

Università degli Studi di Padova
Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica

TESINA

STATO DELL'ARTE DELL'ACCESSIBILITÀ DI STRUMENTI
E SISTEMI SOFTWARE OPEN SOURCE

Relatore: Professor Franco Bombi
Correlatore: Dottor Roberto Valli

Laureando: Hubert Fornasiero

Anno Accademico 2009/2010

Indice

Introduzione.....	4
CAPITOLO 1 Disabilità ed Accessibilità.....	6
1.1 Disabilità.....	6
1.2 Accessibilità.....	7
1.3 Disabilità e strumenti software.....	8
1.3.1 Ausili per disabilità visiva.....	8
1.3.2 Ausili per disabilità uditiva.....	9
1.3.3 Ausili per disabilità motoria.....	9
1.3.4 Ausili per disabilità cognitiva.....	10
1.4 Legislazione italiana.....	11
CAPITOLO 2 Strumenti di accessibilità in ambiente grafico GNOME su sistema operativo GNU/Linux.....	12
CAPITOLO 3 Strumenti di accessibilità in ambiente grafico KDE su sistema operativo GNU/Linux.....	22
CAPITOLO 4 Approfondimento sulla tecnologia TTS (Text-to-Speech).....	26
CAPITOLO 5 Prove di configurazione TTS FESTIVAL ed installazione MBROLA.....	34
CONCLUSIONI.....	38
Biblio/Sitografia.....	40

Introduzione

L'indagine è rivolta a conoscere lo stato dell'arte degli strumenti software di supporto alle tecnologie assistive in ambito open source.

Nel capitolo 1 si darà una fotografia delle principali disabilità e dei loro effetti per metterle in corrispondenza con le tecnologie assistive usate per consentire ai portatori di tali disabilità di interagire con un computer.

Nei capitoli 2 e 3, verranno proposti alcuni esempi delle applicazioni rese disponibili dalle interfacce grafiche dei sistemi operativi GNU/Linux.

Nel capitolo 4 si approfondiranno alcuni concetti teorici riguardanti la tecnologia TTS (Text-to-Speech).

Infine nel capitolo 5 verranno illustrati alcuni aspetti pratici della configurazione dei sistemi TTS riguardanti soprattutto la possibilità di installare altre lingue (nel caso specifico l'italiano) oltre alle lingue di default (normalmente varie versioni della lingua inglese).

CAPITOLO 1 Disabilità ed Accessibilità

Al fine di introdurre le tecnologie assistive in ambito open source, verranno prima esposti i concetti di disabilità ed accessibilità, per poi mettere in relazione le disabilità con gli ausili maggiormente usati per far fronte ad esse. Infine si accennerà all'evoluzione della legislazione italiana in ambito di disabilità ed accessibilità¹.

1.1 Disabilità

Negli anni '80 l'Organizzazione Mondiale della Sanità ha rivisto il concetto di disabilità definendola come una condizione dell'uomo, in cui chiunque può ritrovarsi anche temporaneamente: *incapacità di compiere un'attività nella maniera o a livello considerato normale per un essere umano*.

Negli stessi anni, sul fronte, dell'innovazione tecnologica, l'attenzione si spostava dal piano della persona a quello dello specifico bisogno in cui la persona può trovarsi, dando luogo ad una visione più pratica: se non si è in grado di scrivere una lettera o di aprire una porta, il problema non è dovuto alla propria limitazione, ma alla mancanza di uno strumento tecnico alternativo che consenta di compiere quelle operazioni. Nasce così il concetto di persona con bisogni speciali. La diversità di chi ha bisogni speciali non è più considerata un deficit, ma il punto di partenza per incrementare l'attenzione verso chi anche solo per un periodo temporaneo ha necessità di aiuto per migliorare la qualità della vita.

Sulla stessa linea l'Organizzazione Mondiale della Sanità con la pubblicazione nel 2001 dello storico documento ICF (Classificazione Internazionale della Disabilità, del Funzionamento Umano e della Salute) definisce ora la disabilità non come un attributo della persona, ma come una situazione che nasce dal divario tra lo stato di salute di quella determinata persona e i fattori contestuali dell'ambiente ove la persona vive. Tale situazione può portare ad una limitazione alle proprie capacità di compiere un'attività o al grado di partecipazione che posso esercitare in tale attività.

Nel modello ICF, la tecnologia è il primo nell'ordine di classificazione, tra i fattori contestuali ed ha un ruolo chiave nel rimuovere le situazioni di disabilità. Se una persona usa le mani per scrivere con una penna biro ed una seconda invece, essendo impedita nell'uso delle mani, usa un computer per lo stesso scopo, nessuna delle due è disabile rispetto a questa attività. Se la seconda persona non ha il computer e la prima non ha la penna, entrambe constateranno la propria disabilità, nel senso di essere limitate in tale attività. Ciò che le differenzia è lo strumento (penna o computer) che per la prima persona è questione di scelta, per la seconda di necessità.

E' un nuovo approccio culturale, che colloca la disabilità come una questione di rispetto dei diritti umani. La salute è l'abilità di vivere la propria vita nelle sue piene

¹ http://www.portale.siva.it/servizi/guida/siva_ITA.htm

potenzialità. Ci sono persone per le quali la realizzazione di tale abilità dipende da fattori dipendenti dalla società. Per esempio, se una persona in carrozzina incontra difficoltà a lavorare in un particolare edificio perché sprovvisto di rampe o ascensori, il modello ICF identifica il problema nell'inadeguatezza dell'edificio e non nell'handicap della persona obbligandola a trovare un altro lavoro.

1.2 Accessibilità

Definizione di accessibilità tratta dal Piano di Lavoro del Programma TIDE: Commissione Europea 1995.

"Fruibilità dell'ambiente costruito, dei prodotti e dei servizi di uso generale, da parte della più ampia percentuale di popolazione, incluse le persone anziane e le persone con disabilità."

Uno scalino sul percorso, che impedisce di proseguire il cammino. Un interruttore troppo in alto, che rende inutile l'automazione dell'impianto. Un'informazione inaccessibile, che vieta di andare oltre: sono infiniti e di natura molto varia gli ostacoli che incontra ogni giorno nell'ambiente chi ha limitazioni fisiche, sensoriali o cognitive.

Il concetto di *progettazione senza barriere*, o *progettazione accessibile*, o *progettazione per un'utenza ampliata*, tende oggi a ricadere sempre più nell'ambito disciplinare dell'ergonomia, non limitandosi tra l'altro al solo mondo reale ma estendendosi al mondo virtuale (si pensi ad es. ad applicazioni software progettate in modo che chiunque possa utilizzarle, comprese persone non vedenti o con difficoltà motorie). Anche in materia di ostacoli informatici, infatti, ci sono direttive che vengono in aiuto. Le linee progettuali, dettate dallo standard W3C (World Wide Web Consortium) nella WCAG (Web Content Accessibility Guidelines), indicano quale strada seguire per costruire un portale accessibile, in cui le informazioni siano consultabili da persone con bisogni speciali.

E' importante citare l'articolo 2 della legge Stanca dove si definiscono accessibilità e tecnologie assistive:

“Ai fini della presente legge, si intende per:

- *accessibilità*: la capacità dei sistemi informatici, nelle forme e nei limiti consentiti dalle conoscenze tecnologiche, di erogare servizi e fornire informazioni fruibili, senza discriminazioni, anche da parte di coloro che a causa di disabilità necessitano di tecnologie assistive o configurazioni particolari;
- *tecnologie assistive*: gli strumenti e le soluzioni tecniche, hardware e software, che permettono alla persona disabile, superando o riducendo le condizioni di svantaggio, di accedere alle informazioni e ai servizi erogati dai sistemi informatici.”

1.3 Disabilità e strumenti software

Verrà di seguito illustrato quali ausili vengono impiegati in relazione alle singole disabilità e come in certi casi il computer stesso venga usato come mezzo educativo.

1.3.1 Ausili per disabilità visiva

Nel caso di non vedenti, il senso mancante deve essere sostituito da altri. L'informazione scritta deve cioè essere trasferita su un diverso canale sensoriale: il tatto o l'udito. I dispositivi maggiormente utilizzati a tale scopo sono la sintesi vocale e lo screen reader.

La sintesi vocale è un apparecchio esterno/interno al computer che è in grado di farlo parlare, trasformando in suono il contenuto dello schermo. E' composto da una parte hardware (scheda audio, amplificatore, diffusori acustici) e da una parte software. E' compito di quest'ultima definire le regole di pronuncia delle parole (ogni sintesi vocale dovrà essere adattata alla lingua usata) e adeguarsi alle diverse esigenze dell'utente. Un testo può essere letto in vari modi. Si può avere una lettura per singolo carattere, oppure parola per parola, oppure ancora riga per riga. Inoltre si può avere una lettura continua scorrevole, con punteggiatura, con indicazione degli attributi del testo (lettere maiuscole, corsivi, sottolineature), con spelling integrale.

Lo screen reader è un programma software che è in grado di stabilire quale parte dello schermo debba essere evidenziata sulla barra Braille o letta dalla sintesi vocale. La funzione dello screen reader viene completata, dalla presenza (sulla tastiera del PC e sulla barra Braille) di comandi che portano ad evidenziare o ad ascoltare ciò che si desidera: righe, caratteri, parole o parti importanti di una finestra.

Diverso è il caso di persone con una ridotta capacità visiva o ipovedenti. Ad esse la flessibilità del computer consente in modo dinamico sia l'ingrandimento che la variazione di colori e di sfondi di quanto appare sullo schermo.

Gli ingranditori per computer sono essenzialmente programmi residenti in memoria che aumentano le dimensioni dei caratteri sul monitor permettendo in tal modo la visione a persone con gravi minorazioni visive. Non richiedendo apparecchiature particolari, sono installabili su qualsiasi computer. L'ingrandimento riduce la porzione di schermo che può essere consultata. Con un sistema di ricerca (comandato in genere da un mouse) è possibile selezionare la parte del video che interessa. Alcuni hanno la possibilità di essere collegati a sistemi ingrandenti del video (costituiti da una camera a circuito chiuso). Funzionano a schermo diviso. Su una parte apparirà l'immagine del libro ingrandito, sull'altra i caratteri che vengono digitati al computer. In questo modo l'utente ipovedente può procedere alla digitazione di un testo senza mai togliere lo sguardo dallo schermo. Per essere efficace il sistema deve però essere integrato da un leggio elettronico. Si tratta di un apparecchio che sposta automaticamente il testo da leggere sotto l'obiettivo della telecamera alla velocità e nel modo desiderato.

1.3.2 Ausili per disabilità uditiva

Un uso particolarmente efficace in campo riabilitativo del mezzo informatico si è avuto a favore dei sordi profondi. Questi sono coloro che essendo nati privi di udito o avendolo perso in modo totale hanno problemi nell'imparare a parlare o nel regolare le loro emissioni sonore. Infatti essi sono privi del canale sensoriale di riscontro, cioè della possibilità di ascoltarsi mentre parlano. In questo caso, per ausili si identificano dei pacchetti software educativi specifici. In ambito educativo un software adeguato può essere di aiuto per un lavoro sia di gruppo che individuale. Esistono diversi tipi di software tra cui:

- software per facilitare l'apprendimento della letto-scrittura;
- software per apprendere in modo interattivo con il supporto delle immagini la composizione di frasi elementari;
- software per lo sviluppo e potenziamento delle frasi locative;
- software diagnostici, che consentono di effettuare uno screening sulla capacità uditiva dei bambini a partire dalla scuola materna;
- software riabilitativi, che consentono di programmare sessioni di riabilitazione in cui il bambino o l'adulto sordo possono interagire col computer per mezzo di un microfono e avere in risposta un feedback visivo dei segnali emessi.

Un altro aiuto ai non udenti viene dai riconoscitori del parlato, strumenti informatici in grado di ascoltare e trascrivere le parole pronunciate dall'uomo. In questi casi l'aiuto al non udente consiste nel fare apparire sullo schermo del computer la trascrizione di quanto pronunciato al microfono.

1.3.3 Ausili per disabilità motoria

Per quanto riguarda la accessibilità al computer, i problemi maggiori si presentano per i deficit agli arti superiori e riguardano soprattutto l'uso della tastiera e del mouse. Esempi di questi problemi sono la necessità di introdurre tutti i caratteri usando un solo dito e un solo tasto, oppure il dover usare al posto del dito una leva applicata a un casco, o ancora la facilità di commettere errori involontari dovuti a tremolio della mano o alla pressione troppo prolungata del tasto e anche la difficoltà di avere stabilità e precisione nel dirigere il mouse.

Per i casi meno difficili, l'interfaccia utente del sistema operativo presenta alcune soluzioni, quando si utilizza la funzione Accesso facilitato. Questa funzione consente di ritardare o evitare la ripetizione di ogni tasto, di premere separatamente tasti che di solito devono essere premuti insieme, di usare i tasti di spostamento del cursore al posto del mouse e così via.

Se l'utente non è in grado di gestire la tastiera in modo diretto, occorre passare a degli strumenti di input alternativo. Due sono, attualmente, le strade percorribili: i sistemi a scansione e l'immissione a voce.

I sistemi a scansione si servono di un numero limitato di tasti: da uno solo, ad un massimo di cinque o sei. La scrittura non viene quindi effettuata in modo diretto (un tasto per carattere) ma attraverso un procedimento di selezione e conferma. Si tratta di un sistema che avrebbe possibilità di scrittura illimitate. In pratica però viene fortemente condizionato dalla sua lentezza di esecuzione, dovuta ai prolungati tempi di attesa o alla macchinosità di certe operazioni. Sono state ideate negli ultimi anni varie soluzioni, più o meno efficaci, per accelerare la scrittura ma in genere, quello che si guadagna in velocità va tuttavia a scapito della semplicità d'uso.

Nei sistemi comandati a voce al computer viene applicato un microfono, una scheda audio e un software di riconoscimento vocale. In questo modo si consente di riconoscere un certo numero di parole dettate dall'utente e di associarle a comandi relativi al sistema operativo o alle particolari applicazioni. Attualmente i sistemi di riconoscimento sono evoluti al punto di consentire anche la dettatura e trascrizione di testi usando vocabolari di parole molto vasti, praticamente illimitati. Con un computer che riconosce un limitato numero di comandi una persona può comandare un sistema di automazione dell'ambiente in cui si trova (domotica) o la carrozzina di cui si serve. Con un riconoscitore a vocabolario illimitato possono scrivere documenti. Per svolgere la stessa funzione di un singolo tasto si possono infine utilizzare apparecchi alternativi detti sensori. Si differenziano fra loro per la modalità di attivazione (pressione, ma anche spostamento, scuotimento, tocco, soffio), per la forma e dimensione, per il tipo o la forza di movimento richiesto ecc. La gamma di sensori diversi disponibili presso le ausilioteche o ditte specializzate è vastissima, adattabile a qualsiasi capacità motoria residua, purché volontaria.

1.3.4 Ausili per disabilità cognitiva

Nel caso di menomazioni della capacità intellettuale e psicologiche in genere, i problemi di accessibilità riguardano la padronanza logica delle operazioni che si eseguono. Per gli altri deficit il ricorso alla tecnologia informatica rappresenta una specie di protesi, ovvero uno strumento che consente di sopperire ad una funzione organica compromessa. In tal caso il criterio di valutazione è semplicemente quello della funzionalità. Come tale, può essere direttamente accertato dall'utente finale, mentre altrettanto non si può dire nel caso del ritardo mentale.

In questo caso non è il disabile che si serve del computer o che impara dal computer con maggiore o minore efficacia. E' l'insegnante o il terapeuta che si serve di questo apparecchio per interagire didatticamente con la persona disabile.

Per far questo egli deve tuttavia essere in grado di dominare la macchina, ovvero di gestirla e adattarla al proprio progetto didattico: obiettivi, contenuti, linguaggio, tempi, verifiche.

1.4 Legislazione italiana

Nel 1992 viene pubblicata sulla Gazzetta Ufficiale la legge 104/92².

Si tratta di una legge quadro per l'assistenza, l'integrazione ed i diritti delle persone handicappate. La legge affronta in maniera organica tutte le problematiche dell'handicap e rappresenta ancora oggi un punto di riferimento fondamentale per il raggiungimento della qualità dell'integrazione scolastica e lavorativa delle persone con disabilità. Ha come fine principale la rimozione delle cause invalidanti e l'integrazione soprattutto attraverso la ricerca scientifica.

Un esempio importante viene dall'articolo 20 riguardante le prove d'esame nei concorsi pubblici e per l'abilitazione alle professioni, in cui si dà diritto al candidato di specificare nella domanda di ammissione l'eventuale necessità e tipologia di ausili nonché di tempi aggiuntivi per lo svolgimento della prova.

Successivamente con la legge 17 del 1999 sono state introdotte alcune modifiche riguardanti l'istruzione universitaria:

- il diritto ad ausili tecnici per studenti portatori di handicap,
- il diritto ad un tutorato,
- il diritto ad un trattamento individualizzato (previo accordo con il docente) per il superamento degli esami,
- l'istituzione di un docente delegato con funzioni di coordinamento.

Successive modifiche hanno riguardato soprattutto aspetti non pertinenti con questo elaborato come permessi e congedi parentali.

Un'altra importante legge è la legge 4 del 2004 (denominata legge Stanca) contenente disposizioni per favorire l'accesso agli strumenti informatici da parte di utenti disabili, questa legge costituisce un buon punto di partenza anche se mostra alcuni limiti:

- si applica solo ai soggetti di cui all'articolo 3 (enti pubblici ed aziende private concessionarie di servizi pubblici),
- la normativa specifica che il Cnipa valuta il mantenimento del rispetto dei requisiti, non il loro rispetto effettivo. La conseguenza è che la legge non prevede un organismo per la verifica del rispetto dei requisiti, che sono auto certificati, e dunque non controllati da un ente terzo.

² http://www.pubbliaccesso.gov.it/normative/legge_20040109_n4.htm

<http://www.governo.it>

<http://www.parlamento.it>

CAPITOLO 2 Strumenti di accessibilità in ambiente grafico GNOME su sistema operativo GNU/Linux

GNOME³ è il nome di un progetto atto a fornire agli utenti un desktop grafico intuitivo ed una piattaforma per sviluppare applicazioni che si integrino con il desktop. GNOME è un software libero e fa parte del progetto GNU. Il software base GNOME rispetta i principi delle Human Interface Guideline a favore dell'usabilità⁴. GNOME pone molta attenzione anche all'accessibilità e per venire incontro a tali esigenze è stato creato il GNOME Accessibility Project⁵.

Dalla versione 2.0 in poi sono state introdotte migliorie nella navigazione della tastiera (dando la possibilità di muoversi tra tutte le icone del desktop e nel gestore file con la pressione di un solo tasto); nella navigazione del mouse (possibilità di click automatico facendo sostare il puntatore su un'area speciale della finestra applicazione); nel lettore di schermo (tramite la collaborazione con terze parti).

Parte integrante del progetto GNOME è la tecnologia assistiva ORCA⁶, un software diretto ad aiutare persone con varie disabilità, che si compone di screen reader, braille, screen magnifier ed altri ausili software. ORCA si interfaccia con applicazioni che supportano il toolkit AT-SPI (Assistive Technology Service Provider Interface) sviluppato anch'esso nell'ambito del progetto GNOME.

Per le prove svolte durante la scrittura di questo elaborato, è stato installato il sistema operativo open source FEDORA 13 su un PC portatile IBM modello T30 con processore Pentium 4 ed 1GB di RAM.

Si procederà illustrando l'impostazione del lettore di schermo ORCA cui si può accedere come da figura 2.1 selezionando in sequenza i menu *Sistema->Preferenze->Tecnologie Assistive*.

³<http://www.it.gnome.org/index.php/Home>

⁴L'usabilità è definita dall'ISO (International Organisation for Standardisation), come l'efficacia, l'efficienza e la soddisfazione con le quali determinati utenti raggiungono determinati obiettivi in determinati contesti. In pratica definisce il grado di facilità e soddisfazione con cui l'interazione uomo-strumento si compie. Il termine non si riferisce ad una caratteristica intrinseca dello strumento, quanto al processo di interazione tra classi di utenti, prodotto e finalità.

Usabilità. (7 maggio 2010). *Wikipedia, L'enciclopedia libera*. Tratto il 5 luglio 2010, 08:40 da <http://it.wikipedia.org/w/index.php?title=Usabilit%C3%A0&oldid=31923508>.

⁵ <http://projects.gnome.org/accessibility/>

⁶ <http://live.gnome.org/Orca>

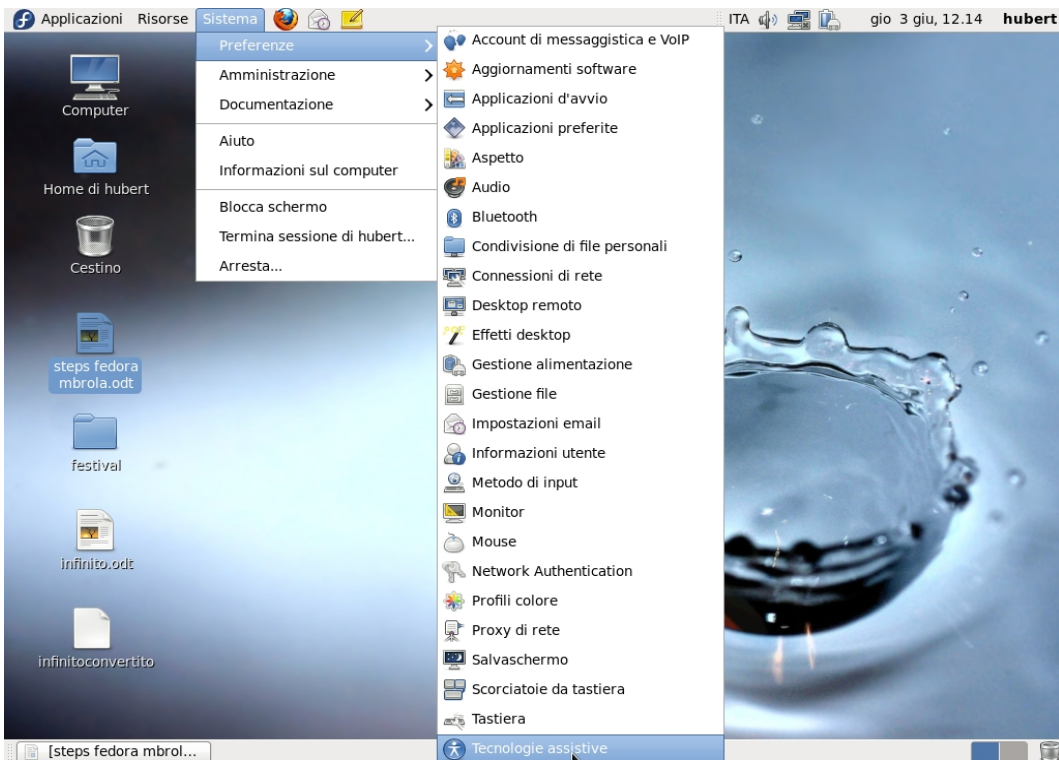


Figura 2.1

nella schermata successiva illustrata in figura 2.2, premendo il pulsante Applicazioni preferite, si apre una schermata in cui si può porre la flag sull'esecuzione di ORCA all'avvio.

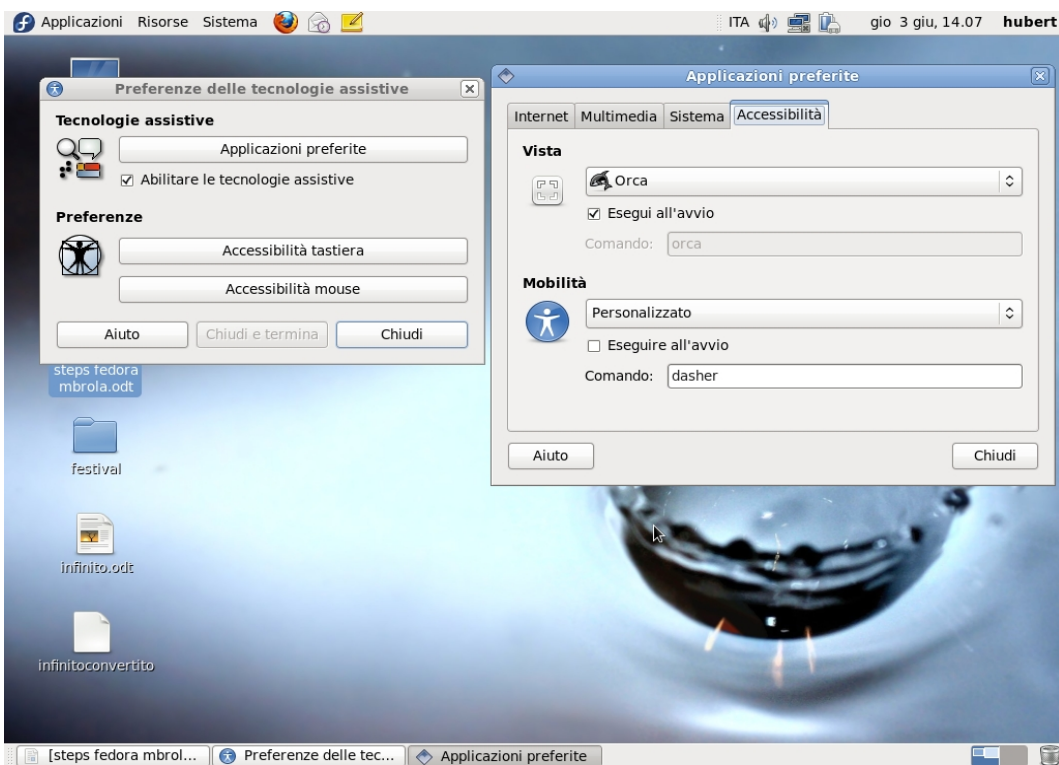


Figura 2.2

al successivo login apparirà la finestra di figura 2.3,

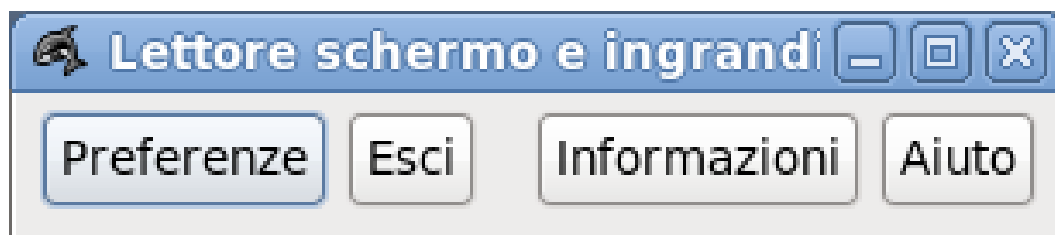


Figura 2.3

cliccando sul pulsante *Preferenze* apparirà la schermata di figura 2.4 con tutte le opzioni di attivabili e potremo lavorare sulle impostazioni di ORCA,

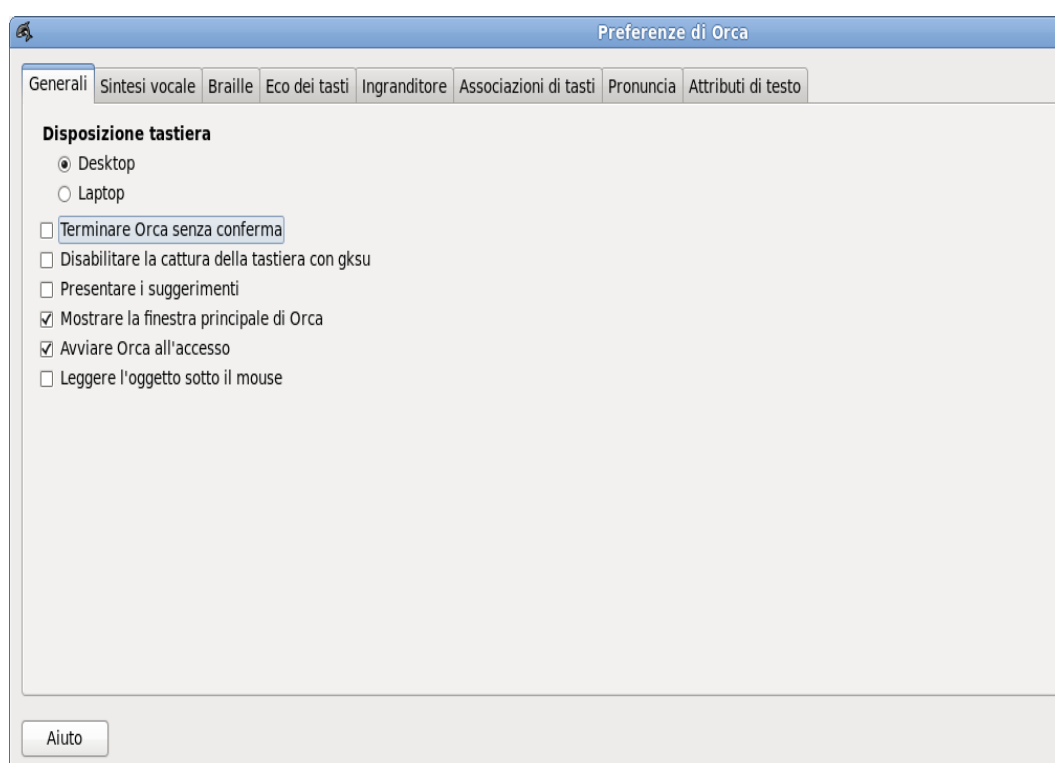


Figura 2.4

Nella schermata di figura 2.4 è possibile tra le altre cose selezionare il tipo di tastiera (desktop/laptop) e attivare la lettura degli oggetti che passano sotto il mouse.

Nella finestra riguardante la sintesi vocale riportata in figura 2.5, è possibile la scelta del sintetizzatore preferito, la voce desiderata e fornire alcune altre impostazioni riguardanti la prolissità, parametro quest'ultimo che influenza il comportamento del lettore in presenza di alcuni dettagli delle finestre delle applicazioni, dei menu a bandierina e della punteggiatura.

Nel nostro caso abbiamo selezionato la voce femminile di FESTIVAL⁷ per MBROLA⁸.

⁷<http://www.cstr.ed.ac.uk/projects/festival/>.

<http://www2.pd.istc.cnr.it/festival/>

⁸ <http://tcts.fpms.ac.be/synthesis/>

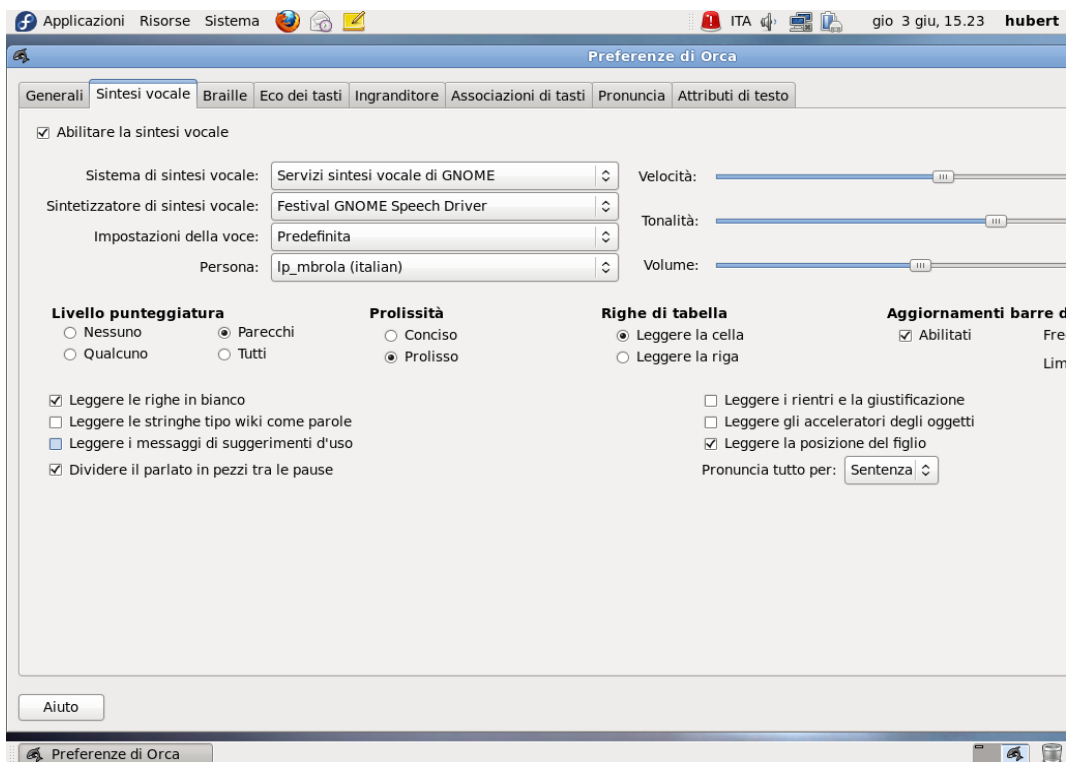


Figura 2.5

E' inoltre possibile come in figura 2.6 abilitare l'eco dei tasti, sempre da questa schermata si ha la possibilità di settare il comportamento dello screen reader in fase di scrittura con un editor di testo, modificando l'eco che si può avere sul singolo carattere (che verrà letto non appena digitato), oppure sulla parola (che verrà letta appena si inserirà uno spazio bianco) e anche sull'intera frase appena scritta.

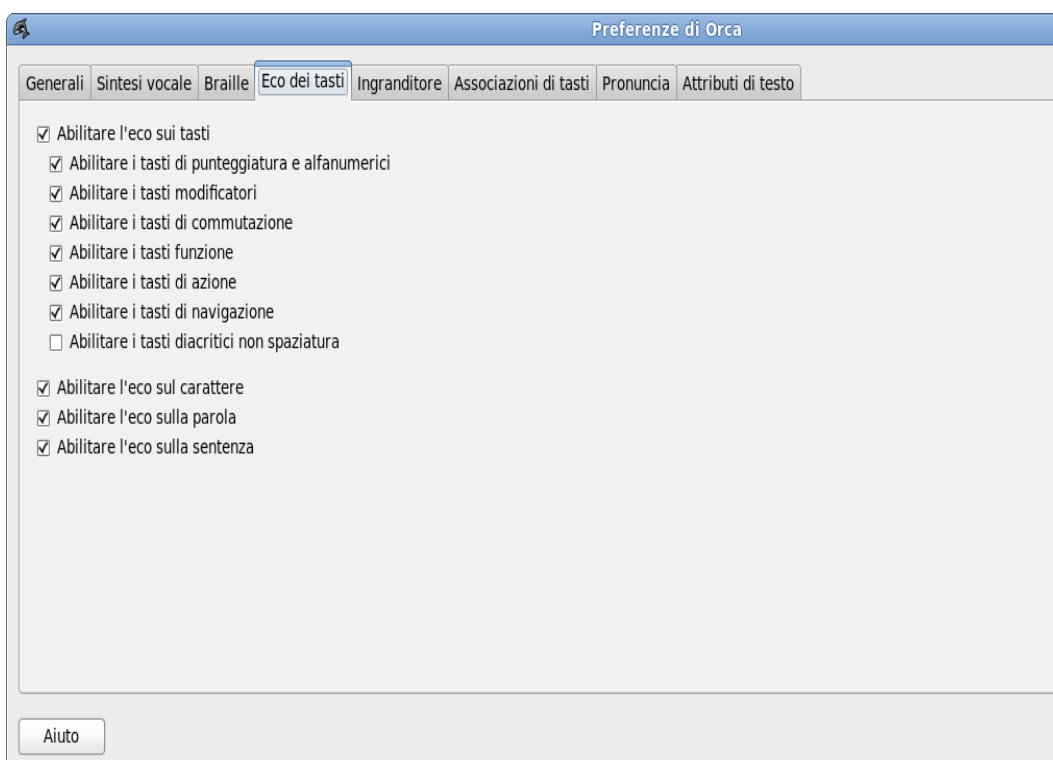


Figura 2.6

Un'altra feature importante è l'ingranditore di schermo di cui si può vedere la finestra di setup in figura 2.7,

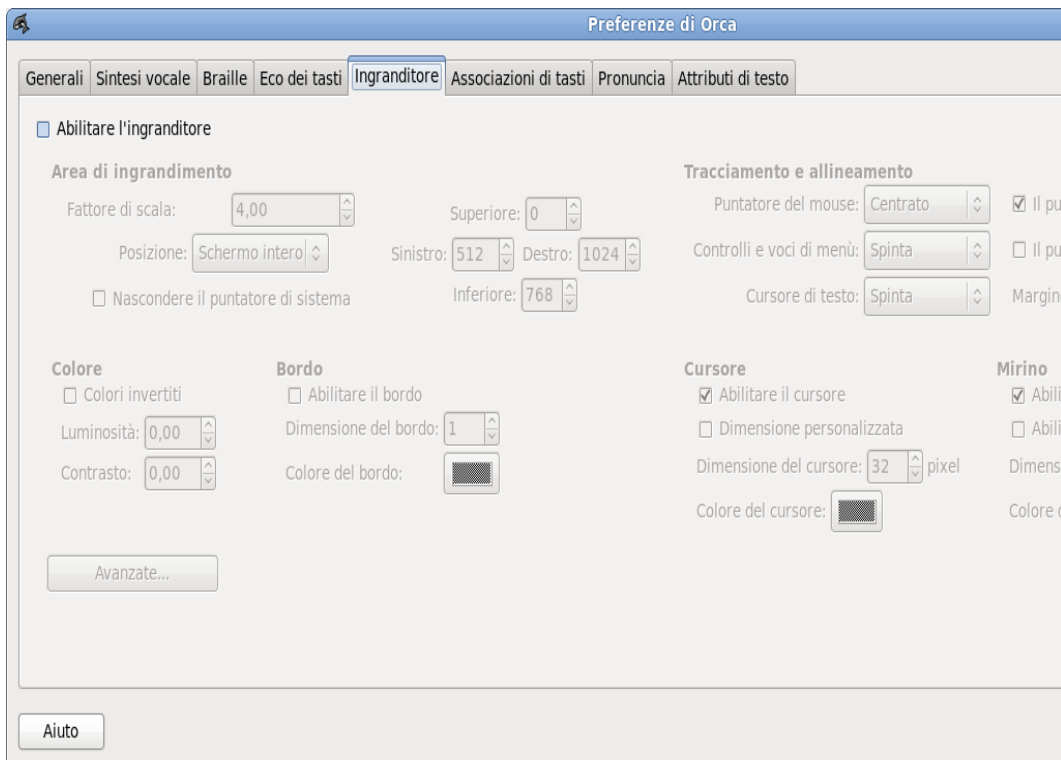


Figura 2.7

l'abilitazione dell'ingranditore con un fattore di scala 4 su schermo intero porta al seguente effetto illustrato in figura 2.8,

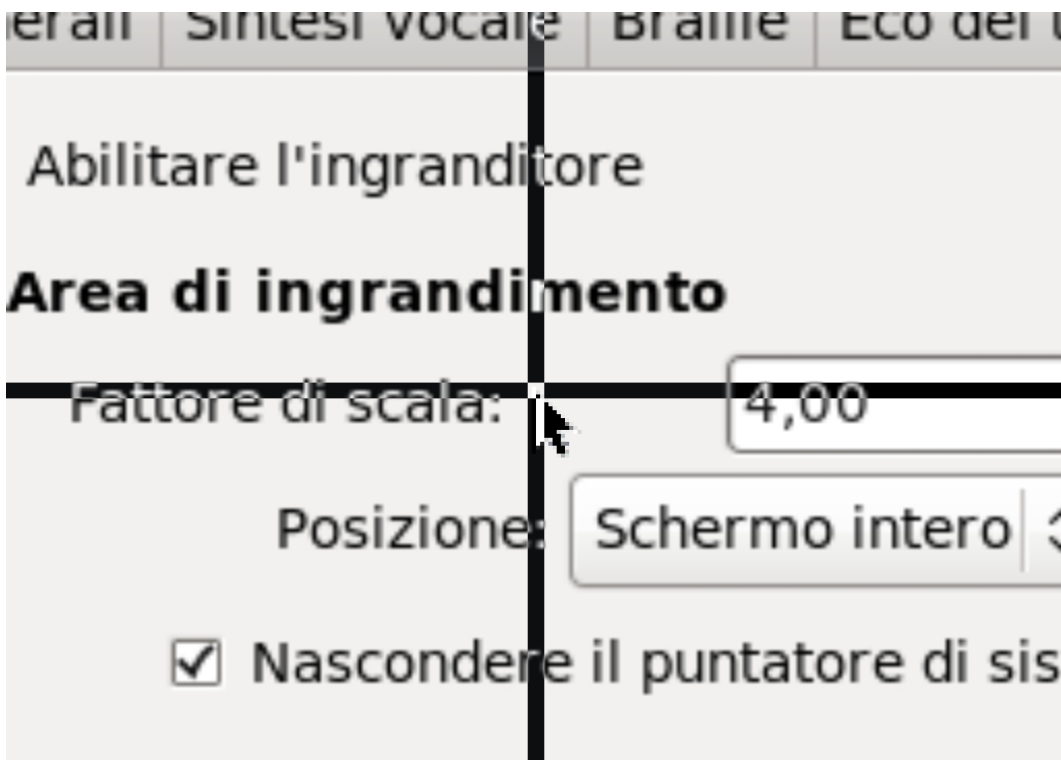


Figura 2.8

Venendo agli ausili per disabilità motorie, è possibile dal menu *Sistema->Preferenze->Tastiera* aprire le schermate di figura 2.9 e 2.10 che consentono di impostare il comportamento della tastiera, regolando la sensibilità dei tasti (ad esempio per evitare pressioni ripetute dovute e tremolio) o simulando le combinazioni di tasti per aiutare utenti impossibilitati a premere più tasti contemporaneamente,

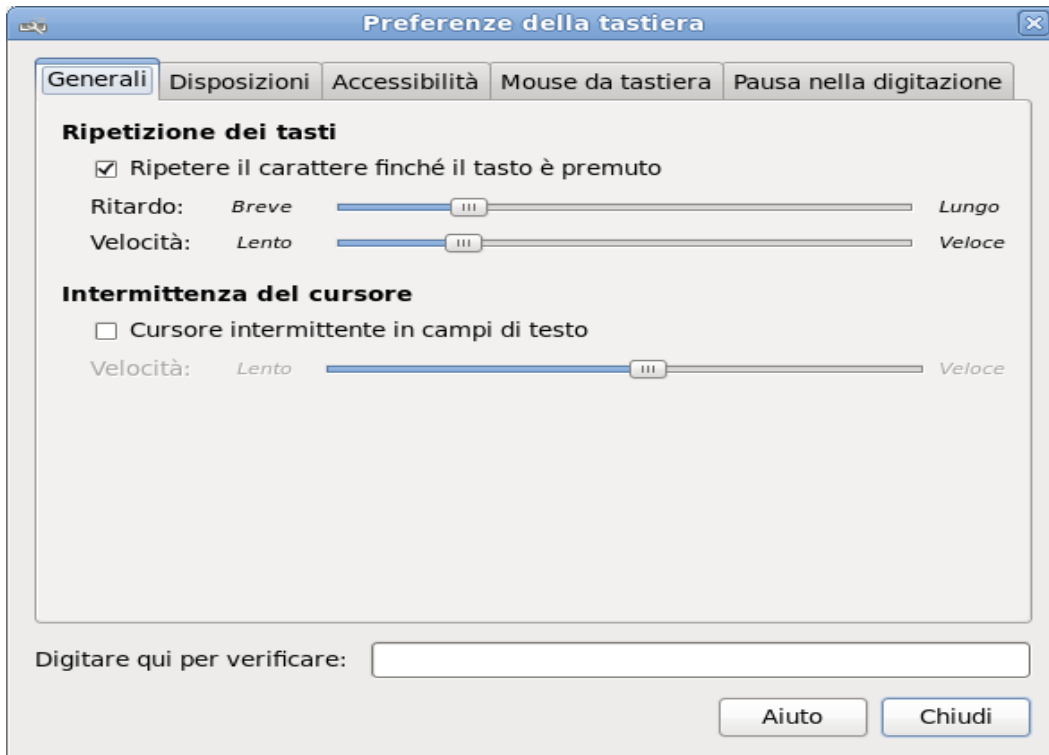


Figura 2.9

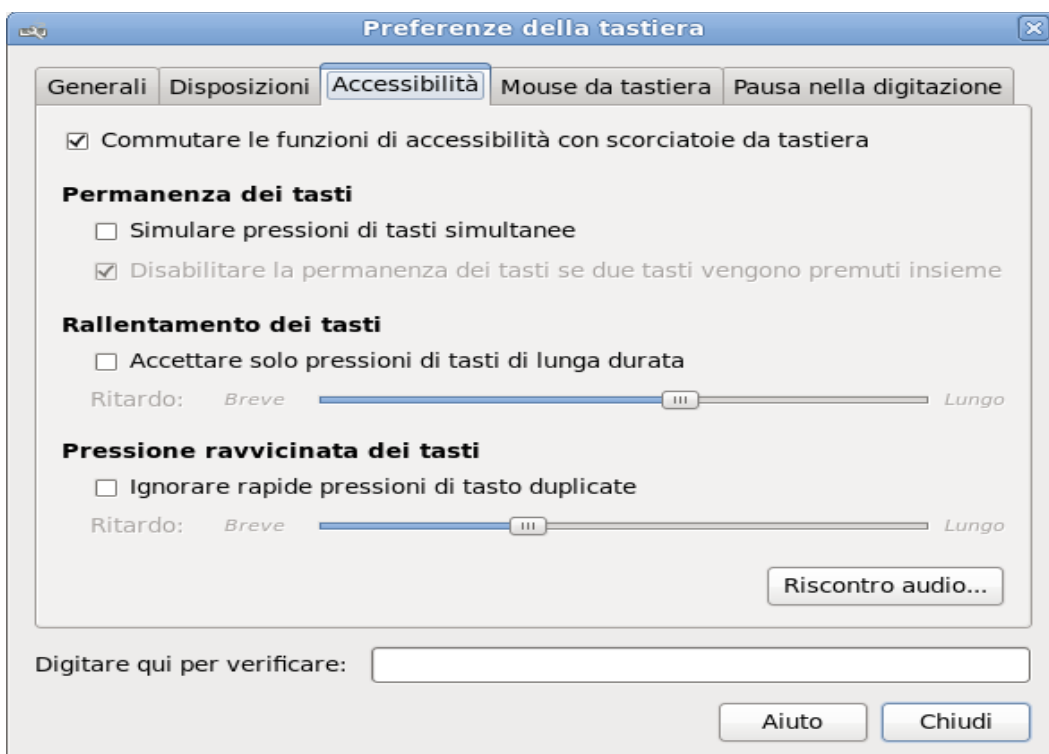


Figura 2.10

Sempre per venire incontro alle disabilità motorie sul desktop GNOME è disponibile anche la tastiera virtuale GOK:

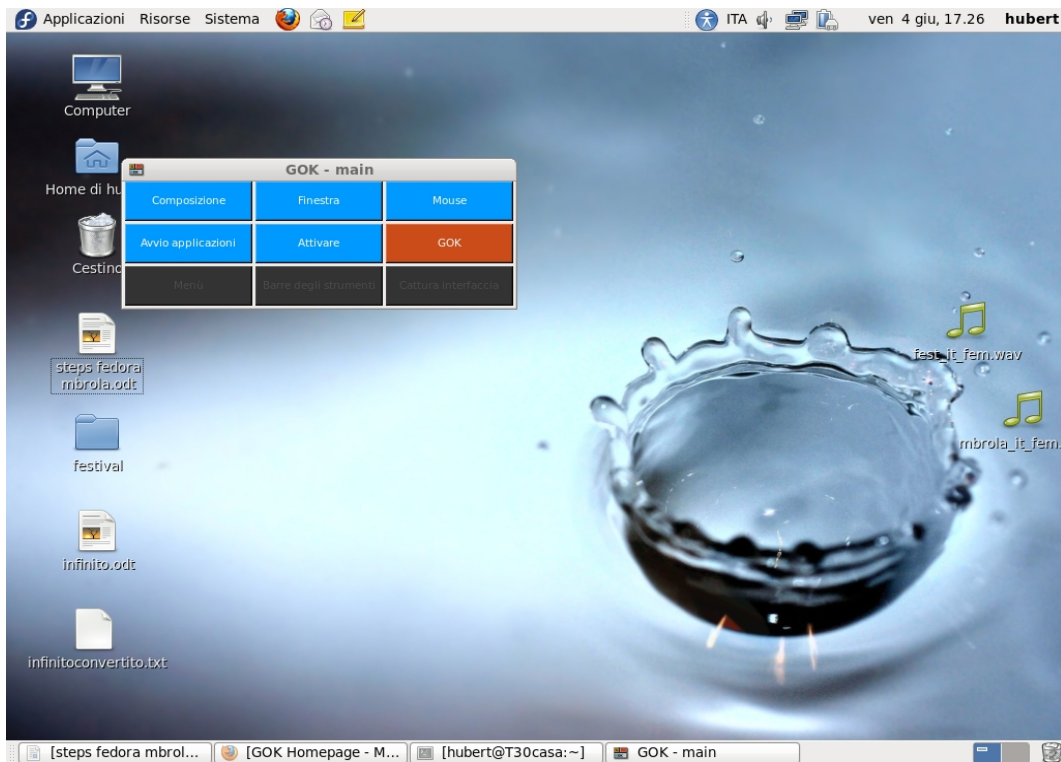


Figura 2.11

Premendo nella finestra GOK-main mostrata in figura 2.11 il pulsante *Avvio Applicazioni*, si può scegliere di avviare un'applicazione a scelta,

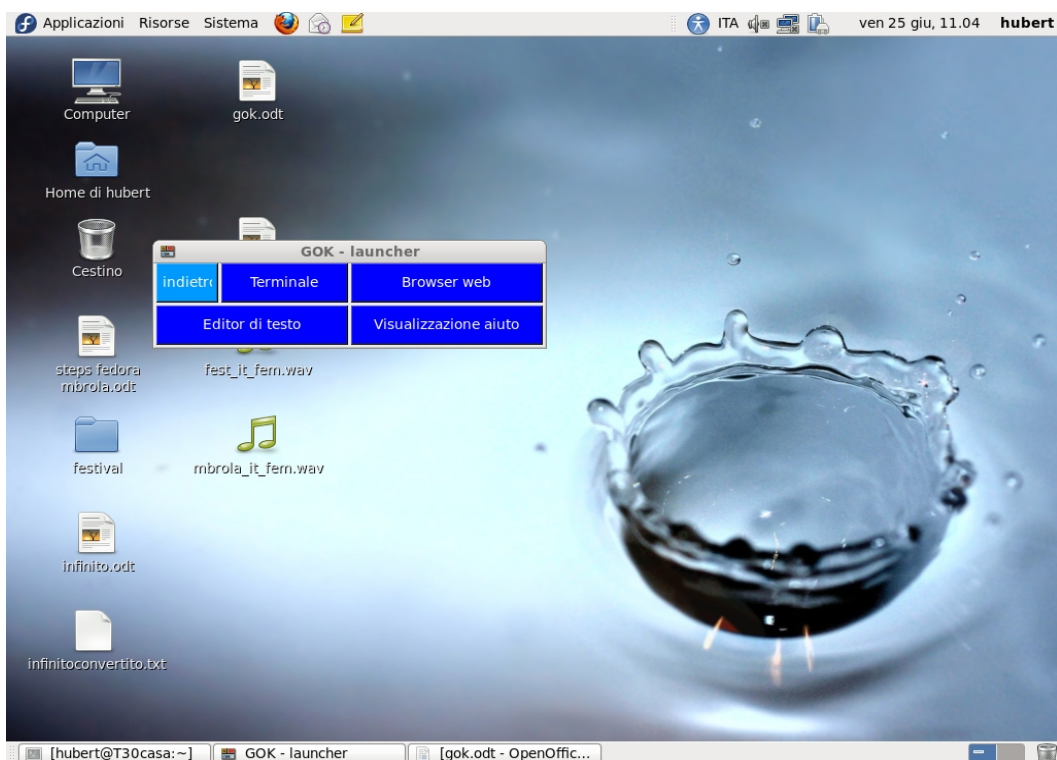


Figura 2.12

In figura 2.13 si può invece vedere l'effetto della pressione del tasto destro del mouse su un'icona del desktop con GOK abilitato, infatti GOK stesso offre la possibilità di operare sull'oggetto selezionato riportando sulla tastiera virtuale le opzioni del menu a bandierina:

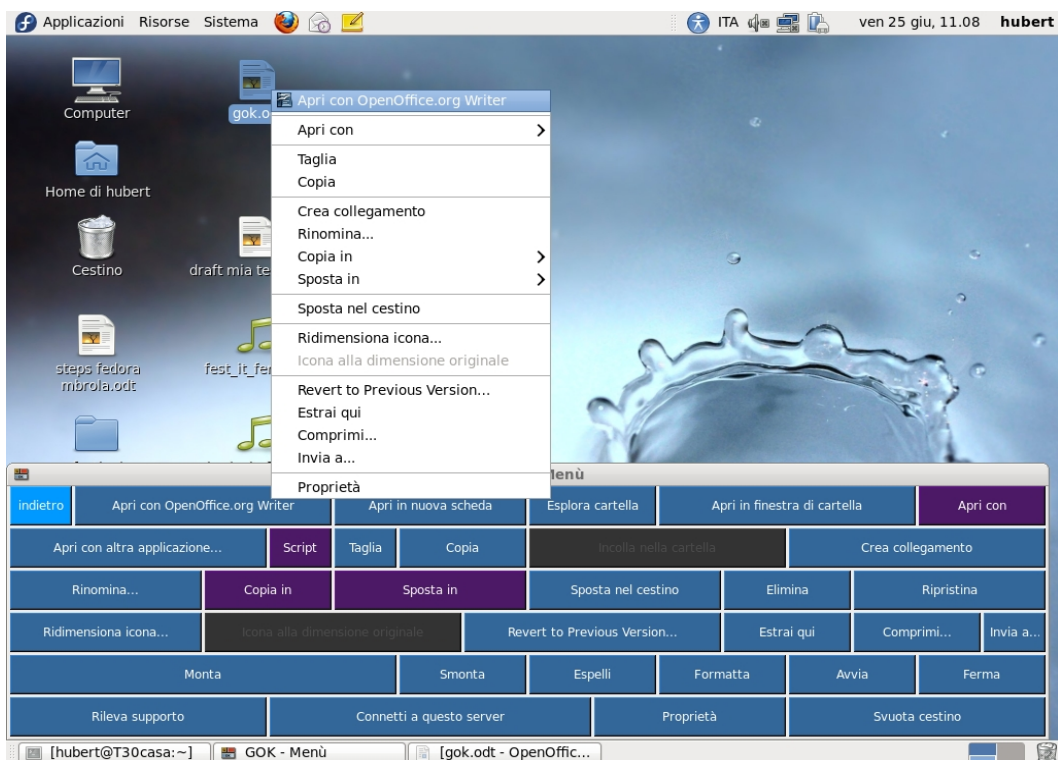


Figura 2.13

Un'altra applicazione notevole che viene in aiuto delle persone con disabilità motorie è il sistema di scrittura facilitata DASHER⁹. Si tratta di un software open source sviluppato da Inference Group; può lavorare sia in situazioni di utilizzo ad una mano (pilotato da joystick, touchscreen, trackball, o mouse) sia in situazioni di utilizzo senza mani (pilotato da head-mouse o da eyetracker). DASHER lavorando con qualsiasi sistema di puntamento consente di introdurre il testo senza bisogno di premere pulsanti, ma solo dirigendosi verso le parole (o il loro completamento) man mano che vengono proposte (provenienti dalla destra dello schermo).

Pur richiedendo un discreto allenamento il principio di funzionamento non è complicato. Facendo riferimento a figura 2.14 a partire dall'origine (intersezione dei 2 assi cartesiani da cui nasce la riga rossa di puntamento) ci si deve spostare verso la lettera (o completamento della parola) che ci interessa; più ci sposta dall'origine più si accelera (le lettere verranno proposte più rapidamente).

⁹<http://www.inference.phy.cam.ac.uk/dasher/>

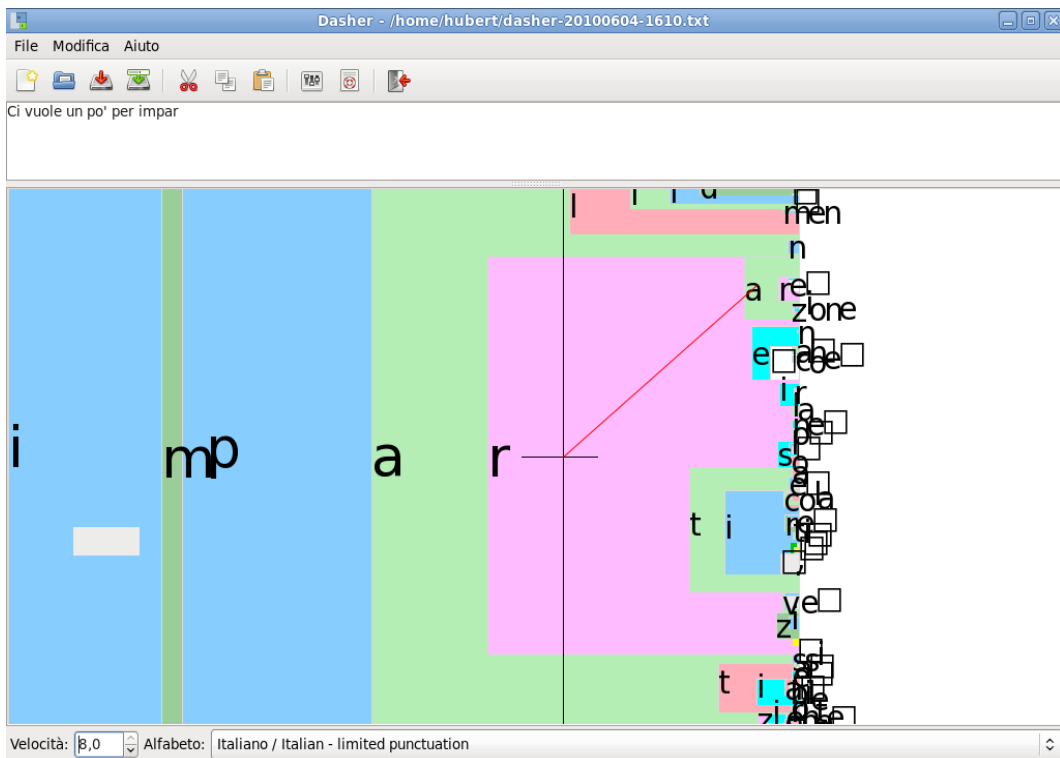


Figura 2.14

Nel momento dello scatto si stava componendo la parola *imparare*, la parte *impar* era già stata scritta (aveva superato l'asse verticale) e si puntava tra i vari completamenti proposti (scelti in base al dizionario selezionato inizialmente) quello cercato cioè *are* seguito dal quadratino vuoto che indica la fine della parola. Oltre a modificare la velocità, si può scegliere il dizionario e modificare altre opzioni come in figura 2.15:

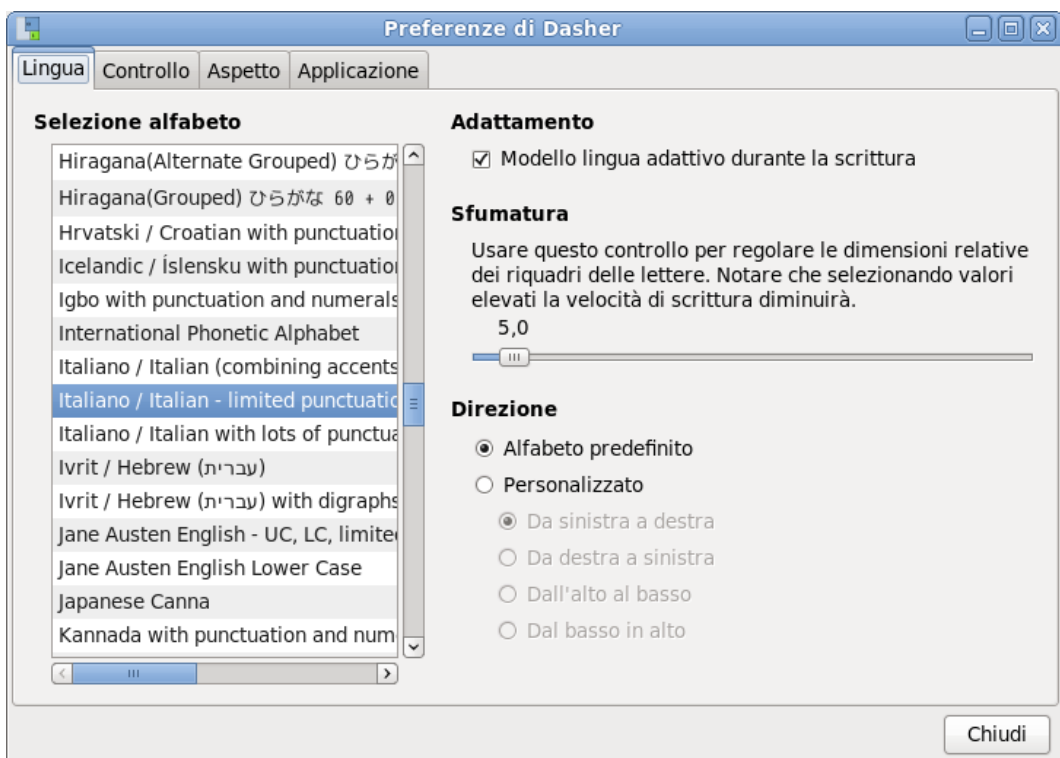


Figura 2.15

CAPITOLO 3 Strumenti di accessibilità in ambiente grafico KDE su sistema operativo GNU/Linux

KDE¹⁰ è una comunità internazionale impegnata nello sviluppo di software open source. Il software KDE offre ad utenti UNIX/Linux un ambiente grafico completo per interagire con il sistema operativo oltre ad applicazioni come Konqueror (web browser), Dolphin (file manager) ecc. All'interno di KDE si è sviluppato il KDEAP (KDE Accessibility Project)¹¹ che ha come obiettivo quello di rendere accessibile l'ambiente desktop KDE.

Fedora 13 installa KDE versione 4.4 di cui vedremo i principali componenti rivolti all'accessibilità.

KTTS: KDE Text-to-Speech System.

KTTS è il sistema Text-to-Speech di KDE, fornisce un'API comune agli sviluppatori KDE per arricchire le proprie applicazioni con la feature TTS. Tra le applicazioni che si interfacciano con KTTS ci sono gli editor di testo e Konqueror.

KTTS per effettuare la sintesi può fare leva su due sintetizzatori vocali: Espeak, Festival.

Espeak usa la sintesi basata sulle formanti (spiegata nel capitolo 4), consentendo così in poco spazio il supporto a molte lingue. Non consente ancora i risultati qualitativi dei sistemi basati su database (tipo i database di difoni che saranno illustrati sempre nel capitolo 4). In particolare, ha difficoltà a gestire gli accenti dell'italiano (il supporto a molti altri linguaggi oltre all'italiano stesso è in via di sviluppo).

FESTIVAL verrà illustrato anch'esso nel capitolo 4.

KTTS viene configurato via KTTSMGR illustrato in figura 3.1.

¹⁰<http://www.kde.org/>

¹¹<http://accessibility.kde.org/aids/>

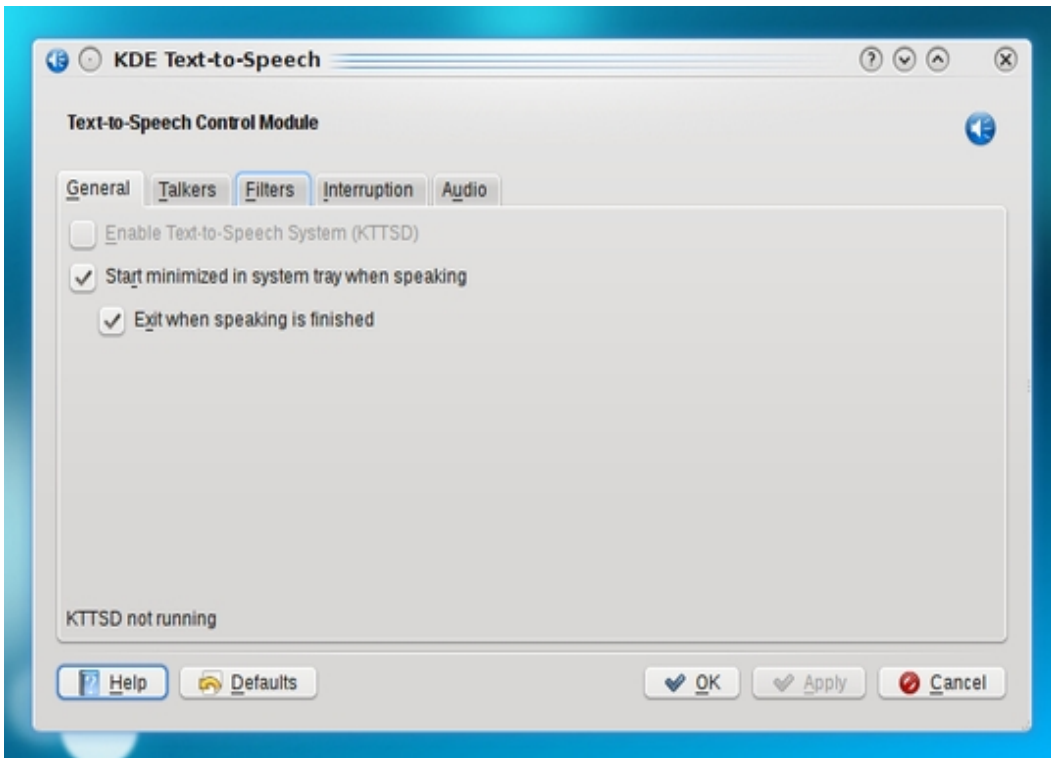


Figura 3.1

KMAG: KDE Screen Magnifier.

Anche questa applicazione viene incontro ad utenti con disabilità visive, consentendo di ingrandire la parte del desktop circostante il puntatore del mouse oppure un'area predefinita dall'utente.

In figura 3.2 si ha uno screenshot di KMAG versione 1.0.



Figura 3.2

Anche su KDE per aiutare utenti con disabilità motorie, sono disponibili settaggi di accessibilità che modificano il comportamento della tastiera e del mouse, consentendo ad esempio di simulare le combinazioni di tasti come in figura 3.3,

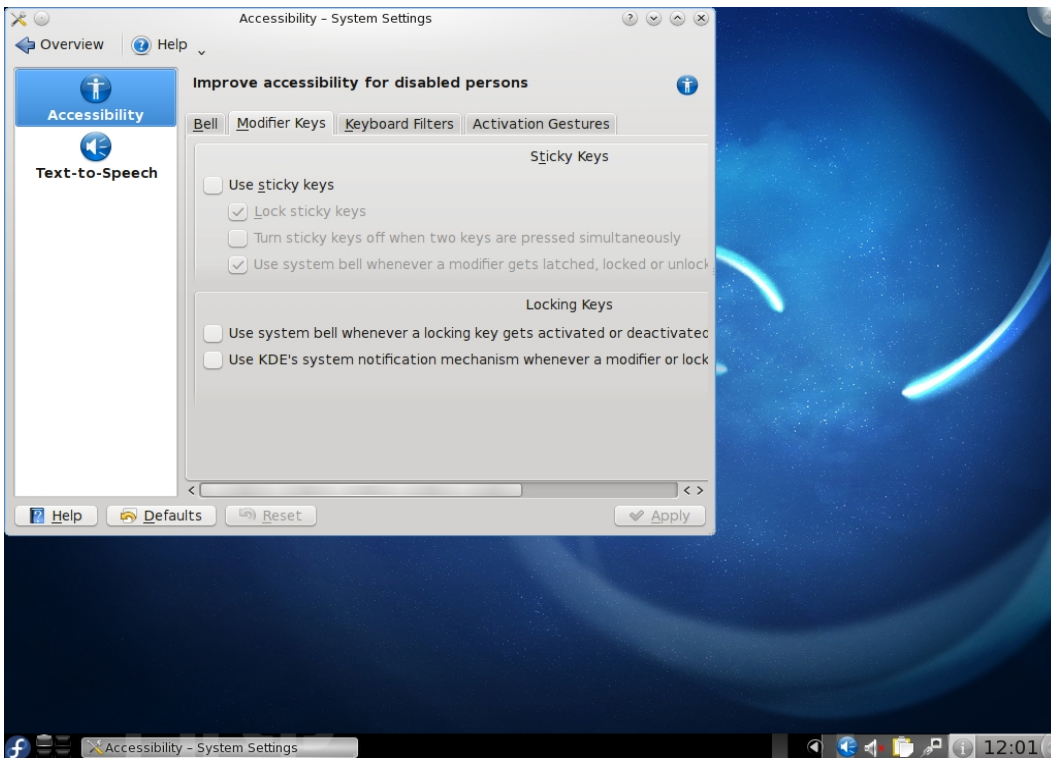


Figura 3.3

mentre in figura 3.4 vi sono i settaggi che riguardano la sensibilità della tastiera.

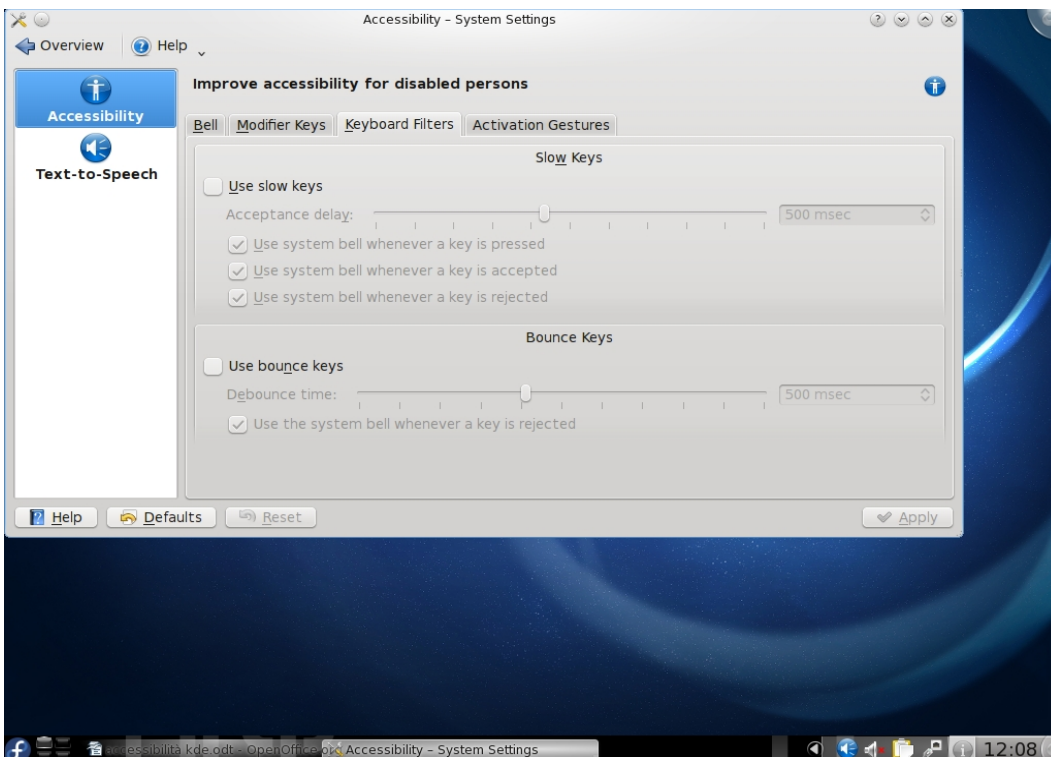


Figura 3.4

Per operare sulla configurazione del mouse si lavora tramite l'applicazione KMOUSE di cui si riporta la schermata di configurazione in figura 3.5.

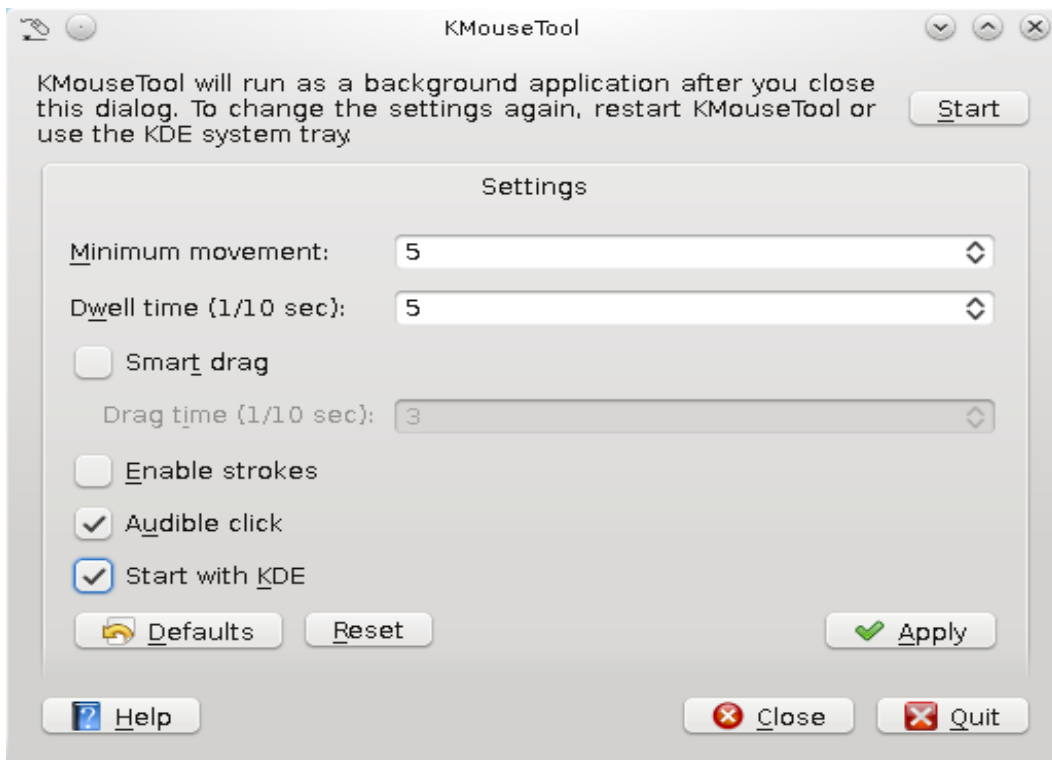


Figura 3.5

CAPITOLO 4 Approfondimento sulla tecnologia TTS (Text-to-Speech)

Un sistema TTS (Text-to-Speech)¹² è un sistema computerizzato in grado di leggere qualunque testo, il quale può essere introdotto in vari modi (es. da operatore o da scanner) e riprodurre la voce umana. Si compone di una parte front-end che converte il testo in simboli fonetici e di una parte back-end (come illustrato in figura 4.1) che trasforma i simboli fonetici in suoni creando effettivamente la voce artificiale.

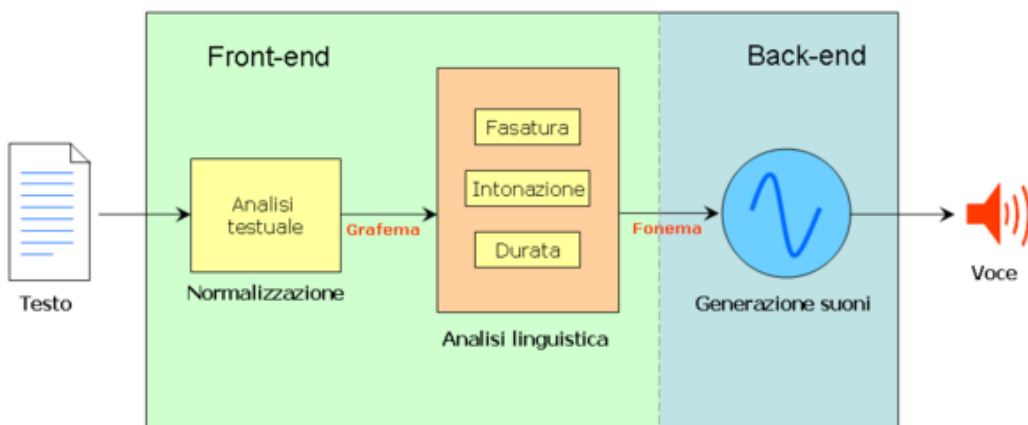


Figura 4.1 http://it.wikipedia.org/wiki/Sintesi_vocale

Come per il meccanismo di lettura umano un sistema TTS (si veda figura 4.2) comprende una parte NLP (Natural Language Processing) che produce la trascrizione fonetica del testo scritto (comprensiva di intonazione e ritmo) ed una parte DSP (Digital Signal Processing) che trasforma i simboli fonetici in parlato.

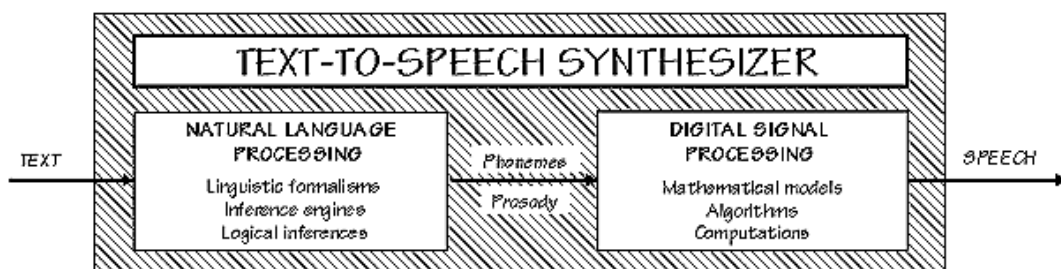


Figura 4.2 http://tcts.fpms.ac.be/synthesis/introtts_old.html

I blocchi costituenti della parte NLP sono il text analyzer block, il letter to sound module ed il prosody generator come appare in figura 4.3.

¹² http://tcts.fpms.ac.be/synthesis/introtts_old.html
http://it.wikipedia.org/wiki/Sintesi_vocale

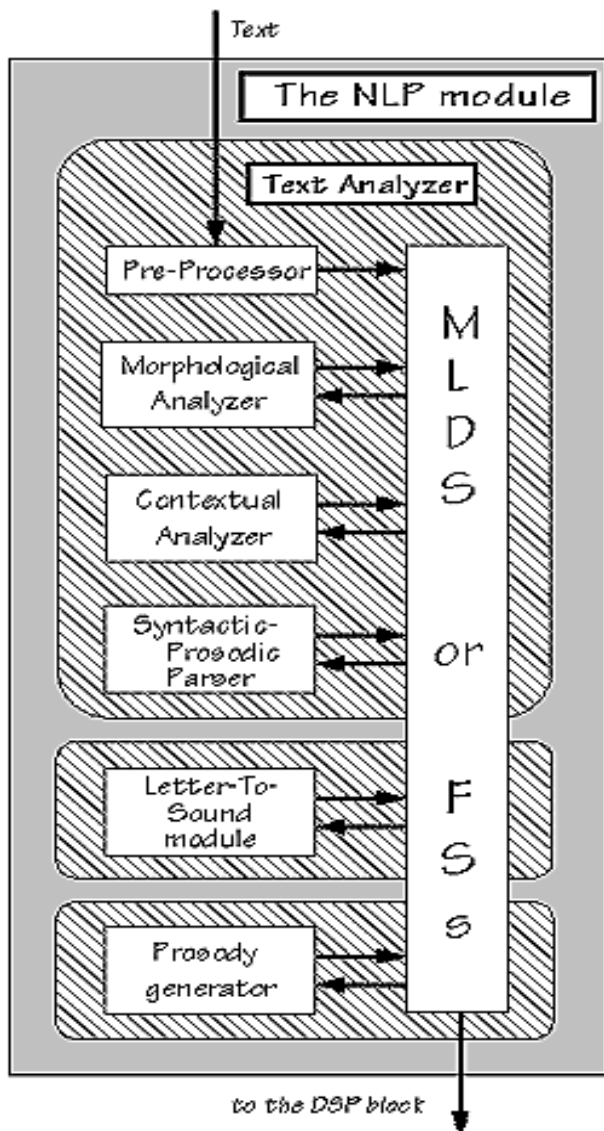


Figura 4.3 http://tcts.fpms.ac.be/synthesis/introtts_old.html

All'interno del text analyzer block, il pre-processor organizza le frasi in una lista utilizzabile di parole, risolvendo il problema di tradurre in parole i numeri, riconoscendo inoltre la punteggiatura e la fine delle frasi; la morfological analysis unit propone tutte le possibili speech categories per ogni singola parola (es. nome, verbo, aggettivo) e divide le singole parole in morfemi (parti elementari)¹³; il

¹³ Nella linguistica descrittiva moderna si chiama morfema il più piccolo elemento di una parola o di un enunciato dotato di significato che non possa essere ulteriormente suddiviso. Nella linguistica storica si preferisce limitare il concetto di morfema a un elemento non ulteriormente suddivisibile che abbia una funzione grammaticale; ciò che resta di una parola una volta tolti tutti i morfemi si chiama semantema. Esempio: le forme verbali possono essere composte da molti morfemi; prendendo come esempio l'espressione si sono inviperite troviamo:

- si: particella della coniugazione riflessiva;
- s- : radice del verbo ausiliare essere;
- ono: desinenza della terza persona plurale;
- in-: prefisso per derivare il verbo da un nome;
- viper-: radice del nome che dà il significato principale;
- i-: vocale tematica della terza coniugazione italiana;

contextual analysis module analizza le parole nel loro contesto ed intrecciando le possibili speech categories delle parole vicine, ne riduce le possibilità; infine il syntactic-prosodic parser analizza la struttura del testo per determinarne la probabile prosodia¹⁴.

Si passa poi al modulo letter to sound che è responsabile della trascrizione fonetica del testo in modo da generare fonemi¹⁵ compatibili con la parte back-end. In genere è realizzato in 2 modi (o un mix dei 2): Dictionary based o Rule based (Figura 4.4).

-t-: suffisso del participio passato;
-e: desinenza femminile plurale.

Morfema. (17 aprile 2010). Wikipedia, L'enciclopedia libera. Tratto il 1 giugno 2010, 15:51 da <http://it.wikipedia.org/w/index.php?title=Morfema&oldid=31464548>.

¹⁴ La prosodia è la parte della linguistica che studia l'intonazione, il ritmo, la durata e l'accento nel linguaggio parlato. Le caratteristiche prosodiche di un'unità di linguaggio parlato (si tratti di una sillaba, di una parola o di una frase) sono dette soprasegmentali, perché sono simultanee ai segmenti in cui può essere divisa quell'unità. Le si può infatti rappresentare idealmente come 'sovrapposte' ad essi. Alcuni di questi tratti sono, ad esempio, la lunghezza della sillaba, il tono, l'accento.

Prosodia. (20 maggio 2010). Wikipedia, L'enciclopedia libera. Tratto il 1 giugno 2010, 15:43 da <http://it.wikipedia.org/w/index.php?title=Prosodia&oldid=32277415>.

¹⁵ Un fonema è un'unità differenziante, indivisibile e astratta di un sistema linguistico. Non è corretto dire che costituisce un fono usato a carattere distintivo in una certa lingua. Secondo lo strutturalismo rappresenta l'unità minima di seconda articolazione, la più piccola e senza significato proprio.

Per individuare i vari fonemi e determinare così lo "spazio fonetico" di una data lingua si usa il criterio della ricerca di coppie minime, che si differenziano solo per un fonema. Per esempio, in italiano i "tipi fonici" p e b sono fonemi perché è possibile trovare almeno una coppia minima di parole che differiscono solo per un tratto distintivo (e.g. 'pelle' - 'belle').

Fonema. (17 giugno 2010). Wikipedia, L'enciclopedia libera. Tratto il 24 giugno 2010, 17:04 da <http://it.wikipedia.org/w/index.php?title=Fonema&oldid=32910369>

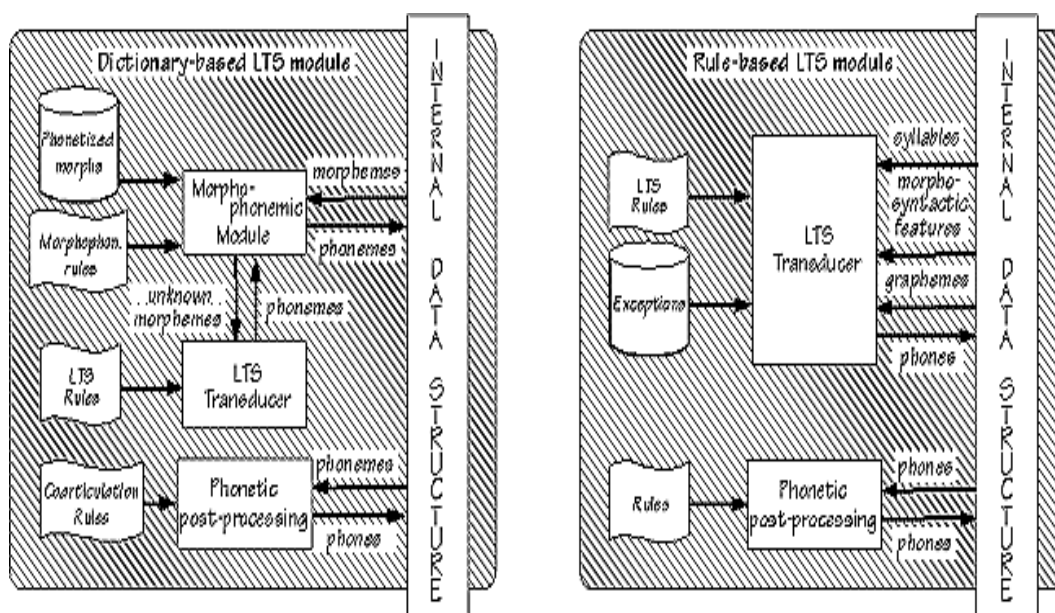


Figura 4.4 http://tcts.fpms.ac.be/synthesis/introtts_old.html

Nel metodo basato su dizionario, si cerca di ottenere un dizionario esaustivo di tutti i possibili fonemi. Per tenerlo di dimensioni più ridotte possibile si introducono nel dizionario i soli morfemi, la pronuncia delle parole da essi costituite si ottiene dalle regole linguistiche di inflessione, derivazione ecc.

Nel metodo Rule based si crea un set di regole *letter to sound* o *grafema-fonema*¹⁶ e si gestiscono le parole che esulano dalle regole con un dizionario delle eccezioni. Essendo la maggior parte delle eccezioni nelle parole di uso più comune, un piccolo dizionario delle eccezioni copre una larga parte delle parole di un testo corrente ad esempio in inglese, un dizionario delle eccezioni di 2000 parole, copre normalmente il 70% delle parole di un testo.

Riferendoci nuovamente alla figura 4.3 il modulo prosody generator si occupa di sovrapporre ai fonemi generati dal modulo letter to sound i dati riguardanti la prosodia del testo scritto.

Le strutture dati FSs (Feature Structures) ed MLDS (Multi Level Data Structures) sono i 2 sistemi più utilizzati di rappresentazione/elaborazione interna dei dati della parte Front-end, comprendono sia la divisione delle parole di una frase in speech categories, sia altre caratteristiche dei grafemi e dei fonemi componenti le parole della frase in esame.

A questo punto subentra la parte back-end del modulo TTS che è la parte DSP rappresentata in dettaglio in figura 4.5. Il suo ruolo per il computer è simile a quello della muscolatura e delle corde vocali nell'uomo per produrre il suono della voce.

¹⁶ Con il termine grafema si indica il segno elementare e non ulteriormente suddivisibile che costituisce il linguaggio scritto, ovvero la singola lettera. Altri esempi di grafemi sono i caratteri cinesi, i numeri e la punteggiatura.

Un primo punto fondamentale di cui tenere conto nella progettazione della parte DSP è il seguente: nel parlato, le transizioni fonetiche, ai fini della comprensione del discorso sono più importanti degli stati stabili. Si può tenerne conto in modo esplicito, imponendo delle regole formali che descrivono l'influenza reciproca dei fonemi oppure in modo implicito, creando un database di possibili concatenazioni di fonemi che verranno usate come unità acustiche al posto dei singoli fonemi.

Queste 2 filosofie portano a due classi di sintesi per sistemi TTS: basati sulle regole e basati sulla concatenazione.

I DSP per regole, spesso sono basati sulle formanti e descrivono il parlato come evoluzione dinamica di circa 60 parametri riguardanti, formanti, larghezza di banda e forma d'onda della voce. Attualmente le regole da introdurre per conferire naturalezza alla voce sintetizzata non sono ancora note. Questo tipo di sintetizzatore non utilizza campioni della voce umana ma ricrea la voce per elaborazione basandosi su un modello acustico e per tale motivo viene detta anche *sintesi per formanti*. Questa tecnica consiste nella generazione di forme d'onda di cui si modulano alcuni parametri acustici come la frequenza fondamentale, i toni e i livelli di rumore. Molti sistemi di sintesi basata sulle regole generano una voce dal suono artificiale e molto metallico che non può essere scambiata per una voce umana. Questa tecnica di sintesi non ha però come obiettivo la massima naturalezza e presenta una serie di vantaggi rispetto alla sintesi concatenativa. La sintesi basata sulle regole infatti è decisamente intelligibile anche ad alte velocità, non presentando i piccoli stacchi acustici tipici dei sistemi a sintesi concatenativa: la sintesi ad alta velocità è molto usata per i sistemi di lettura dello schermo per l'uso dei computer da parte delle persone ipovedenti.

Inoltre i sistemi di sintesi basata sulle regole sono gestiti da programmi di dimensione più contenuta non dovendo utilizzare un database di campioni vocali. Questa caratteristica ne consente l'impiego in sistemi embedded, dove la capacità di memoria e la potenza di calcolo del microprocessore possono essere limitate. Infine, i sistemi di sintesi basata sulle regole possono controllare tutti gli aspetti del linguaggio vocale, generando un'ampia varietà di prosodie e intonazioni e veicolando così non soltanto il contenuto del testo ma anche effetti emotivi e toni di voce. Tra gli esempi di sintesi basata su regole con un controllo molto accurato dell'intonazione, sia pure non in tempo reale, si trovano i lavori svolti alla fine degli anni settanta per il gioco *Speak & Spell* prodotto da Texas Instruments e per le console di videogiochi prodotte da SEGA all'inizio degli anni ottanta.

La sintesi concatenativa, come dice il nome, si basa sulla concatenazione ossia la combinazione di frammenti di voce registrati. In generale questa metodologia produce il risultato di sintesi più naturale, tuttavia la differenza tra le variazioni naturali della voce umana e le tecniche di frammentazione automatica delle forme d'onda può talvolta generare dei piccoli disturbi udibili.

Esistono tre sotto-tipi principali di sintesi concatenativa: la sintesi per campioni unitari, la sintesi per difoni, la sintesi per applicazioni specifiche.

La sintesi per campioni unitari si appoggia su grandi database di voci registrate. Durante la creazione del database ogni emissione registrata viene frazionata in uno o più di questi segmenti: suoni isolati, sillabe, morfemi, parole, frasi e periodi completi.

I frammenti subiscono poi un allineamento forzato che ne consentirà una successiva ricombinazione con stacchi molto meno percettibili. I campioni sonori vengono indicizzati nel database sulla base della frammentazione e di parametri acustici quali la frequenza fondamentale (tono musicale), la durata, la posizione all'interno della sillaba e i suoni adiacenti. Durante la sintesi in tempo reale l'emissione finale viene generata sulla base di un algoritmo di decisione ad albero pesato che identifica la miglior sequenza tra i campioni candidati scelti dal database. Questo tipo di sintesi produce i risultati di maggior naturalezza perché riduce al minimo le operazioni di elaborazione digitale (DSP) sui campioni registrati. Le elaborazioni digitali infatti spesso alterano la resa del suono sintetizzato rendendola meno naturale: alcuni sistemi usano tecniche DSP solo per ammorbidire le transizioni tra i campioni sonori in fase di concatenazione.

La sintesi per difoni¹⁷ utilizza un database di suoni di dimensioni minime contenente tutti i difoni tipici di un determinato linguaggio. Il numero dei difoni dipende dalle caratteristiche fonetiche del linguaggio: per esempio, la lingua spagnola comprende circa 800 difoni mentre il tedesco ne conta circa 2500. Con questa tecnica viene memorizzato nel database un unico campione per ciascun difono. Durante l'elaborazione in tempo reale, ai difoni selezionati viene sovrapposta la prosodia della frase usando tecniche DSP come la codifica lineare predittiva, PSOLA (Pitch-Synchronous Overlap and Add) oppure MBROLA. La qualità della voce risultante in genere è inferiore rispetto a quella ottenuta per sintesi per campioni unitari ma suona più naturale rispetto a quella ottenuta con la sintesi basata sulle regole. I difetti della sintesi per difoni consistono in piccoli stacchi tra i suoni, tipici del meccanismo di concatenazione, e in un effetto di voce metallica come nella sintesi basata sulle regole. In effetti, rispetto a queste tecniche la sintesi per difoni non presenta vantaggi significativi a parte la dimensione ridotta del database di appoggio. Per questo motivo l'impiego di questa tecnica per applicazioni commerciali è in fase di declino mentre continua a essere impiegata nella ricerca grazie alle molte implementazioni software gratuite disponibili.

La sintesi per applicazioni specifiche è di fatto un tipo particolare di sintesi per campioni unitari in cui si usa un database di frasi intere pre-registrate. Si usa principalmente per applicazioni in cui i testi da sintetizzare sono limitati alle esigenze di un settore specifico, come per esempio gli annunci ferroviari o aeroportuali. La tecnologia è semplice da implementare ed è in uso da tempo in applicazioni di tipo commerciale e in dispositivi tipo le sveglie parlanti o le calcolatrici con voce. La naturalezza di questi sistemi è molto elevata grazie al fatto che il numero di frasi componenti è limitato e riproduce molto fedelmente la prosodia e l'intonazione delle

¹⁷ Un difono è una transazione tra due suoni diversi. Come regola generale si indica con difono quella porzione di segnale vocale relativa alla sequenza di due fonemi che va da $\frac{1}{2}$ del primo fonema a $\frac{1}{2}$ del successivo. Ci sono però alcune eccezioni come ad esempio nelle geminate dove ai difoni destro e sinistro si aggiunge un difono centrale che va da $\frac{1}{4}$ a $\frac{3}{4}$ del fonema geminato. In fonetica, la consonante geminata indica un suono consonantico la cui durata sia apprezzabilmente più lunga di quella delle consonanti ordinarie, dette brevi o scempie.

registrazioni originali. D'altro canto, questi sistemi si limitano a riprodurre parole e frasi contenute nel loro database e possono sintetizzare solo le combinazioni predefinite, per cui non possono essere estesi per un uso generalizzato.

Gli aspetti più influenti sulle prestazioni di un sintetizzatore per concatenazione sono: la scelta dei segmenti, il tipo di sintesi del segnale cui si fa riferimento.

I segmenti devono coprire tutti gli eventi possibili, devono essere facilmente concatenabili per non rendere troppo pesante il lavoro del blocco di concatenazione, non devono occupare troppo spazio di memoria. Questo impone la ricerca di un compromesso tra diverse necessità. Segmenti abbastanza lunghi, ridurrebbero di molto i punti di concatenazione tra gli stessi aumentando di molto la qualità della sintesi, analogamente, per tenere conto di tutti le possibilità di utilizzo di un segmento durante un discorso, lo stesso dovrebbe essere memorizzato in più varianti.

Spesso dunque si usano di difoni che sono abbastanza limitati in numero. Per contro i difoni richiedono un alto numero di punti di concatenazione, ed un buon algoritmo per la stessa.

Riguardo ai meccanismi di sintesi, essi possono essere divisi in due grandi gruppi: Production models ed i Phenomenological models.

I production models forniscono un modello matematico dei fenomeni vocali (legati alle corde vocali, agli effetti nasali della voce ed altro), tra questi emergono gli LPC (Linear Prediction Coding) ed i sintetizzatori basati su Formanti.

I phenomenological models non tengono conto degli effetti umani sul segnale producendo una voce completamente artificiale; tra questi i più importanti sono hybrid Harmonic/Stochastic (H/S) model di [Abrantes et al. 91] ed il Time-Domain Pitch-Synchronous-Overlap-Add (TD-PSOLA) one [Moulines & Charpentier 90].

Quest'ultimo è un time domain algorithm, con un buon risultato qualitativo ed un basso costo computazionale (mediamente 7 operazioni per campione), il primo ha un costo computazionale circa 10 volte superiore.

L'algoritmo MBROLA [Dutoit 93,96] unisce la pulizia di concatenazione tipica dell'algoritmo hybrid Harmonic/Stochastic con il basso costo computazionale di PSOLA.

Uno schema per un DSP basato sulla concatenazione, è quello illustrato in figura 4.5:

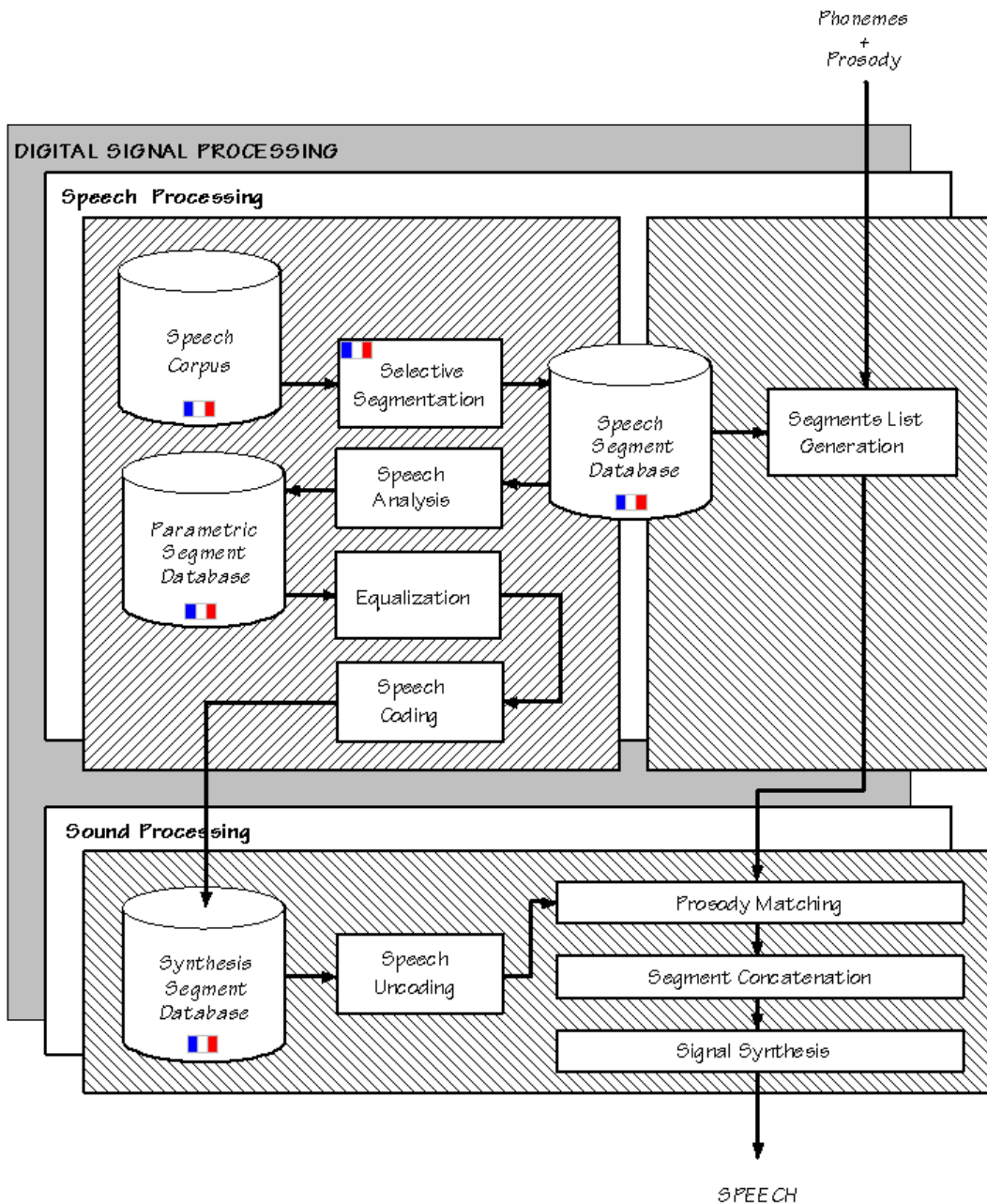


Figura 4.5 http://tcts.fpms.ac.be/synthesis/intro_tts_old.html

La parte speech processing tramite il modulo Segments List Generation interfaccia la parte NLP (Front-end) con il DSP (Back-end) deducendo una prima lista di segmenti. Nella parte sound processing il Prosody Matching module, cerca nel DB quei segmenti i cui parametri più si adattano e li modifica per adeguarli alla prosodia prima di passarli al modulo che si occupa della concatenazione il quale (sempre tramite modifica di alcuni parametri) elimina anche le discontinuità tra gli stessi. Si ottiene così la sequenza finale di parametri da passare al modulo che si occupa puramente della sintesi vocale.

CAPITOLO 5 Prove di configurazione TTS FESTIVAL ed installazione MBROLA

FESTIVAL è un sistema TTS sviluppato da Centre for Speech Technology Research CSTR in Edinmburgh, Scotland, UK; offre inoltre un ambiente di sviluppo e ricerca per tecniche di speech sysnthesis. Si basa sulla concatenazione di difoni ed usa la tecnica di sintesi Residual-Excited LPC.

ISTC-SPFD CNR ha curato la registrazione del database per la voce italiana maschile, mentre ITC-irst si è occupato della registrazione del database per la voce italiana femminile. Per le registrazioni delle voci si è fatto ricorso a lettori/attori professionisti come Claudio Carini. Contestualmente sono stati rivisti per la lingua italiana i moduli di analisi del testo. Le stesse componenti sono state anche riviste per utilizzare MBROLA.

MBROLA è un algoritmo di sintesi vocale basato sulla concatenazione di difoni, non è un sistema TTS completo in quanto fornisce solo l'algoritmo di sintesi e (per alcune lingue) il database di difoni.

Il progetto è dello TCTS Lab of the Faculté Polytechnique de Mons (Belgium).

E' stata effettuata una comparazione tra FESTIVAL e MBROLA producendo 2 files in formato wave risultanti dalla lettura della poesia *L'infinito* di Leopardi tramite il comando text2wave. Dai 2 files si può notare il miglioramento nei punti di congiunzione dei difoni con MBROLA rispetto a FESTIVAL.

Vengono riportati di seguito i dettagli di installazione e configurazione di strumenti TTS su un pc portatile IBM T30 con installato il sistema operativo linux open source Fedora 13.

L'installazione di default di Fedora 13 con interfaccia grafica GNOME prevede la versione 1.96 di Festival e Flite:

```
[root@T30casa hubert]# yum info festival
Loaded plugins: presto, refresh-packagekit
Installed Packages
Name           : festival
Arch           : i686
Version        : 1.96
Release        : 16.fc13
Size           : 8.8 M
Repo           : installed
From repo      : anaconda-InstallationRepo-201005130056.i386
Summary        : Speech synthesis and text-to-speech system
URL            : http://www.cstr.ed.ac.uk/projects/festival/
License        : MIT and GPL+ and TCL
Description    : Festival is a general multi-lingual speech synthesis
system developed at CSTR. It offers a full text to speech system
with various APIs, as well as an environment for development and
```

research of speech synthesis techniques. It is written in C++ with a Scheme-based command interpreter for general control.

```
[root@T30casa hubert]# yum info flite
Loaded plugins: presto, refresh-packagekit
Available Packages
Name           : flite
Arch           : i686
Version        : 1.3
Release        : 14.fc12
Size           : 5.7 M
Repo           : fedora
Summary        : Small, fast speech synthesis engine (text-to-speech)
URL            : http://fife.speech.cs.cmu.edu/flite/
License        : MIT
Description    : Flite (festival-lite) is a small, fast run-time
speech synthesis engine developed at CMU and primarily designed
for small embedded machines and/or large servers. Flite is
designed as an alternative synthesis engine to Festival for voices
built using the FestVox suite of voice building tools.
```

le voci installate di default sono in inglese:

```
[root@T30casa hubert]# yum search festvox
Loaded plugins: presto, refresh-packagekit
Matched: festvox
festvox-awb-arctic-hts.noarch : Scottish-accent US English male
speaker "AWB" for Festival
festvox-bdl-arctic-hts.noarch : US English male speaker "BDL" for
Festival
festvox-clb-arctic-hts.noarch : US English female speaker "CLB"
for Festival
festvox-jmk-arctic-hts.noarch : Canadian-accent US English male
speaker "JMK" for Festival
festvox-rms-arctic-hts.noarch : US English male speaker "RMS" for
Festival
festvox-slt-arctic-hts.noarch : US English female speaker "SLT"
for Festival
festvox-kal-diphone.noarch : American English male speaker "Kevin"
for Festival
festvox-ked-diphone.noarch : American English male speaker "Kurt"
for Festival
flite.i686 : Small, fast speech synthesis engine (text-to-speech)
```

non prevede MBROLA

```
[root@T30casa hubert]# yum info mbrola
Loaded plugins: presto, refresh-packagekit
Error: No matching Packages to list
```

Dopo aver scaricato le componenti necessarie dal sito:

<http://www2.pd.istc.cnr.it/festival/>

eseguiamo le operazioni di installazione e configurazione.

Per installare lo schema italiano per FESTIVAL si è usato il comando:

```
[root@T30casa share]# tar xvfz
/home/hubert/Desktop/festival/italian_scheme/italian_scm.tar.gz
```

mentre l'installazione del lessico italiano di FESTIVAL richiede il seguente comando:

```
[root@T30casa share]# tar xvfz
/home/hubert/Desktop/festival/italian_lexicon/festlex_IFD.tar.gz
```

per installare le voci maschili e femminili per FESTIVAL e per MBROLA sono stati usati i seguenti comandi:

```
[root@T30casa share]# tar xvfz
/home/hubert/Desktop/festival/voci_base/voce_maschile_PC/festvox_p
c_diphone.tar.gz
```

```
[root@T30casa share]# tar xvfz
/home/hubert/Desktop/festival/voci_base/voce_femminile_LP/festvox_
lp_diphone.tar.gz
```

```
[root@T30casa share]# tar xvfz
/home/hubert/Desktop/festival/voci_per_MBROLA/maschile_PC/festvox_
pc_mbrola.tar.gz
```

```
[root@T30casa share]# tar xvfz
/home/hubert/Desktop/festival/voci_per_MBROLA/femminile_LP/festvox
_lp_mbrola.tar.gz
```

si sono poi copiati i dizionari italiani per MBROLA:

```
[root@T30casa /]# cp -r
/home/hubert/Desktop/festival/italian_db_per_MBROLA/it3
/usr/share/festival/lib/voices/italian/italian
[root@T30casa /]# cp -r
/home/hubert/Desktop/festival/italian_db_per_MBROLA/it4
/usr/share/festival/lib/voices/italian/italian
[root@T30casa /]#
```

Come testo di prova per la lettura si è usato *L'infinito* di Leopardi, settando come voce di default per FESTIVAL la voce maschile festvox_pc_diphone. Dopo una prima prova è stato ottenuto il seguente messaggio:

```
[root@T30casa festival]# festival --tts
/home/hubert/Desktop/infinito.txt
LTS_Ruleset italian_downcase: no rule matches:
LTS_Ruleset: # □ *here* □ □ S e m p r e #
closing a file left open: /home/hubert/Desktop/infinito.txt
```

Il problema è stato risolto ricodificando il file, scegliendo al momento del salvataggio con GEDIT l'opzione codifica caratteri = occidentale (ISO-8859-1) come illustrato in figura 5.1.



Figura 5.1

CONCLUSIONI

Come esposto nel capitolo 1, è forte l'impegno da parte degli organismi di governo nazionali e sovranazionali ai fini di garantire una adeguata qualità della vita alle persone affette da disabilità.

La tecnologia ha un ruolo non secondario nel perseguire questo risultato, tramite lo sviluppo di ausili hardware o software sempre più efficaci.

Mentre sui costi (per il singolo individuo o per la società) degli ausili hardware non si può fare molto, un importante impatto dal punto di vista economico oltre che tecnologico sulla fruibilità degli ausili software può avere lo sviluppo di tali sistemi in ambito open source.

Attualmente la gamma degli ausili software forniti in ambito open source è assai vasta, ed in grado di fornire aiuto nei più svariati casi.

Dopo aver superato (grazie anche all'uscita di Fedora 13) alcune difficoltà legate all'installazione e configurazione dei prodotti su hardware un po' obsoleto come quello in dotazione, si sono potute apprezzare le funzionalità dei software rivolti all'accessibilità.

Un esempio importante di applicazione della filosofia open source si ha da MBROLA che non fornisce un TTS completo, ma solo una parte back-end che consente lo sviluppo da parte di terzi della relativa parte front-end. Questo aiuta molto la diffusione della tecnologia TTS a quante più lingue possibile.

Un altro esempio è GNOME con il toolkit AT-SPI. Anche KDE con il progetto KDEAP si sta muovendo nella stessa direzione.

Sulla base di questi esempi si può affermare che la filosofia open source è basilare per il miglioramento di questi ausili e per la loro diffusione al maggior numero di utenti possibile.

Biblio/Sitografia

Disabilita':

http://www.portale.siva.it/servizi/guida/siva_ITA.htm

<http://www.dienneti.it/handicap/index.htm>

<http://www.pariopportunita.it/>

<http://www.usabile.it/>

Legislazione:

http://www.pubbliaccesso.gov.it/normative/legge_20040109_n4.htm

<http://www.governo.it>

<http://www.parlamento.it>

legge 104 del 1992:

http://www.governo.it/GovernoInforma/Dossier/agevolazioni_disabili/L_104.pdf

legge 4 del 2004:

http://www.governo.it/GovernoInforma/Dossier/linee_guida_disabili/legge_4_2004.pdf

[legge_4_2004.pdf](http://www.governo.it/GovernoInforma/Dossier/linee_guida_disabili/legge_4_2004.pdf)

legge 17 del 1999:

http://www.parlamento.it/leg/ldl_new/sldlelenco011999ordcron.htm

Strumenti accessibilità open source:

<http://www.it.gnome.org/index.php/Home>

<http://projects.gnome.org/accessibility/>

<http://live.gnome.org/Orca>

<http://www.inference.phy.cam.ac.uk/dasher/>

<http://www.kde.org/>

<http://accessibility.kde.org/aids/>

Sistemi TTS:

http://tcts.fpms.ac.be/synthesis/introtts_old.html

http://it.wikipedia.org/wiki/Sintesi_vocale

<http://www-swiss.ai.mit.edu/~jaffer/Scheme.html>

<http://ditelo.itc.it/>

P. Cosi e J.P. Hosom, "High Performance "General Purpose" Phonetic Recognition for Italian",

Proceedings of International Conference on Spoken Language Processing (ICSLP-2000), Beijing, Cina, 16-20 October, 2000, Vol. II, pp. 527-530.

B. Angelini, F. Brugnara, D. Falavigna, D. Giuliani, R. Gretter e M. Omologo, "A Baseline of a Speaker Independent Continuous Speech Recognizer of Italian", Proceedings of EUROSPEECH 93, Berlin, Germany, 1993.

A.J. Fourcin, G. Harland, W. Barry e V. Hazan, Eds., Speech Input and Output Assessment, Multilingual Methods and Standards, Ellis Horwood Books in Information Technology, 1989.

M. Federico, D. Giordani, and P. Coletti, "Development and evaluation of an Italian broadcast news corpus", Proceedings of 2nd International Conference on Language Resources and Evaluation, Athens, Greece, 2000.

M. Fanty, J. Pochmara e R.A. Cole, "An Interactive Environment for Speech Recognition Research", Proceedings of International Conference on Spoken Language Processing (ICSLP-92), Banff, Alberta, October 1992, 1543-1546.

Festival:

<http://www.cstr.ed.ac.uk/projects/festival/>.

<http://www2.pd.istc.cnr.it/festival/>

<http://www.itc.it/Index.aspx>

Festvox:

<http://www.festvox.org/>.

CSLU SPEECH TOOLKIT:

<http://cslu.cse.ogi.edu/tools.htm>.

Praat:

<http://www.fon.hum.uva.nl/praat>.

Tcl/Tk:

<http://sol.brunel.ac.uk/tcl/Tcl.html>.

MBROLA:

<http://tcts.fpms.ac.be/synthesis/>.

An Introduction to Text-to-Speech Synthesis . Autore : Thierry Dutoit . Kluwer Academic Publishers . ISBN 1-4020-0369-2

MPIRO:

Multilingual Personalized Information Objects. European Project IST-1999-10982

Version: <http://www.ltg.ed.ac.uk/mpiro/>