

Facciamo parlare il PC


 SOFTWARE
SUL
CD/DVD

■ Se siete stanchi di scorrere il testo con il mouse e di scrivere con la tastiera, ecco come utilizzare gli strumenti integrati in KDE per dettare il testo e farlo leggere direttamente al vostro computer

Tra le varie forme di interazione con il PC al momento disponibili, la più diretta è senza dubbio quella vocale, ma al momento è anche quella più difficile da implementare. Fortunatamente, gli sviluppi più recenti di tale tecnologia consentono letteralmente di "parlare" con il proprio computer in maniera molto naturale e gestire diverse operazioni come la lettura e la scrittura di testi, l'esecuzione di applicazioni e altri aspetti legati alla gestione del PC semplicemente con l'utilizzo della voce.

Tutto aumenta notevolmente la produttività dell'utenza e rende un attimo più "umano" il calcolatore, ma soprattutto abbatte finalmente le "barriere tecnologiche" che separano ancora oggi il PC dai portatori di handicap: pensate a persone con gli arti paralizzati o mancanti, utenti ciechi e così via. Le applicazioni possibili da questo punto di vista sono infinite.

La tecnologia che sta dietro a tutto questo non è certamente nuova (sono decenni che IBM lavora in questo senso) e in ambiente Windows sono presenti da tempo software per la sintesi vocale (*IBM ViaVoice*, *Dragon Naturally Speaking*, ecc.) che consentono di dettare direttamente testi all'interno dei word processor, farli leggere al PC e avviare applicazioni semplicemente utilizzando la voce.

Anche la comunità Open Source si è mossa in questo senso, in particolare gli sviluppatori di KDE che nell'ultima versione (3.4) del desktop hanno integrato, tra le diverse funzionalità, un completo sistema per la sintesi vocale. In realtà la comunità si è occupata da diverso tempo del "problema" proponendo diversi sistemi (*Festival*, *Mbrola*, ecc.), ma finalmente gli sviluppatori di KDE sono riusciti a integrare queste tecnologie e rendere il loro utilizzo semplice e naturale.

COME FUNZIONA LA SINTESI VOCALE

I software di sintesi vocale permettono di generare un segnale elettrico corrispondente alla voce umana a partire da un testo in formato elettronico e di ascoltarlo attraverso la scheda audio del PC. La generazione della voce avviene in real time, in un tempo molto inferiore alla durata del segnale prodotto (circa un centesimo di secondo). In teoria, sarebbe quindi possibile per un calcolatore leggere molti testi contemporaneamente, bufferizzarli da qualche parte e riproporli in un secondo momento. Questi software per la sintesi vocale sono molto complessi e richiedono anni di sviluppo sia per la scrittura del codice sia per la raccolta dei segmenti vocali e dei dizionari. Linguaggi naturali, come ad esempio l'italiano, presentano molte eccezioni e, a causa di questa ricchezza

semantica dovuta anche all'evoluzione storica, non è molto semplice racchiudere in dizionari digitali i dittonghi e le sillabe parlate, a maggior ragione la difficoltà si presenta per i cognomi, i nomi di luogo e di persona. Per generare la voce, a partire da un testo, occorre che quest'ultimo sia per prima cosa analizzato frase per frase e "normalizzato", cioè occorre accordarsi sulle convenzioni della pronuncia (numeri, abbreviazioni, ecc.) e in questa fase vengono anche cercate nel vocabolario utente

KDE
3.4

LICENZA: GNU GPL

SITO WEB: www.kde.org

TIPO: Desktop

UN PO' DI STORIA

La ricerca in questo settore è partita nel 1950, quando nei laboratori Bell furono ideati degli elaboratori in grado di riconoscere parole isolate (ovviamente scandite e separate da pause). Negli anni successivi, i progressi sono stati scarsi, fino a quando nel 1971, il Dipartimento della Difesa statunitense varò un progetto di ricerca sul "parlato continuo" (senza pause). Da quel momento la strada per pilotare i computer a voce è iniziata! Da quegli anni, grazie all'aumento vertiginoso della potenza di calcolo, sono nati numerosi prodotti di vocal synthesis e speech recognition. In questo ambito, analizzeremo i sintetizzatori vocali che permettono di leggere testi e fornire informazioni sul contesto di lavoro. Questa funzione si rivela particolarmente utile per i disabili, agevolando la navigazione su Internet (si parla molto di accessibilità dei siti web) o all'interno di menu e finestre.



eventuali pronunce particolari. Si procede, quindi, alla divisione delle frasi in parole e alla ricerca di queste nel vocabolario interno per risolverne la pronuncia; inoltre, si definisce l'accentazione e si effettua la scansione delle pause tra una parola e l'altra nel modo più naturale possibile. Dopo aver diviso le parole, occorre individuare le sillabe, intese come "suoni pronunciati adiacenti". Sulle sillabe si effettua poi l'annotazione del tono più o meno alto (prosodia) con cui devono essere pronunciate affinché venga riprodotta, il più fedelmente possibile, la voce naturale (questo è sicuramente il passaggio più critico di tutto il procedimento). Dalla divisione in sillabe si passa alla divisione in "polifoni", cioè insiemi di uno o più suoni presenti come tali nel database dei segmenti vocali del sintetizzatore e di ogni polifono vi sono più istanze nel database, uno per ogni diversa impostazione tonale. Il software si incaricherà di scegliere di volta in volta l'istanza più adatta a realizzare una particolare sillaba in una particolare posizione tonale. Una volta scelta l'istanza del singolo polifono più vicina alle caratteristiche tonali da realizzare, è necessario adattarla ai parametri desiderati tramite l'elaborazione numerica dei segnali, perché non si otterranno mai esattamente le caratteristiche di tono e di durata necessarie. Da quest'ultimo passo deriva il termine "sintesi vocale". I polifoni così "sintetizzati" saranno poi concatenati, con opportune dissolvenze per formare il segnale vocale finale prodotto dal sintetizzatore. Il segnale così ottenuto verrà poi inviato alla scheda audio tramite algoritmi di bufferizzazione intelligenti che gestiranno al meglio il flusso dei dati.

REQUISITI FONDAMENTALI

Al fine di poter usufruire di tutte le funzionalità descritte nel corso dell'esempio è necessario disporre dei seguenti software:

- KDE versione \geq 3.2, meglio se 3.4;
- Un motore di sintesi vocale, ad esempio Festival o Mbrola;
- Il programma *sox* (opzionale);
- Il software *xsltproc* (opzionale);
- Il sistema *Gstreamer \geq 0.8.7 (opzionale);*

Il programma *sox* non è indispensabile ma servirà nel caso in cui abbiate la necessità di variare la velocità di pronuncia.

Il programma *xsltproc* è richiesto per il supporto SSML e per utilizzare i filtri di trasformazione XML offerti da *kttsmgr*.

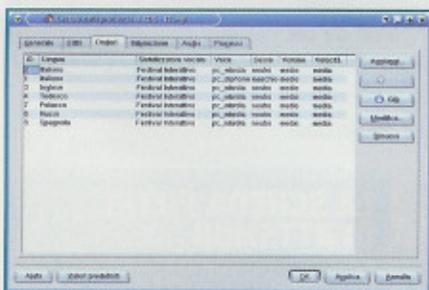
Infine, anche *Gstreamer* non sarà indispensabile ma potrà essere utilizzato come valida alternativa ad *ARTs*, utilizzato di default da KDE.

IL SISTEMA DI SINTESI VOCALE IN KDE

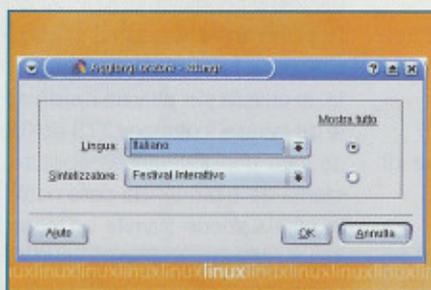
KDE include, all'interno delle sue innumerevoli funzionalità, un sottosistema per la pronuncia suddiviso in varie applicazioni: *kttsd*, *kttsmgr*, *libkhtml-kttsdplugin* e *kttexteditor_kttsd*. Il demone per la sintesi vocale di KDE, *kttsd*, fornisce tutti i servizi TTS (*Text To Speech*) alle varie applicazioni tramite DCOP. Il programma *kttsmgr* è invece dotato di un'interfaccia grafica e serve per configurare e gestire il demone *kttsd*. Inoltre, è disponibile *libkhtmlkttsdplugin*, un plugin di Konqueror che fornisce all'utente la possibilità di pronunciare, tutto o una parte, del testo di un sito web. Infine, *kttexteditor_kttsd* che permette di pronunciare, tutto o una parte, di un file di testo. Le funzionalità principali, offerte agli utenti da questo sottosiste-

Gli "oratori" di KDE

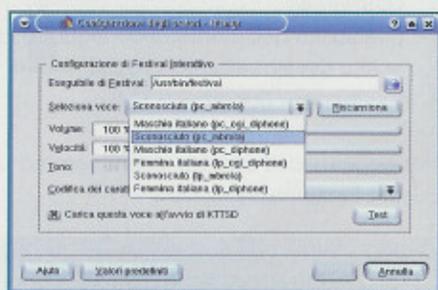
Aggiungere e gestire i plugin per scegliere la voce e la lingua con cui verrà letto il testo



1 NUOVO ORATORE Questa è la scheda di configurazione più importante di *kttsmgr* dove potrete aggiungere un nuovo oratore, ossia un plugin di sintesi vocale (la voce e la lingua con cui verrà letto il testo). Cliccando sul pulsante "Aggiungi", apparirà la finestra "Aggiungi oratore" dove sarà possibile scegliere, tra quelli disponibili, la lingua e il sintetizzatore desiderato.



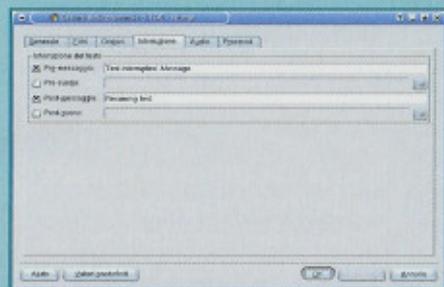
2 VOCI DISPONIBILI A questo punto *kttsmgr* interrogherà il sintetizzatore prescelto sulle voci disponibili, indicando anche il tempo di attesa. Dopo aver effettuato quest'operazione, verranno elencate nella scheda "Oratori" tutte le voci disponibili trovate da *kttsmgr* con la possibilità di agire su queste cliccando, semplicemente, sul pulsante "Modifica".



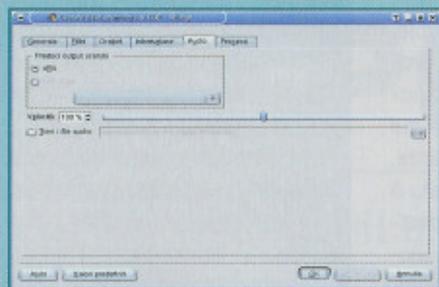
3 MODIFICA ORATORE Attraverso la finestra di configurazione degli oratori sarà possibile indicare il percorso dell'eseguibile del sintetizzatore utilizzato, selezionare la voce desiderata e modificare il volume, il tono e la velocità di pronuncia del testo. Infine, è possibile scegliere se caricare la voce all'avvio e effettuare il test di verifica, prima di confermare.

Uso avanzato di kttsmgr

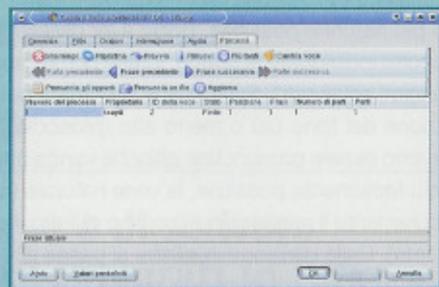
Selezione del sistema di riproduzione dell'audio e gestione dei processi di lettura



1 MESSAGGI INTERROTTI Possiamo specificare le azioni da far eseguire a *kttsmgr* ogni volta che un processo di testo verrà interrotto da un altro messaggio con maggiore priorità. La massima priorità sarà assegnata all'output del lettore dello schermo, seguito dagli avvertimenti che hanno la priorità successiva ed infine dai messaggi. Per un utilizzo base di *kttsmgr* non è necessario configurare alcun parametro aggiuntivo .SSI.



2 LA SEZIONE AUDIO È possibile scegliere il sistema di riproduzione dell'audio tra *aRTs* e *Gstreamer* nonché la velocità della pronuncia, ma solo se è stato installato il programma *sox*. Infine, abilitando la casella "Tieni i file audio" il programma copierà i file audio temporanei nella directory indicata, per poter essere riutilizzati successivamente per altri scopi. Il nome dei file avrà la forma *kttsd-P-NF*, dove *P* è il numero del processo e *NF* il numero della frase.



3 PROCESSI DI LETTURA Cliccando sulla scheda "Processi" possiamo avviare, eliminare, mettere in pausa i processi di lettura del testo, mandare avanti o indietro una frase, leggere un testo contenuto negli appunti di KDE oppure aprire un intero file di testo da leggere. La frase che verrà pronunciata, apparirà in un riquadro nella parte bassa di questa finestra. In questo modo è possibile verificare facilmente l'esattezza del testo letto.

ma di KDE, possono essere così riassunte:

- pronuncia del testo contenuto in un file;
- pronuncia degli eventi di notifica di KDE;
- pronuncia del contenuto degli appunti;
- pronuncia del testo di una pagina Web in Konqueror o parte di questo;
- pronuncia del testo nell'editor Kate o parte di questo;
- uscita audio selezionabile, a scelta, tra *aRTs* e *Gstreamer*;
- possibilità dell'utente di configurare i filtri;
- scelta dei sintetizzatori vocali;
- trasformare documenti XML;

Nel seguito dell'articolo esamineremo alcune di queste funzionalità.

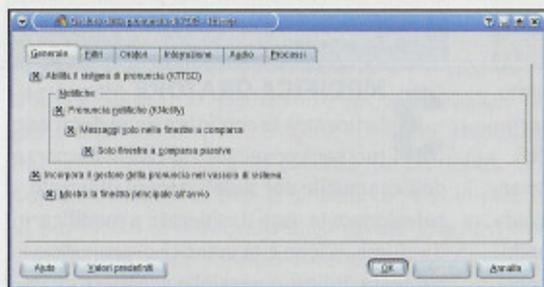


Fig. 1 • L'interfaccia di configurazione di *kttsmgr*

COME SI UTILIZZA KTTSMGR

Dopo esserci assicurati di aver installato tutto il software necessario, è arrivato il momento di avviare *kttsmgr*, il programma che utilizzeremo per dare voce al nostro Pinguino in ambiente KDE. Proseguiamo, quindi, digitando da riga di comando:

```
kttsmgr
```

oppure come alternativa, possiamo avviare *kttsmgr* dal menu principale di KDE. Al primo avvio si presenterà una finestra di configurazione composta da sei schede. È necessario abilitare, con un segno di spunta, la casella per attivare il sistema KTTS, in modo tale da avviare il demone TTS di KDE (*kttsd*). Più in basso sarà possibile anche abilitare il sistema della pronuncia delle notifiche e la segnalazione tramite un'icona presente nella barra di stato di KDE ogni volta che *kttsmgr* sarà in esecuzione. L'ultima casella ci permetterà di scegliere se avviare o meno la schermata di configurazione all'avvio di *kttsmgr*.

SINTESI VOCALE DALLA RIGA DI COMANDO

È possibile utilizzare il sistema di sintesi vocale anche al di fuori dell'ambiente

LA SCHEDA FILTRI DI KTTSMGR

I filtri sono spesso utilizzati per pre-elaborare i testi prima di inviarli al motore di sintesi vocale e possono essere utili per migliorare la pronuncia, per le abbreviazioni o per aggirare alcuni limiti dei sintetizzatori vocali. Dal momento che si tratta funzionalità avanzate di KTTS, per un utilizzo base di *kttsmgr*, non saranno necessari.

grafico KDE, eseguendo il programma direttamente dalla riga di comando.

Aprirete una console ed eseguite:

```
festival
```

Se tutto è installato correttamente, verrà visualizzato a schermo qualcosa di simile:

```
Festival Speech Synthesis System 1.95:beta
                                     July 2004
Copyright (C) University of Edinburgh,
1996-2004. All rights reserved.
For details type `(festival_warranty)'
festival>
```

Ciò indicherà che Festival è pronto in attesa di un comando da eseguire. Provate ora a digitare il seguente comando:

```
(SayText "Hello, Linux Magazine!")
```

ascolterete la fedele riproduzione della frase posta all'interno dei doppi apici.

Ovviamente, le funzionalità di festival non finiscono qui. È possibile far leggere al PC anche interi file di testo utilizzando ad esempio:

```
(tts "/path/completo/al_file" nil)
```

Nel caso in cui abbiate installato anche il software

COS'È E COSA OFFRE FESTIVAL

Si tratta di un framework scritto in C++ e sviluppato dal CSTR (*Centre for Speech Technology Research*) dell'Università di Edimburgo, per fornire una struttura generale per la

realizzazione di sistemi di sintesi vocale. La versione corrente è la 1.95-beta e offre diverse funzionalità, alcune delle più importanti possono essere così riassunte:

- supporto per i compilatori GCC versione 3.2 e 3.3;
- supporto al compilatore Intel 8.0;
- supporto al sistema operativo MacOS X;
- miglior supporto ai testi in lingua inglese, spagnolo e gallese, tedesco, finlandese, slovacco, russo, italiano, francese ed altre minori.
- moduli indipendenti della lingua, configurabili dall'utente;
- supporto Mbrola (disponibile anche per la lingua italiana);
- retrocompatibilità con le versioni precedenti di Festival.

Mbrola, è possibile cambiare la voce semplicemente tramite il comando:

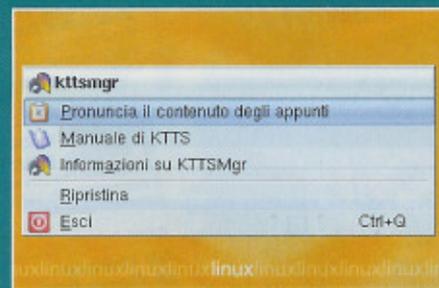
```
(voice_lp_mbrola)
```

Questo permetterà di selezionare il sintetizzatore Mbrola e di utilizzare, ad esempio, una voce femminile. Le funzionalità offerte da Festival vanno oltre, è possibile approfondire il suo utilizzo tramite la documentazione online oppure utilizzando le pagine di manuale del programma utilizzando il comando: (help).

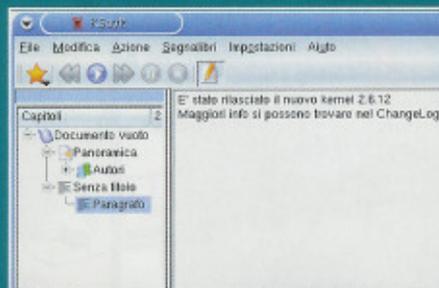
Matteo Mattei, Giovanni Scafora

Il Pinguino parla!

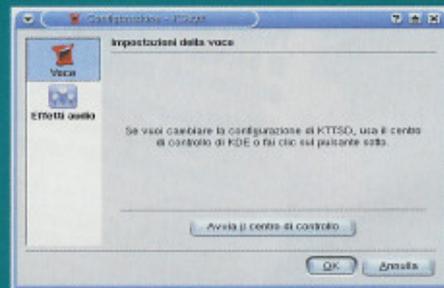
Letture di testo e personalizzazione del sistema di sintesi vocale



1 LETTURA DEGLI APPUNTI Avviamo il browser web Konqueror, selezioniamo e copiamo il testo di una pagina Web. Clicchiamo, adesso, con il tasto destro del mouse sull'icona di *kttsmgr*, residente nel vassoio di KDE, verrà mostrato il menu a tendina delle opzioni di *kttsmgr*. Selezionando "Pronuncia il contenuto degli appunti", ascolteremo la lettura del testo che in precedenza avevamo salvato negli appunti.



2 ASCOLTIAMO IL TESTO Un altro programma, incluso nella suite di KDE e utile al nostro scopo, è *ksayit*. Il programma *ksayit*, come *kttsmgr*, è avviabile sia da riga di comando sia attraverso il menu d'avvio di KDE e si presenta con una schermata molto semplice ed intuitiva. Sarà solo necessario scrivere un testo nell'apposito spazio e cliccare sul pulsante "Play" per poter ascoltare il testo immesso.



3 VOCE PERSONALIZZATA Per concludere, avviando la configurazione di *ksayit* dal menu "Impostazioni", sarà possibile cambiare la configurazione di *kttsd*, usando il centro di controllo di KDE, e abilitare o disabilitare diversi effetti audio, applicabili alla pronuncia, disponibili nel vostro sistema. In questo modo possiamo personalizzare il sistema di sintesi vocale fin nei minimi dettagli per adattarlo alle nostre esigenze.