CAPITOLO 4

SOTTOTITOLAZIONE DI TELEFONATE E VIDEOCONFERENZE

CAPITOLO 4

SOTTOTITOLAZIONE DI TELEFONATE E VIDEOCONFERENZE

4.1 INTRODUZIONE

Il progetto Voice, come ricordato in precedenza, vedeva tra gli obiettivi del suo terzo ed ultimo anno di svolgimento la sperimentazione di sistemi di riconoscimento vocale per sottotitolare una conversazione telefonica.

Su questa sperimentazione, partita in tono minore, è stato possibile concentrarsi solo ora, sulla scia dei successi ottenuti nelle conferenze e durante le lezioni scolastiche. Si è quindi affrontata la sfida più grande: sottotitolare una conversazione sordo-udente facendo comparire sul monitor del computer posto in casa del sordo le parole del suo interlocutore.

A questo proposito ricordo che moltissimi audiolesi sono in grado di parlare, motivo per cui la barriera principale per una conversazione telefonica rimane per essi quella di sentire il proprio interlocutore. Prima che con i sottotitoli, questo tipo di problema è stato affrontato cercando soluzioni tecniche in grado di interfacciare il telefono con la protesi acustica.

Questo tipo di soluzione funziona parzialmente con i telefoni di vecchia concezione, che avevano un grosso magnete nella cornetta, e con le protesi acustiche che hanno la possibilità di captare il suono per induzione magnetica (piuttosto diffuse).

L'avvento dei telefoni cellulari GSM ha creato ulteriori difficoltà, dovute alle interferenze che il telefonino stesso produce nella protesi acustica (simili a quelle che percepiamo ponendo il telefonino vicino ad una televisione o ad una radio).

Il problema, nonostante le possibilità di comunicazione alternative, è tuttora molto sentito nella comunità degli audiolesi, ed è stato affrontato con approcci differenti. Il DTS[55], ormai tecnologicamente sorpassato, è stato per molto tempo la soluzione più diffusa e viene impiegato ancora oggi in servizi come il 'servizio ponte'[56] che consente una chiamata ad un sordo tramite operatore. Un esempio dell'evoluzione di questo tipo di soluzioni viene da Düsseldorf, dove la Caritas locale ha finanziato delle cabine telefoniche per sordi 'segnanti', che attraverso un videotelefono collegato con un interprete gestuale, consente di effettuare una conversazione con un udente.

[55] Dispositivo telefonico per sordi che consente un dialogo testuale tra due apparecchi dello stesso tipo.

[56] Servizio ponte: il sordo contatta tramite DTS o FAX un operatore che gli fa da 'interprete' con l'udente. (in Italia a Firenze:055/6505120)

In questo anno sono comparsi sul mercato diversi nuovi sistemi di videoconferenza, ed in particolare dei video telefoni basati sulle medesime tecnologie, che sfruttano tutte le potenzialità della linea ISDN[57] oggi in via di diffusione anche nelle abitazioni private. Per questo motivo è stata condotta una sperimentazione nell'ambito di videoconferenze e di video telefonate, sia per testarne le possibilità in termini di lettura labiale e comunicazione gestuale, sia per definire i bisogni degli utenti in materia di sottotitoli in questo specifico ambito.

Nel presente capitolo verranno illustrate le fasi della sperimentazione nei vari ambiti ed i diversi test che hanno portato a dei risultati intermedi per la sottotitolazione della telefonata, che si sono rivelati importanti al fine di separare i diversi aspetti in vista di raggiungere l'obiettivo finale. È da sottolineare ancora una volta che sin dall'inizio lo spirito del progetto è stato quello di utilizzare tecnologie esistenti e facilmente reperibili per dimostrare le potenzialità di sistemi di riconoscimento vocale nel mondo della disabilità. Ciò significa in concreto che non sono stati sviluppati, per scelta, prototipi hardware o software innovativi, con lo scopo di dimostrare le potenzialità di tecnologie oggi comuni; nella speranza di stimolare la ricerca da parte dell'industria e/o gli enti che sarebbero in grado di ottenere rapidi e significativi miglioramenti se investissero in questo campo.

Prima di affrontare nello specifico le varie fasi della sperimentazione, é necessaria una breve introduzione riguardante i sistemi di riconoscimento vocale oggi disponibili ed i loro principi base di funzionamento.

4.1.1 RICONOSCIMENTO VOCALE DISCRETO.

Questo tipo di riconoscimento vocale esiste ormai da parecchi anni, e ha la possibilità di funzionare con hardware piuttosto datati, in quanto l'entità dell'elaborazione richiesta è modesta. All' oratore si richiede di dettare facendo una piccola pausa tra una parola e l'altra ed il riconoscimento avviene parola per parola, ovvero il computer analizza i suoni tra le pause, cercandone la corrispondenza con uno dei vocaboli presenti nel suo dizionario. L'obbligo di dover separare le parole ha parzialmente frenato la diffusione di questo tipo di sistemi nell'ambito della dettatura, in quanto non consente all'oratore di concentrarsi sul contenuto del discorso.

Questi software possono essere usati anche nella 'modalità comandi' (presente anche nei sistemi continui) per controllare il PC con la voce. Tipicamente si possono usare una serie di comandi per le operazioni più comuni (es: [Apri MS Word] [Chiudi finestra] [Menu avvio - Programmi - WinZip]). I comandi vanno pronunciati senza pause e la frase viene interpretata come un'unica parola. Inoltre la ricerca avviene su un vocabolario molto ristretto (i soli comandi appunto) e questo fa si che nel caso il comando non venga riconosciuto con precisione, non viene eseguito nulla, evitando il rischio di comandi errati.

[57] ISDN: linea telefonica digitale che consente di avere due canali da 64 kb/s che possono essere usati in contemporanea arrivando ad una banda di 128kb/s

Questa modalità presenta anche il grande vantaggio di essere pressoché indipendente dal tipo di voce perché non utilizza un profilo vocale personalizzato per ogni utente, ma data la limitatezza del vocabolario su cui viene condotta la ricerca, garantisce buoni risultati anche con voci diverse. I vocaboli (comandi) sono stati scelti in modo da essere il più possibile differenti l'uno dall'altro in modo da minimizzare il rischio di riconoscimento di un comando troppo simile.

Nel prodotto della Dragon[58] è inoltre presente un tool di sviluppo (Vocabulary Manager) che permette sia di creare nuovi comandi, sia di associare delle macro ad ogni comando, per eseguire operazioni specifiche.

Vedremo che l'indipendenza dal tipo di voce, permessa dalla 'modalità comando', ha giocato un ruolo importante anche nelle nostre sperimentazioni.

4.1.2 RICONOSCIMENTO VOCALE CONTINUO.

L'avvento di questo tipo di tecnologia, che consente di dettare ad un computer parlando normalmente (senza pause forzate), ha aperto nuove interessanti strade e possibilità di impiego. Lo sviluppo di questi sistemi è stato reso possibile anche dalla disponibilità di computer sempre più potenti, dal momento che il numero di elaborazioni richieste cresce esponenzialmente. Infatti non vengono riconosciuti più i singoli vocaboli, ma frasi intere, ed il riconoscimento è tanto più accurato quanto più lunga e fluida è la frase. Questo perché il 'motore del riconoscimento' si basa sul confronto tra un profilo vocale specifico per ogni utente ed i vocaboli che compongono la frase pronunciata, applicando poi dei modelli probabilistici che scelgono il vocabolo più probabile rispetto al contesto. Con questi prodotti è perciò necessario e fondamentale addestrare il sistema a riconoscere la propria voce, consentendone l'utilizzo anche a soggetti con lievi difetti di dizione o inflessioni ed accenti particolari. Leggendo brani pre-impostati il sistema è in grado di creare un profilo vocale per la persona; inoltre, proponendo al sistema brani contenenti vocaboli propri o specifici di una terminologia particolare, si addestra il sistema a riconoscere il proprio modo di parlare.

L'addestramento base dura una ventina di minuti (solo quattro per l'ultimissima release di Dragon) e, pur non essendo un impegno gravoso per un oratore di una conferenza, diventa un problema nel caso della telefonata dove il chiamante è ignoto. Nel caso di una conferenza il piccolo sforzo iniziale richiesto all'oratore per addestrare il sistema è ripagato dalla correttezza e dalla fluidità del sottotitolo e, non ultima, dalla disponibilità del testo del discorso pronto per la stampa o la memorizzazione nel momento stesso in cui la conferenza termina.

Tuttavia questi sistemi, adottando i suddetti modelli probabilistici, possono generare frasi composte da parole ortograficamente corrette, ma in alcuni casi senza senso. Un esempio può essere la seguente frase: "il progetto Voice per la sottotitolazione automatica a favore degli audiolesi" è stato trascritto in: " il progetto avesse per la società nazionale romantica favore degli audiolesi".

[58] Dragon Dictate Power Edition

Questa fase, frutto di una cattiva dizione e di un'eccessiva velocità di dettatura, è composta da parole corrette che originano una frase senza senso.

I migliori risultati si hanno in tutti quei campi che abbiano una terminologia specifica e ripetitiva, come ad esempio per la refertazione in ambito medico, settore in cui hanno oggi grande diffusione. Anche nei prodotti di riconoscimento vocale continuo è disponibile la modalità 'comando' descritta al paragrafo precedente, ma il suo funzionamento è riconducibile a quello dei sistemi discreti, e dati i costi molto maggiori sia in termini di hardware necessario, sia di acquisto del tool di sviluppo, si è utilizzato il prodotto discreto per le sperimentazioni basate su tale caratteristica.

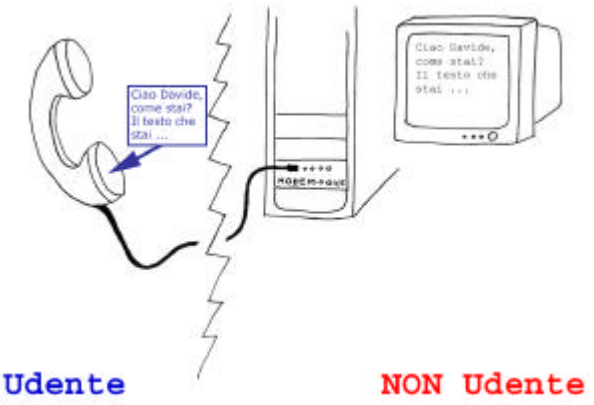
4.2 SOTTOTITOLAZIONE DI UNA TELEFONATA CON UN SISTEMA CONTINUO.

Si è tentato di sottotitolare una conversazione telefonica utilizzando un software di riconoscimento vocale continuo, Dragon NaturallySpeaking, collegando un modem-voce con la scheda audio del PC. Il fatto di aver impiegato un modem-voce è dettato da motivi di praticità, in quanto si sarebbe potuto far entrare il segnale telefonico, tramite un collegamento dedicato, direttamente nella scheda audio. La presenza del modem-voce ha tuttavia facilitato le prove, grazie anche ai software di gestione delle telefonate e delle caselle vocali ad esso associati.

Le impostazioni presenti all'interno della scheda 'multimedia' del pannello di controllo di Windows, portano a credere di poter selezionare come input ed output audio direttamente il modem Voice, senza bisogno di alcun tipo di collegamento 'esterno' tra modem e scheda audio. Questo purtroppo non è vero, a causa di differenze tra le frequenze di campionatura della scheda audio utilizzate dal software di riconoscimento vocale e le frequenze del modem. Il software infatti, progettato per prendere l'input sonoro da una scheda audio e campionare a 11.025 kHz, 16 bit, Mono, risulta incompatibile con i 9600 kHz del modem-voice.

Si è allora optato per collegare con dei cavetti l'uscita LINE OUT del modem, da cui proviene il segnale telefonico, con l'ingresso MIC della scheda audio; e parallelamente collegando l'uscita LINE OUT della scheda audio con l'ingresso MIC del modem. Per realizzare questi collegamenti sono stati inoltre impiegati diversi tipi di filtri da applicare sugli spinotti per evitare fischi e fruscio eccessivi.

A questo punto il sistema avrebbe dovuto funzionare nel seguente modo: l'audioleso sta al computer, collegato alla linea telefonica tramite il modem voce, con in aggiunta a quanto visto un microfono per rispondere vocalmente nel caso in cui sia in grado di parlare, mentre l'udente sta dall'altra parte della linea telefonica con un qualsiasi telefono.



Questo sistema ha evidenziato immediatamente limiti e difficoltà difficilmente superabili con gli strumenti disposizione:

- il segnale della linea telefonica è di scarsa qualità, mentre il riconoscimento vocale richiede una buona qualità dell'input tanto che i test di dettatura al PC avevano già evidenziato differenze tra microfoni di qualità differenti.

- altro limite era presentato dalla presenza di un unico canale per la voce di entrambi gli interlocutori. Questo crea, nel caso di un audioleso, anche potenziali problemi di sincronizzazione del dialogo, ma soprattutto, la voce del sordo va a mescolarsi e sovrapporsi alla voce dell'udente. Ciò significa ad esempio, che anche prevedendo una conversazione con una persona nota, che abbia preventivamente addestrato il computer del sordo a riconoscere la sua voce, ci sarebbe il problema che, viaggiando sul suo stesso canale, la risposta del sordo o dell'eventuale sintesi vocale verrebbe analizzata dal riconoscimento vocale, aggiungendo al testo parole a sproposito perché generate su un profilo vocale altrui.

Preso atto di queste difficoltà, si è deciso di fare dei test che permettano di analizzare singolarmente i vari aspetti del problema.

4.3 TRASCRIZIONE DI UN MESSAGGIO DA SEGRETERIA TELEFONICA

Si è perciò deciso di provare ad eliminare una delle variabili del problema, quella della gestione del dialogo, con i relativi problemi di sovrapposizione delle voci ed i rischi di interferire nel riconoscimento della voce dell'oratore principale.

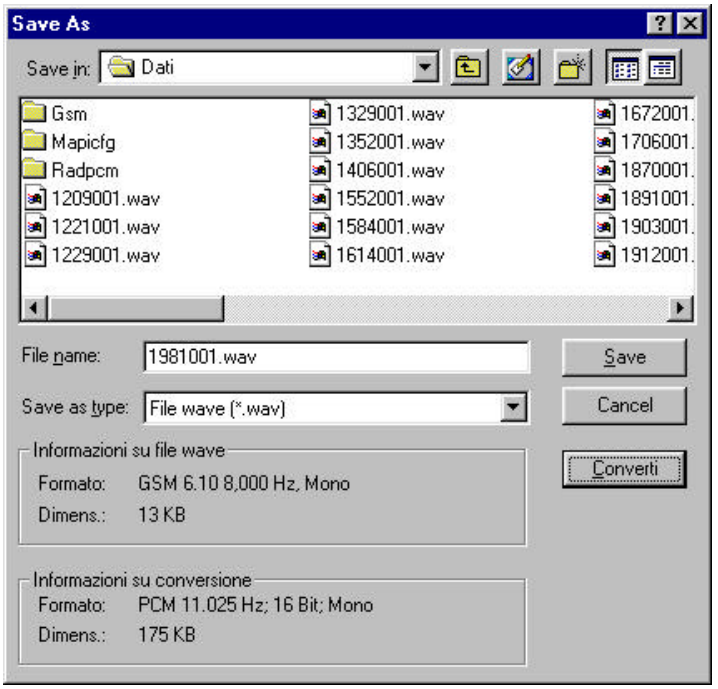
La situazione è molto simile a quella, già ben collaudata, del riconoscimento di un messaggio dettato ad un apparecchio di memorizzazione, quale ad esempio il mini registratore digitale Voice IT.

La differenza tra la situazione del riconoscimento di un messaggio registrato su registratore portatile e quello registrato dalla segreteria telefonica virtuale del modem-voice è la qualità sonora del messaggio.

Il procedimento è il seguente: attraverso un software di gestione delle caselle vocali, come ad esempio Talk Works di Symantech, viene registrato il messaggio sotto forma di file audio, che viene successivamente rielaborato dal software di riconoscimento vocale continuo (Dragon NaturallySpeaking o IBM ViaVoice) che lo traduce in testo.

In questa fase è inoltre possibile superare i problemi relativi alle frequenze di campionatura.

La segreteria telefonica virtuale infatti salva i messaggi in file .wav campionati a 9600 Mhz, GSM, Mono. Tramite un utility (audio editor) di Talk Works il file viene convertito in PCM 11.025 kHz, 16 bit, Mono, ovvero nel formato richiesto dal software di riconoscimento vocale. A questo punto il file viene analizzato (tramite la funzione 'trascrivi') dal riconoscitore vocale che lo trasforma in testo. In questa fase è anche possibile selezionare il profilo vocale desiderato, o addirittura, presupponendo che l'utente sordo non sia in grado di riconoscere la voce di chi ha lasciato il messaggio, fare diverse prove cambiando profilo vocale, e scegliendo poi la miglior trascrizione.



I risultati della trascrizione dei messaggi lasciati in segreteria telefonica sono stati decisamente buoni nel caso in cui venga utilizzato il profilo vocale appropriato. Anche i messaggi registrati da un telefono cellulare sono stati interpretati correttamente. Utilizzando invece il profilo vocale di un'altra persona il testo non risulta comprensibile. Questo evidenzia che il rumore e la scarsa qualità del segnale telefonico non sono l'unico problema che affligge il riconoscimento vocale 'in diretta' di una telefonata, ma che anche gli altri aspetti

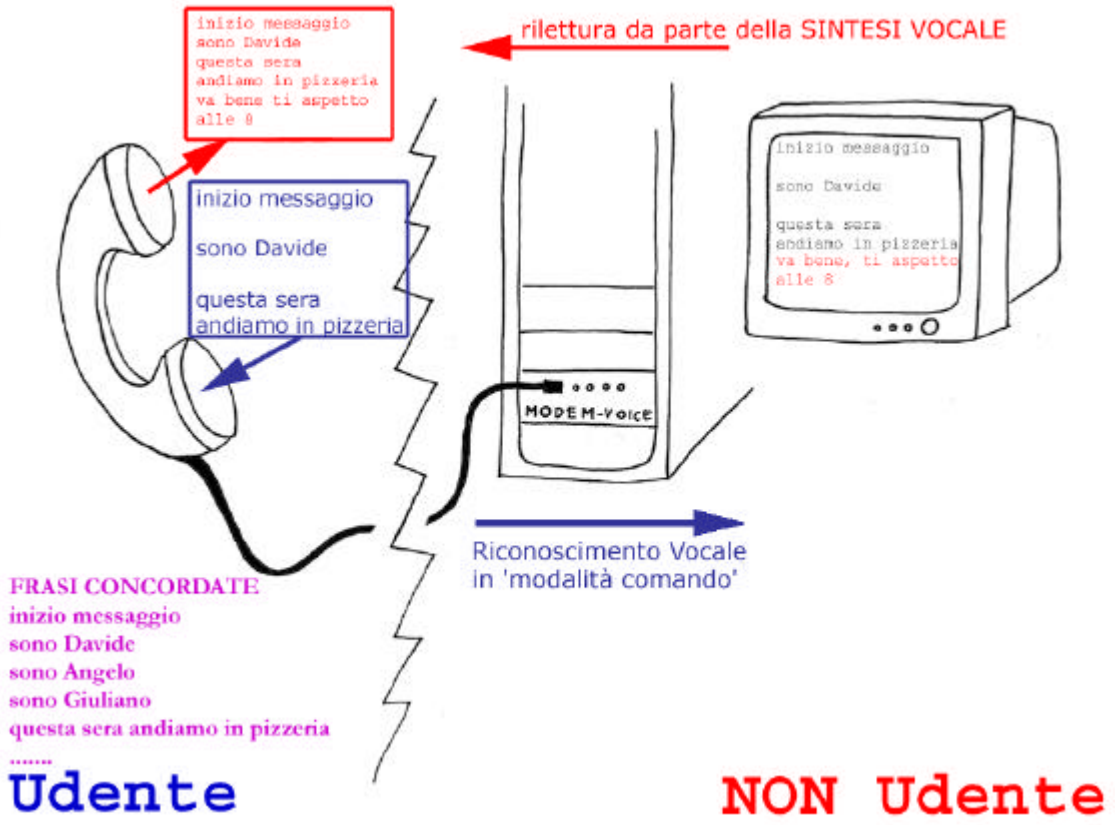
quali il campionamento, la sincronizzazione e la dipendenza dall'utente, assenti in questo test, contribuiscono al risultato finale negativo.

La scelta del profilo vocale corretto potrebbe essere fatta anche grazie ad un sistema di gestione delle caselle vocali (con Talk Works stesso ad esempio) che faccia scegliere un profilo a chi telefona (es: "Se sei Giuliano premi 1, se sei Davide premi 2, etc.") memorizzando in una particolare directory il messaggio e permettendo al sordo di selezionare il profilo vocale corretto per la trascrizione.

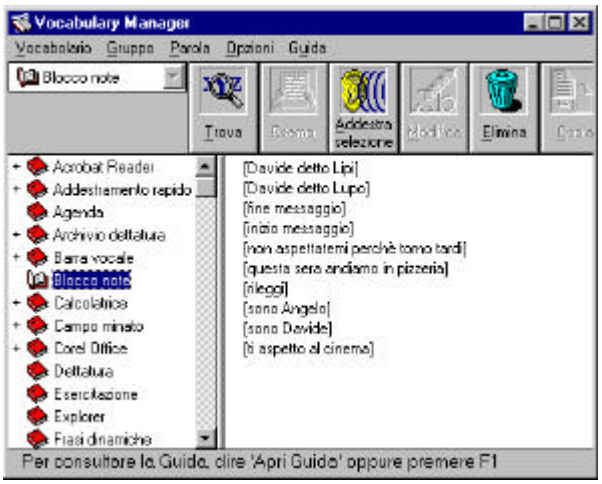
Per affrontare il problema del dialogo e della sincronizzazione degli interlocutori si è messo a punto un test che fosse indipendente dal particolare profilo vocale ovvero dall'utente, basato su un l'utilizzo di un prodotto di riconoscimento 'discreto' impiegato in modalità 'comandi'.

4.4 SOTTOTITOLAZIONE DI UNA TELEFONATA CON VOCABOLARIO RISTRETTO PRESTABILITO

Si è pensato di realizzare un sistema per consentire una telefonata tra un sordo ed un udente sfruttando in positivo quello che era stato considerato uno dei limiti principali dei sistemi di riconoscimento utilizzati in modalità 'comando', ovvero la capacità di riconoscere solo parole singole, con l'effetto di una mancata trascrizione in caso di dubbio.



A tale scopo è stato impiegato un software 'discreto' in modalità 'comando': Dragon Dictate Professional Edition. Questo sistema, come spiegato in precedenza, ha il vantaggio/limite di essere indipendente dallo specifico utente. In questa situazione rimanevano predominanti i problemi di sincronizzazione tra i due interlocutori e di dare un feedback all' udente circa la frase comparsa sul video del sordo.

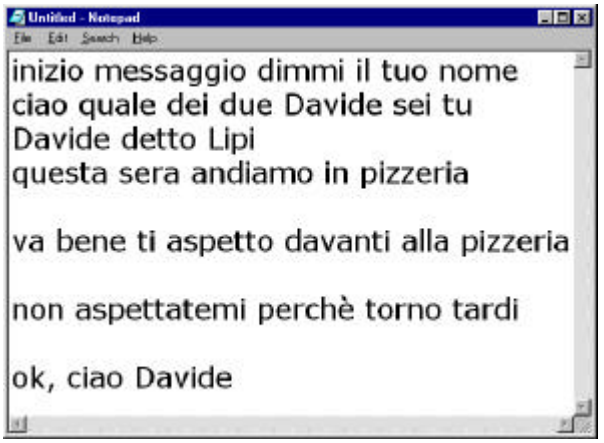


Si è superato il problema nel seguente modo: l'udente pronuncia una delle frasi concordate con il sordo. Il software avrà precedentemente memorizzato l'intera frase come un unico vocabolo, per cui la riconoscerà solo se viene pronunciata per intero correttamente.

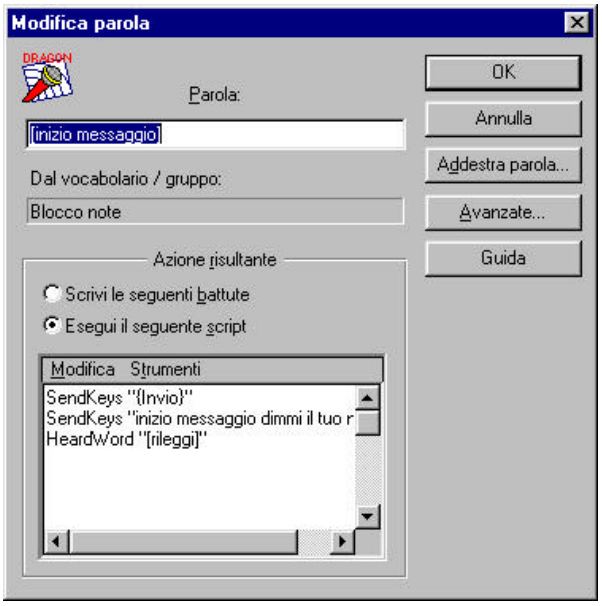
Tramite il Vocabulary Manager si definisce l'insieme dei comandi riconosciuti e si associa a loro una macro che viene eseguita quando viene riconosciuto il comando

Nel qual caso la frase comparirà sul video del sordo. Al termine di ogni frase (comando), viene automaticamente eseguita una macro istruzione che seleziona il testo appena scritto e lo fa leggere alla sintesi vocale.

Ecco il testo come compare sul Pc del sordo, che può scrivere le risposte e premere [F12] per far leggere la frase alla sintesi vocale.



Questo è reso possibile dalle macro istruzioni programmabili messe a disposizione nel Vocabolary Manager.



In questo modo si è ottenuta la certezza per l'interlocutore udente che la frase comparsa sul video remoto fosse proprio quella voluta. Il sordo può anche sfruttare la sintesi vocale per rispondere per iscritto. Allo stesso modo è possibile impostare un messaggio iniziale che avvisi l'interlocutore del particolare sistema in uso. La macro istruzione consente anche di dare una risposta prestabilita alla frase scelta: ad esempio, se dico "Sono Davide", il sistema, riconoscendo la frase esegua la macro che fa rispondere alla sintesi vocale "Ciao Davide, dimmi pure..".

Questo sistema, ovviamente un po' limitato per un uso quotidiano, ha tuttavia permesso di superare temporaneamente il problema della dipendenza dall'utente e di dare un feedback su quanto riconosciuto dal software concentrandosi sugli altri problemi. Si è anche pensato ad eventuali sviluppi di questo sistema, ovvero la possibilità, tramite una pagina web, di concordare ed aggiornare in base alle esigenze di ognuno, le frasi ammesse, creando un vocabolario personalizzato per ogni utente o tipologia di utenti.

Le prove effettuate hanno evidenziato ancora una volta che il problema della sottotitolazione di una telefonata in diretta è di natura composita: il sistema funziona solamente dopo pazienti regolazioni empiriche dei volumi di ingresso/uscita di modem e scheda audio, e i risultati cambiano drasticamente al variare dell'apparecchio telefonico.

Inoltre collegando l'uscita della scheda audio all'ingresso del modem per trasmettere quanto generato dalla sintesi vocale si generano disturbi e 'rumore elettrico' che infastidiscono molto il riconoscimento. Paradossalmente si ottengono migliori risultati nel trascrivere, con un sistema più complesso e raffinato, un messaggio lasciato in segreteria telefonica piuttosto che riconoscere 'al volo' dei comandi prestabiliti appartenenti ad un vocabolario ristrettissimo. Questo perché nel caso della segreteria telefonica c'è la possibilità di 'trattare' opportunamente il file audio e convertirlo riducendo alcuni disturbi dati dalla linea telefonica. Questo fatto conferma che il problema è la somma di tanti problemi di diversa natura, che affrontati singolarmente trovano soluzioni praticabili, ma che sommandosi originano problemi difficilmente superabili.

4.5 SISTEMI DI VIDEOCONFERENZA E VIDEOTELEFONI

Nell'ambito progetto Voice, ed in particolare in quello delle attività in collaborazione con le associazioni di audiolesi per promuovere l'utilizzo delle nuove tecnologie, sono stati organizzati degli incontri e dei test di funzionamento di alcuni sistemi di videoconferenza e di videotelefoni che sfruttano le potenzialità delle nuove linee ISDN.

La sperimentazione è stata in primo luogo condotta per verificare le possibilità comunicative che vengono offerte alla persona sorda, ovvero verificare la possibilità di utilizzare il linguaggio dei segni o la possibilità della lettura labiale.

A questo scopo è stato invitato al CCR-ISIS un gruppo campione di utenti, composto da soggetti con vari gradi di sordità, di diverse età, alcuni dei quali in grado di parlare fluentemente, altri che utilizzano la lingua dei segni. In questo modo si è testata innanzitutto la fluidità e la definizione dell'immagine trasmessa, che, per la prima volta nella storia di questi apparecchi, si è rivelata sufficiente sia per la lettura labiale che per il linguaggio gestuale. La lettura labiale è da sempre una delle prove più difficili da superare per questi apparecchi, in quanto sia un numero insufficiente di fotogrammi al secondo, sia la scarsa nitidezza data da una bassa risoluzione, causa l'impossibilità di comprensione da parte del sordo.

Gli apparecchi dell'ultima generazione utilizzano entrambi i canali messi a disposizione dalla linea ISDN (in pratica vengono effettuate due telefonate contemporaneamente) che consente così di inviare 128kb di dati al secondo. Inoltre il video viene trasmesso assieme all'audio in formato compresso, secondo un protocollo standard (H320) che garantisce la compatibilità con un vasto numero di apparecchi differenti per caratteristiche prestazioni (da un semplice videotelefono ad un sistema di videoconferenza con telecamere automatiche a controllo vocale). Tale protocollo, oltre a trasmettere dati compressi, ripartisce la banda in base alle esigenze, ovvero in modo 'intelligente'. Se ad esempio si sta trasmettendo l'immagine di un oratore e tutto lo sfondo è fisso, viene trasmessa solo la parte di immagine che ha subito una modificazione ovvero la parte raffigurante il volto, senza inviare i dati relativi allo sfondo. Questo si traduce in un miglioramento della fluidità e della definizione dell'immagine, in particolare delle labbra che sono la parte dell'immagine che maggiormente ci interessa.

L'altra prova che ha impressionato positivamente il pubblico ed i tecnici è stata quella relativa alla sincronizzazione tra l'immagine e l'audio. Uno dei potenziali punti deboli è infatti il possibile ritardo nella trasmissione dell'immagine rispetto a quella del parlato. Se tale ritardo diventa percepibile può disturbare la comprensione da parte di un sordo protesizzato che compensa il deficit uditivo con la lettura labiale.

Parallelamente si è deciso di tentare anche la sottotitolazione della videoconferenza. Un primo test ha riguardato la trasmissione di sottotitoli generati in locale. Da una parte si è messo un gruppo di sordi con un videotelefono, dall'altra un udente con il prototipo Voice già impiegato per le conferenze: un PC dotato del software di riconoscimento vocale e dello specifico programma VOICE meeting che acquisisce l'immagine dalla telecamera che

riprende l'oratore e visualizza sotto di essa i sottotitoli. A questo punto l'uscita video del PC è stata collegata all'ingresso video del videotelefono, che ne esclude la piccola telecamera incorporata. In questo modo attraverso il videotelefono è stata trasmessa l'immagine dell'interlocutore udente corredata dal sottotitolo direttamente sul piccolo display del videotelefono ricevente, o su uno schermo TV ad esso collegato. Così si sono potute valutare le eventuali difficoltà nella lettura del sottotitolo attraverso questo sistema, che si è rivelato ottimo anche per questi impieghi. Inoltre con tale prova sono stati verificati in questa situazione specifica quelli che erano i 'bisogni degli utenti' in materia di sottotitoli, cioè la dimensione, il colore, il tipo di carattere, lo sfondo utilizzati nei sottotitoli. La raccolta dei bisogni degli utenti è una delle importanti funzioni svolte nell'ambito del progetto Voice, che fa poi da tramite fra le associazioni e gli enti che si occupano di sottotitolazione, fornendo utili ed autorevoli indicazioni in materia. Questa sperimentazione tuttavia, essendo basata sulla sottotitolazione 'in locale', ha eliminato tutti i problemi sin qui evidenziati concentrando l'attenzione sulla visualizzazione del sottotitolo, senza però preoccuparsi di come esso viene generato.

La prova di sottotitolazione della videotelefonata 'in remoto', ovvero con il PC con il software di riconoscimento vocale installato presso la postazione del sordo per generare i sottotitoli di ciò che dice il suo interlocutore dall'altra parte del filo, è invece analoga al caso della sottotitolazione della telefonata, e perciò estremamente difficoltosa per tutti i fattori sopra esposti.

Anche se la presenza del doppio canale ISDN e della grande quantità di dati che può trasmettere potrebbero far pensare di avere a disposizione un segnale sonoro qualitativamente superiore a quello del telefono tradizionale, in realtà la situazione non è affatto cambiata. In questi sistemi il protocollo di trasmissione ripartisce la banda a disposizione a seconda dell'occupazione, per cui alla traccia audio (compressa con un formato molto simile all'MP3) viene riservato uno spazio estremamente ridotto (10% della banda) in favore dell'immagine che occupa ovviamente moltissimo spazio.

In futuro se, come sperato, saranno comprese le potenzialità del riconoscimento vocale e l'importanza della sottotitolazione, si potrebbe tentare la via di studiare protocolli alternativi e specifici che, ad esempio, trasmettano l'immagine in B/N anziché a colori e permettano di dedicare una 'fetta' di banda più ampia a favore della qualità audio, migliorando i problemi ad essa connessi, senza tuttavia risolvere tutte le difficoltà già evidenziate parlando di telefonate.

4.6 SOTTOTITOLAZIONE TELEVISIVA IN DIRETTA

Nell'ambito della sperimentazione per la sottotitolazione di una telefonata, sono state poste le basi per future sperimentazioni nella sottotitolazione televisiva. Questa possibilità è nata anche grazie ai contatti con i responsabili del sistema di sottotitolazione Televideo RAI ottenuti durante le conferenze e workshop internazionali organizzati nella prima fase del progetto. I buoni risultati

ottenuti con le conferenze garantiscono interessanti prospettive per sottotitolare documentari o servizi giornalistici, essendoci la possibilità di far eseguire l'addestramento al giornalista prima del servizio, mentre restano molti problemi aperti circa l'impiego in film e trasmissioni caratterizzati da dialoghi veloci e con molte voci diverse.

L'idea, ancora allo stato embrionale, è quella di utilizzare un comune sistema di riconoscimento vocale per PC per sottotitolare un giornalista mentre legge un notiziario tv, correggendo in tempo reale il testo prodotto e mandando in onda poi il tutto in lieve differita (30 sec - 2 min).

Riuscire a proporre una dimostrazione credibile in tal senso significherebbe andare ben oltre la semplice sensibilizzazione al problema di fronte ai responsabili della sottotitolazione.

Il sistema prevede il riconoscimento vocale di quanto detto dallo speaker televisivo, l'invio della trascrizione ad una persona che velocemente ed in tempo reale corregga i pochi errori commessi da sistema automatico, mandando il tutto in onda pochi istanti dopo. Questa sarebbe una grandissima innovazione nell'attuale processo di creazione dei sottotitoli, che porterebbe a risparmiare tempo, persone, e denaro, con la possibilità di offrire al pubblico un numero molto maggiore di trasmissioni sottotitolate (oggi la RAI sottottitola in media un film ogni giorno ed un telegiornale).

I punti di forza di questo sistema stanno nelle seguenti considerazioni:

- gli speaker televisivi sono professionisti con una dizione chiara, molto migliore di quella dei normali utenti dei prodotti di riconoscimento vocale;

- la qualità sonora dell'input audio che viene dato al riconoscimento vocale è ottima, grazie alle apparecchiature utilizzate in televisione;

- il testo del notiziario è disponibile prima dell'inizio della diretta TV ed è quindi possibile farlo analizzare al riconoscimento vocale che eventualmente richiederà l'addestramento delle poche parole nuove in esso contenute;

Con questi presupposti ci si può aspettare ragionevolmente di avere una trascrizione pressoché priva di errori, per cui il lavoro del correttore di bozze dovrebbe essere estremamente limitato, tanto da consentire una differita che va da i due minuti ai trenta secondi.

I test di questo prototipo sono stati rimandati a causa della momentanea indisponibilità di un apparecchio specifico per realizzare la differita, apparecchio professionale estremamente costoso, che hanno in dotazione solamente gli studi televisivi dei network più grossi. Sono perciò in corso degli accordi per poter effettuare delle prove presso gli studi televisivi di Rai o di Mediaset.

Questo progetto rimane tuttavia un esempio estremamente significativo delle potenzialità non ancora sfruttate dei sistemi di riconoscimento vocale, e per questo si pone come uno dei principali possibili sviluppi futuri dello spirito che ha animato il progetto Voice.