

Progetto CLIPS
Corpora e Lessici di Italiano Parlato e Scritto

W1- a5
(Strumenti **H**ardware e **S**oftware)

Title: *Wavesurfer per CLIPS*

Document No: CLIPS/W1-a5/SHS/003

Document Name: NADL009

Status: pubblico

Date: 10/3/2006

Authors: *Massimo Petrillo*

Wavesurfer per CLIPS

Introduzione

Le attività previste dalla workpackage W1-a5 del progetto CLIPS sono consistite nello sviluppo di un'unica applicazione in grado di effettuare l'analisi automatica del segnale (estrazione pitch e calcolo della curva stilizzata, calcolo delle formanti, ecc.) insieme all'etichettatura del segnale.

Scelte effettuate

La base di partenza di tutto lavoro svolto è un sistema di etichettatura sviluppato presso il KTH di Kungl Tekniska Hogskolan di Stoccolma chiamato Wavesurfer, progettato e realizzato da Kåre Sjölander and Jonas Beskow (vedi www.speech.kth.se/wavesurfer/ per ulteriori informazioni).

La presente documentazione è quindi mirata alla sola esposizione delle modifiche apportate al programma originario. Non rientra, quindi, negli scopi di questo documento descrivere le caratteristiche generali dell'ambiente di sviluppo e della applicazione realizzata dal KTH, informazioni reperibili al sito sopra citato.

La scelta di Wavesurfer si fonda sulla possibilità di concentrare gli sforzi solo sul miglioramento e l'adattamento del software disponibile, piuttosto che la scrittura di un nuovo software da zero. Lo sviluppo di un nuovo sistema avrebbe comportato, inoltre, delle fasi di collaudo con tempi non compatibili con il resto del progetto. In questo modo il collaudo sarà volto prevalentemente alle modifiche effettuate.

L'applicazione è scritta in parte in TCL/TK, in parte in linguaggio C. Alcuni tool per l'analisi del segnale sono stati scritti in Matlab.

Dal punto legale l'adozione di questo software è possibile grazie alla licenza sotto la quale viene distribuito; essa permette infatti l'uso gratuito del software e la possibilità di modificarne il codice sorgente. In appendice viene riportata la licenza d'uso.

In questa ottica, Wavesurfer è risultato particolarmente adatto alle nostre esigenze in quanto altamente configurabile senza modificarne il codice, così le modifiche apportate riguardano solo, almeno in linea di principio, l'inserimento di nuove funzioni e non la modifica di quelle esistenti.

Modifiche apportate

Le modifiche sono essenzialmente di due tipi:

- facilitazioni nell'inserimento di dati ed etichette non previste dagli autori del prodotto originale e protezione da modifiche accidentali;
- elaborazioni del segnale.

Sono stati definite, inoltre, due configurazioni distinte per la segmentazione segmentale e per quella prosodica (ancora in corso). All'apertura di ogni file sarà richiesto di specificare con quale lavorare.

Facilitazioni nell'inserimento di dati ed etichette

Sono state apportate le seguenti modifiche allo scopo di rendere difficile modifiche accidentali dei dati inseriti

- Disabilitazione delle numerose opzioni di configurazione previste dal software originale. Un utente, non necessariamente troppo smaliziato, avrebbe potuto modificare la configurazione del software in modo da renderlo inutilizzabile per i fini del progetto. Questa scelta permette, invece, di avere una stessa configurazione per tutti gli operatori che useranno il software, con evidenti semplificazioni nella manutenzione. E' stata prevista, tuttavia, la possibilità di riabilitare queste funzioni per gli utenti avanzati (per informazioni consultare il manuale) .
- Semplificazione dell'inserimento di etichette. Gli autori originari avevano previsti due diversi menu contestuali¹ per la gestione delle etichette e per la configurazione del software: la pressione del tasto destro attivava il secondo menu mentre per maneggiare le etichette era necessario premere insieme il tasto *shift* sulla tastiera ed il tasto destro del mouse. Questa scelta non facilitò l'uso da parte di un'operatore poiché all'operazione più frequente era associata la combinazione più complessa, viceversa un'operazione piuttosto rischiosa come la configurazione del software era accessibile troppo semplicemente. Si è proceduto quindi ad associare alla pressione del tasto destro il menu di gestione delle etichette, mentre il menu di configurazione è stato reso disponibile alla pressione simultanea del tasto *shift* e del tasto sinistro del mouse².

¹ I "menu contestuali" sono quelli attivati dalla pressione del tasto destro del mouse

² Questa combinazione è inusuale per un menu contestuale, quindi viene ridotta la possibilità di errore

- Blocco dei livelli di trascrizione. Al menu di configurazione è stata aggiunta la voce “Lock pane”/”Unlock pane” il cui compito è di inibire l’uso del menu contestuale sul livello in questione con l’impossibilità di modificarne le etichette. La successiva invocazione del comando riabilita il menu contestuale.
- Immissione etichette da menu. Questa funzione è stata prevista con il duplice scopo di velocizzare l’inserimento di etichette e di favorire l’uniformità tra i vari operatori. Per le esigenze del progetto sono necessari due tipi di menu per le etichette: dipendenti dal livello e dipendenti dal segnale esaminato. Il primo tipo è utile, ad esempio, per il livello di trascrizione fonetica per il quale il menu è fisso rispetto ai segnali da analizzare; viceversa per i livelli standard e di parola il menu deve variare a seconda del segnale esaminato. Sono quindi previsti dei file aggiuntivi (oltre ai wav) per configurare questi menu.
- Conversione da trascrizione ortografica a fonetica. Un modulo sviluppato dal Politecnico di Bari, per conto della Fondazione “Ugo Bordoni”, si occupa di questa conversione allo scopo di facilitare la creazione dei menu per l’inserimento delle etichette.
- Allineamento facoltativo delle etichette tra livelli adiacenti. Nel menu contestuale sono state aggiunte le voci “Insert label up” e “Insert label down” che provvedono all’inserimento in una dei livelli adiacenti una etichetta allineata con quella corrente. Questo allineamento temporale, non essendo effettuato *a mano*, risulta più preciso.
- Inserimento di etichette in corrispondenza della selezione corrente.
- Allineamento di etichette alla selezione corrente.

Analisi del segnale

L’analisi automatica del segnale riguarda essenzialmente il calcolo delle formanti e l’estrazione dei parametri prosodici. Il calcolo delle formanti viene effettuata da una libreria a caricamento dinamico (DLL) scritta in C. Questa libreria è stata denominata FRM4SNK. Mentre l’estrazione dei parametri prosodici è effettuata da un programma scritto in Matlab (vedi [D’Anna, Petrillo 2001]).

Il calcolo delle formanti viene fatto usando l’analisi LPC del segnale. Il modello usato è quello Auto-Regressivo del 35-esimo ordine. I coefficienti vengono calcolati su finestre di 23ms sovrapposte di 12ms. I massimi della funzione di trasferimento del filtro così creato corrispondono alle formanti.

File usati e generati da Wavesurfer per CLIPS

Ogni file sonoro da analizzare deve essere accompagnato da un file di trascrizione ortografica avente lo stesso nome ma con estensione “.wr_”. Questo file deve essere presente nella stessa cartella del file audio. Tutti i file generati dal programma verranno posti nella stessa cartella del file audio, tutti con il medesimo nome del sonoro con estensione variabile.

Nella tabella che segue viene riportato l’elenco dei file generati dal programma. Tutti questi file hanno lo stesso nome ed estensione diversa.

Estensione	Descrizione	Formato
.st_	File di trascrizione fonetica del file sonoro Usato per la creazione del menu delle etichette	Testo separato da spazi
.add	Informazioni aggiuntive nella configurazione segmentale	TIMIT
.wr	Segmentazione ortografica nella configurazione segmentale	TIMIT
.std	Segmentazione fonetica standard per parole nella configurazione segmentale	TIMIT
.phn	Segmentazione fonetica standard nella configurazione segmentale	TIMIT
.acs	Segmentazione acustica nella configurazione segmentale	TIMIT
.frm	Formanti	File di testo ogni riga indica un frame, il formato è il seguente: posizione[ms] formante1[Hz] formante2[Hz] ...
.ptc	Pitch	File di testo ogni riga indica un frame, il formato è il seguente: posizione[ms] pitch[Hz]
<i>Da definire</i>	<i>File di segmentazione nella configurazione prosodica</i>	TIMIT

Il formato TIMIT prevede file di testo le cui righe sono associate ad un’etichetta secondo il seguente formato:

campione_di_inizio campione_di_fine etichetta

I primi due campi sono dei numeri interi, l’ultimo è una stringa non contenete spazi.

Il campione di inizio di una etichetta coincide con quello di fine della precedente.

Bibliografia

- J.C. Junqua, J.P. Haton “Robustness in automatic speech recognition” Kluwer Academic Publishers 1996
- M. Petrillo “Algoritmi di divisione del segnale verbale in unità sillabiche” tesi di laurea, 2000
- L. D’Anna, M. Petrillo “APA, un prototipo di sistema automatico per l’analisi prosodica”, atti del GFS Padova, 2001, in stampa
- The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus,
http://www ldc.upenn.edu/readme_files/timit.readme.html
- <http://www.speech.kth.se/wavesurfer>
- <http://www.speech.kth.se/snack>

Appendice

Licenza

Copyright (C) 2000-2001 Jonas Beskow and Kåre Sjölander

The following terms apply to all files associated with the software unless explicitly disclaimed in individual files.

The authors hereby grant permission to use, copy, modify, distribute, and license this software and its documentation for any purpose, provided that existing copyright notices are retained in all copies and that this notice is included verbatim in any distributions. No written agreement, license, or royalty fee is required for any of the authorized uses. Modifications to this software may be copyrighted by their authors and need not follow the licensing terms described here, provided that the new terms are clearly indicated on the first page of each file where they apply.

IN NO EVENT SHALL THE AUTHORS OR DISTRIBUTORS BE LIABLE TO ANY PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OF THIS SOFTWARE, ITS DOCUMENTATION, OR ANY DERIVATIVES THEREOF, EVEN IF THE AUTHORS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

THE AUTHORS AND DISTRIBUTORS SPECIFICALLY DISCLAIM ANY WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NON-INFRINGEMENT. THIS SOFTWARE IS PROVIDED ON AN "AS IS" BASIS, AND THE AUTHORS AND DISTRIBUTORS HAVE NO OBLIGATION TO PROVIDE MAINTENANCE, SUPPORT, UPDATES, ENHANCEMENTS, OR MODIFICATIONS.