentazione durante le registrazioni in modo da non introdurre distorsioni nei file registrati;
- è necessario accendere il portatile “di operatore” con una sola scheda VXpocket inserita nello slot inferiore del bus PCMCIA e inserire l'altra nello slot superiore del bus PCMCIA solo quando Windows 98 è in esecuzione;
- è necessario lanciare in esecuzione il programma DIGIWMIX.exe e usarlo per settare i guadagni delle schede VXpocket (pagina 11/16) e salvare tali settaggi come configurazione di default;
- è necessario collegare la scheda VXpocket inserita nello slot inferiore del bus PCMCIA con i (pre)amplificatori 1 e 2 e la scheda VXpocket inserita nello slot superiore del bus PCMCIA con i (pre)amplificatori 3 e 4 in modo da rispettare gli assegnamenti di cui alle tabelle di pagina 11/16:
- il canale 0 deve essere sempre collegato al microfono di tipo headset e deve sempre avere il più alto rapporto segnale/rumore (SNR), in caso contrario il software lo segnala con un messaggio che indica una possibile discrepanza di

- canale (channel mismatch) per cui è necessario controllare gli assegnamenti dei canali e, se necessario, scambiare i cavi collegati alle schede VXpocket;
- è consigliabile usare comandi da tastiera per iniziare/interrompere le registrazioni dal momento che i click del mouse possono venire registrati.

### ***Registrazione del parlante***

E' necessario scegliere lo stesso scenario nel menù "New SAM 6.1 Speaker Registration" e nel menù "Scenarios" (pagina 13/16) per evitare che valori errati delle etichette (SAM-labels) vengano inseriti nei file delle etichette (SAM-label-files).

Per ciascun parlante è necessario indicare:

- numero della sessione (e.g. 003);
- nome del parlante (SA003);
- identificativo del parlante, anno di nascita, sesso, nazionalità e regione oltre a un commento sullo stato emotivo e alla tipologia (adulto o bambino).

Oltre a tali dati per ciascun parlante è necessario scegliere i seguenti codici da insiemi predefiniti:

- codice di scenario;
- luogo di registrazione;
- posizione di registrazione;
- dimensione dell'ambiente e
- nome del file di script.

### ***Warnings e messaggi***

Si hanno messaggi di warning che comunicano all'operatore la necessità di compiere certe azioni o il verificarsi di condizioni anomale nei casi seguenti:

- se si inizia una sessione di registrazione senza eseguire la registrazione della risposta impulsiva della stanza seguita dalla registrazione della "parola silenziosa";
- nel caso della registrazione della "parola silenziosa" può comparire un messaggio di warning dal momento che in effetti non si esegue nessuna registrazione, messaggio che deve essere ignorato;
- se il canale 0, quello cui è collegato il microfono di tipo headset, ha un SNR più basso degli altri canali, condizione che può indicare un errato collegamento dei canali (segnalato anche da una luce rossa accanto al livello del canale incriminato) per cui è necessario controllare le connessioni e ripetere la registrazione;
- nel caso che il software non sia in grado di discriminare il rumore dal parlato ovvero non è in grado di calcolare il valore del SNR, ciò è indice di un problema tecnico con un canale (ad esempio una connessione interrotta, segnalata anche da una luce rossa accanto al livello del canale incriminato) per cui è necessario controllare l'hardware e ripetere la registrazione;
- nel caso si verifichi una saturazione di clipping per cui è necessario ripetere la registrazione;

- nel caso si chiuda una sessione di registrazione senza aver registrato tutti gli elementi del cosiddetto “prompt sheet” che contiene gli elementi che il parlante vede e deve leggere;
- nel caso si saltino le registrazioni della risposta impulsiva dell’ambiente perché già fatte o non necessarie;
- nel caso si vada oltre il tempo previsto per la durata del parlato spontaneo.

### 3 I parlanti e la struttura del corpus

I parlanti sono caratterizzati in base al sesso e alla fascia d'età mentre gli ambienti dove devono essere effettuate le registrazioni sono caratterizzati come:

- Ufficio;
- Intrattenimento e
- Luogo pubblico.

Oltre a tali ambienti dovrebbero essere considerate le registrazioni dei bambini e quelle in automobile che però non verranno prese in esame dato che in esse non risulta coinvolta l'Unità Operativa di Pisa.

Per ciascuna località (le località di pertinenza dell'Unità Operativa di Pisa sono due e sono Firenze e Milano) nell'ambiente "Ufficio" devono essere registrati 33 parlanti, 17 uomini e 16 donne, nell'ambiente "Luogo pubblico" devono essere registrati 33 parlanti, 16 uomini e 17 donne, mentre nell'ambiente "Intrattenimento" i parlanti da registrare sono 14, 7 uomini e 7 donne.

Per ciascuna località i parlanti da registrare sono pertanto 82 (=34 + 34 + 14) di cui 41 uomini e 41 donne.

Le fasce di età dei soggetti sia uomini sia donne sono quella che va dal cambio di voce (da considerarsi non inferiore ai 18 anni per evitare problemi legali) fino a 30 anni, quella compresa fra i 31 e i 45 anni e, infine, quella che parte da 46 anni.

La distribuzione per fasce di età e per ambienti dei soggetti da registrare avviene come illustrato dalle tabelle seguenti.

#### *Uomini*

Età	18-30 anni	31-45 anni	più di 46 anni
Ufficio	7	7	3
Intrattenimento	3	3	1
Luogo pubblico	6	7	3

#### *Donne*

Età	18-30 anni	31-45 anni	più di 46 anni
Ufficio	6	7	3
Intrattenimento	3	3	1
Luogo pubblico	7	7	3

Vediamo ora la struttura del Corpus relativamente ai soggetti adulti.

Il Corpus si compone di tre parti ovvero:

- parlato spontaneo a sua volta suddiviso in parlato spontaneo su dieci argomenti che sono stati scelti per quel parlante da un insieme di trenta (durata almeno 5' per parlante) e parlato elicitato (durata stimata 7' per parlante),
- parlato letto (durata stimata 15' per parlante) composto da 30 frasi foneticamente ricche e da almeno 5 parole foneticamente ricche
- parole chiave (durata stimata 33' per parlante).

Ciascun parlante risulta, pertanto, essere impegnato per un totale approssimativo di 60' (=12'+15'+33') cui va aggiunto il tempo necessario per l'esecuzione della registrazione della "parola silenziosa" preceduta dal settaggio degli amplificatori fatto facendo leggere al parlante delle frasi di test.

Il parlato elicitato si compone (per ciascun parlante) delle risposte a 17 domande suddivise, per tipologia di argomento, in 8 categorie (data di nascita attuale e relativa: 3 domande; orari: 2 domande; nomi di città: 2 domande; nomi propri di persona o di aziende: 3 domande; spelling del proprio nome: 1 domanda; risposte si/no: 2 domande; lingua parlata: 1 domanda; numeri di telefono: 3 domande).

Per il parlato letto si prevede che ciascun parlante legga 30 frasi foneticamente ricche e aventi una lunghezza media compresa fra 5 e 10 parole e produca 5 parole foneticamente ricche scelte da un insieme di 300 parole.

Per quanto riguarda le parole chiave, il compito (per ciascun parlante) è suddiviso in due sottocompiti rispettivamente della durata stimata di 8' e 25'.

Il primo di tali sottocompiti prevede la lettura di parole e frasi di tipo "general purpose". Si hanno 14 categorie di comandi che, sulla base di ripetizioni, danno luogo a 31 termini per ciascun parlante.

La suddivisione dei 31 termini fra le diverse categorie dei comandi avviene come specificato dalla tabella seguente:

<b>Categorie dei comandi</b>	<b>Numero di termini per parlante</b>
Cifre isolate	4
Sequenze random di cifre isolate	1
Stringhe di cifre (20 cifre in tutto)	4
Numeri di telefono	1
Numeri naturali decimali da 10 a 1000000	3
Quantità di denaro	1
Orari (1 in forma analogica e 1 in forma digitale)	2
Date (1 in forma analogica, 1 in forma relativa e 1 in forma digitale)	3
Parole scandite (lunghezza media 7 caratteri)	3
Nome e cognome	1
1 nome di città e uno di strada da un set di 275	2
Si/No	2
1 indirizzo e-mail (da un set di 550) e 1 indirizzo www (da un set di 150)	2
Caratteri di tastiera speciali (da un set di 20)	2

Il secondo di tali sottocompiti prevede la lettura di parole ed espressioni per applicazioni specifiche.

In questo caso si hanno 9 categorie di comandi:

- comandi di base per apparecchiature elettroniche,
- comandi per la navigazione di menù,
- comandi per l'editing,
- comandi per il controllo dell'output,
- comandi per il browsing e il messaging in internet,
- comandi per le funzioni di un organizer,
- comandi per un navigatore satellitare,
- comandi per il controllo di un'autovettura e
- comandi per sistemi audio-video

per ciascuna delle quali ogni parlante deve pronunciare 208 termini scelti a caso da un set di 452 termini.

#### ***4 Ulteriori istruzioni per l'esecuzione delle registrazioni***

Le registrazioni vengono effettuate usando il software “SONY Speech Recording Studio” (SSRS).

I passi da seguire per registrare un parlante sono i seguenti:

- a si avviano i due componenti del software (la parte Speech Recording Studio e la parte Prompter, che visualizza gli elementi che il parlante deve leggere);
- b si connette il Recording al Prompter;
- c si crea un nuovo file di parlante (Speakerfile);
- d si immettono le necessarie informazioni e si sceglie una lista di elementi (prompting list);
- e si esegue la registrazione;
- f si salva la sessione e
- g si riparte dal passo a per il parlante successivo.

Il software SSRS consente di registrare una sessione della lettura di frasi memorizzate in un file in formato testo.

Si compone di due parti:

- una per la registrazione e il controllo della sessione
- una che gestisce la visualizzazione degli elementi ad uso del parlante..

Di ogni produzione (frase letta) che sia già registrata può essere visualizzata la forma d'onda ed è possibile ascoltarla, reregistrarla o cancellarla.

Sul display della forma d'onda sono visualizzati parametri quali:

- SNR;
- offset

e altri che permettono di valutare la bontà della registrazione.

E' possibile lanciare i due programmi (registratore e visualizzatore) sia sullo stesso computer sia su due computer distinti connessi via TCP/IP.

#### ***Lanciare il software***

Si manda in esecuzione la parte recording mediante l'icona “Start Recording Studio” e poi la parte prompter o sulla macchina locale o su una remota connessa via TCP/IP.

Il software crea due finestre:

- la finestra di prompt che è identica a quella che visualizzata al parlante e
- la finestra della forma d'onda (waveform window) che visualizza varie informazioni in merito al correntemente registrato file audio.

A questo punto è necessario collegare il recording con il prompter usando l'icona o il comando opportuni (“Connect Prompter” del menù “File”) e fornendo l'indirizzo o

il nome del computer su cui gira il prompter e ciò anche se i due girano sullo stesso computer.

A questo punto si deve iniziare una nuova sessione di registrazione e per fare ciò si deve:

- aggiungere un nuovo parlante (icona “new” o comando “New Speaker” nel menù “File”) mediante una finestra di dialogo in cui è necessario immettere una serie di informazioni (nome del parlante, identificativo, anno di nascita, sesso, lingua prima, dialetto, nome del file di script che contiene gli elementi che saranno visualizzati al parlante);
- creare una nuova cartella per il nuovo parlante;
- creare lo Speaker Information file che contiene la condizione (fisica e emotiva) del parlante e il nome di chi fa la registrazione;
- utilizzare la finestra di controllo (control window) per visualizzare gli elementi che il parlante deve leggere, registrare e cancellare i segnali audio.

A questo punto si può procedere a registrare ciò che il parlante dice e che coincide con ciò che vede sullo schermo (la cosiddetta “prompt list” o “prompting list”) e che è visibile anche all’operatore.

Quando la registrazione è interrotta l’operatore può navigare nella “prompting list” in modo da ripescare una voce, rivisualizzarla al parlante e registrarla di nuovo tutte le volte che sia necessario.

### ***Le opzioni del programma***

E’ possibile modificare i settaggi di default del programma utilizzando una finestra di dialogo ad hoc.

I settaggi modificabili comprendono:

- le opzioni della directory;
- i settaggio di controllo della forma d’onda e
- i settaggi del dispositivo di I/O

ma la loro trattazione esula dagli scopi delle presenti note.

## 5 *Minima moralia*

Le registrazioni da fare nelle località di Firenze (o uno dei comuni dell'interland) e di Milano (o uno dei comuni dell'interland) prevedono la scelta di:

- un ufficio (office);
- una abitazione (entertainment);
- un luogo pubblico (public place) chiuso e
- un luogo pubblico (public place) aperto.

I soggetti (devono avere almeno il Diploma, da cui anche la necessità che abbiano almeno 18 anni) da registrare sono 14 (7 + 7) per l'abitazione e 33 per l'ufficio (17 + 16) e 17 per gli altri due luoghi.

Data una abitazione (entertainment) si sceglie una stanza di superficie tale da consentire di scegliere due posizioni:

1 CLOSE\_WALL

1 FAR\_WALL

in modo da poter registrare il numero necessario di soggetti.

Dato un ufficio (office) si sceglie una stanza di superficie tale da consentire di scegliere quattro posizioni:

2 CLOSE\_WALL

2 FAR\_WALL

in modo da poter registrare il numero necessario di soggetti.

Per quanto riguarda i luoghi pubblici

- al chiuso: devono avere superficie superiore a 100mq per cui si possono scegliere Bar grandi, hall di stazioni ferroviarie, supermercati o grandi aule universitarie;
- all'aperto: devono essere molto grandi (tipo 25m\*25m, in modo da poter scegliere posizioni in cui il muro più vicino disti almeno 10 m) per cui si possono scegliere terrazze, parchi pubblici, piazze o piazzette e cortili dell'Università.

Per ciascuna località è necessario trovare 2 luoghi pubblici al chiuso e 2 o tre luoghi pubblici all'aperto.

Data la filosofia con cui si esegue la registrazione, i soggetti possono chiedere chiarimenti durante le sessioni di parlato spontaneo ed elicitato e, in misura minore, durante le altre e possono, se stanchi, chiedere di fare brevi pause fra una sessione ed un'altra.

Sia per il parlato spontaneo libero sia per il parlato elicitato il parlante e ha un argomento o una domanda guida che deve leggere ma non pronunciare.

La registrazione, in questi casi, deve iniziare al momento in cui il prompt (argomento o domanda) viene inviato al monitor del parlante in modo che questi possa leggerlo e deve essere interrotta circa mezzo secondo dopo che il parlante ha finito (o è stato interrotto).

Nel caso di parlato spontaneo libero il parlante dovrebbe parlare in media 30'' (5\*60/10) per argomento e lo stesso dicasi per ciascuna domanda del parlato elicitato.

Per le registrazioni in luogo pubblico aperto è necessario evitare il formarsi di capannelli di curiosi per cui sarebbe opportuno scegliere terrazze o luoghi simili.

I tempi dati (a parte quello del parlato spontaneo libero che è un vincolo stretto) sono delle stime anche se sarebbe bene cercare di tenere la durata della sessione al di sotto di 1h.

E' necessario registrare al minimo 80 parlanti per località + 2 di riserva, da utilizzare in caso di problemi.

E' necessario fare il backup su CD-R (mediante masterizzatore portatile) in doppia copia di tutto il materiale registrato durante una giornata

Le registrazioni si devono ripetere, nel caso del parlato letto e della lettura di parole chiave, solo se il soggetto commette errori di lettura.

Si deve ripetere la sessione se il software va in crash oppure se cambiano radicalmente le condizioni ambientali.