TECNICHE DI SCRIPTING PER L'ANALISI ACUSTICA

Lorenzo Cioni
Laboratorio di Linguistica, Scuola Normale Superiore, Pisa

1. INTRODUZIONE

Scopo del presente lavoro è quello di illustrare l'utilità e l'uso di due tipologie di tecniche di scripting, che diremo rispettivamente tecniche sincrone e tecniche asincrone, per l'esecuzione di operazioni ripetitive per l'analisi acustica di segnali di parlato.

Nei paragrafi che seguono vengono inizialmente svolte alcune considerazioni sulla necessità di standardizzare l'esecuzione di sequenze di operazioni, si passa quindi alla definizione delle tecniche di scripting asincrone e sincrone per terminare con la presentazione di due case study relativi ai programmi Multi-Speech™ e ESPS™.

ESPS sta per "Entropic Signal Processing System", software package che rappresenta uno strumento potente e versatile per l'elaborazione dei segnali in ambiente Unix. ESPS può essere visto come una estensione di Unix ed è composto da più di duecento programmi utilizzabili dall'utente singolarmente o all'interno di script.

Multi-Speech è un prodotto software utilizzabile per l'acquisizione, l'analisi acustica e il playback di segnali di parlato e mette a disposizione degli utenti un ricco set di operazioni che possono essere raggruppate all'interno di script che, in ambiente Multi-Speech, sono denominati macro.

La differenza sostanziale, dal punto di vista del presente lavoro, risiede nelle differenze concettuali che esistono fra gli script di ESPS e le macro di Multi-Speech dal momento che i primi sono una estensione delle seconde: in prima approssimazione si può affermare che se una macro è una lista di comandi che permette agli utenti di ripetere sempre nello stesso ordine una serie di operazioni uno script in più ha strutture di controllo e istruzioni per l'accesso ai file.

2 L'ESECUZIONE DI OPERAZIONI RIPETITIVE PER L'ANALISI ACUSTICA

Le operazioni che, di solito, sono eseguite sui file di parlato sono, in genere, operazioni di analisi (calcolo dello spettro e dello spettrogramma, estrazione di F0 e analisi LPC) o di calcolo di durate nel dominio del tempo (VOT e Rise Time, ad esempio) eseguite su file o porzioni di file individuate da etichette posizionate dall'utente in modo da evidenziare le sezioni dei file in cui sono localizzati i fenomeni acustici di interesse. L'esecuzione di tali operazioni su grossi corpora di dati per la validazione di teorie linguistiche richiede l'adozione di cautele procedurali che sono tanto più gravose quanto più ampia è la mole dei dati da sottoporre ad analisi.

Una macro o uno script sono, in prima approssimazione, dei file di comandi ovvero dei file di testo contenenti successioni di comandi che vengono eseguiti sempre nello stesso ordine e con gli stessi valori dei parametri significativi.

L'adozione di macro e di script consente, pertanto, la ripetizione di successioni di comandi sempre nello stesso ordine. In tal modo si ha la possibilità di analizzare senza difficoltà grossi corpora di dati e di implementare protocolli di ricerca che, applicati a tali corpora, danno origine a dei risultati che non risentono dell'ordine di esecuzione delle singole operazioni. Si ha così la certezza di non aver introdotto variazioni accidentali nei parametri utilizzati per l'analisi o nella successione delle operazioni, variazioni che possono falsare i risultati delle analisi portando, al limite, alla validazione erronea di teorie linguistiche.

3 TECNICHE DI SCRIPTING SINCRONE E ASINCRONE

Le tecniche di scripting[i] utilizzabili per l'esecuzione di operazioni di cui al punto precedente possono essere classificate come asincrone oppure sincrone.

Nel caso di tecniche sincrone l'esecuzione delle operazioni contenute nello script o nella macro è guidata da azioni dell'utente che esegue o lo script o la macro dall'interno di un programma per il DSP (per Multi-Speech è il programma stesso, nel caso di ESPS è possibile usare sia EnSig[ii] sia xwaves, [5]). L'utente, in questo caso, imposta lo spazio di lavoro, carica uno o più file necessari, setta su di essi dei marker che individuano istanti temporali ed usa tali marker per eseguire operazioni che producono nuovi file e operazioni che li visualizzano arricchendo lo spazio di lavoro ed eventualmente creando risultati numerici che possono essere salvati in file di testo.

Nel caso ti tecniche asincrone l'utente esegue uno o più script in modo disgiunto dal programma per il DSP ottenendo, a partire da file di solito contenenti porzioni di parlato, file il cui contenuto è il risultato di operazioni di analisi e che possono essere visualizzati, nel caso specifico, sia con comandi di ESPS sia all'interno di EnSig e/o xwaves. Nel caso di xwaves ([7]) un possibile protocollo che può essere utilizzato è il seguente: l'utente visualizza un file di parlato, posiziona su di esso etichette che vengono salvate in file di testo e alle quali corrispondono marker temporali e poi usa tali marker (e gli associati valori di posizione) in script ad hoc per l'esecuzione di operazioni su quel file di parlato (e altri file etichettati in modo simile).

4 CASE STUDY 1: MULTI-SPEECH

Multi-Speech ([6]) permette di usare esclusivamente tecniche sincrone dal momento che le macro sono eseguibili solo dall'interno del programma e non sono previsti strumenti né per l'accesso ad insiemi di file individuati da pattern né per l'accesso a porzioni di file individuate da marker inseriti dall'utente in altri file (vedi oltre).

La creazione di una macro richiede da parte dell'utente per prima cosa la creazione di un file di configurazione che contiene sia la descrizione dello spazio di lavoro sia i valori dei parametri utilizzati dai comandi di acquisizione, di analisi e di visualizzazione. Le macro possono essere poi create sia usando la modalità di LEARN di Multi-Speech sia inserendo comandi in file di testo utilizzando un qualunque editor. La prima modalità presenta una serie di limitazioni ma è utile in fase di progettazione e di apprendimento.

Per illustrare alcune caratteristiche delle macro di Multi-Speech si riportano qui di seguito tre macro usate da Silvia Calamai per alcuni suoi lavori di analisi del vocalismo tonico e atono nella varietà pisana di italiano ([1]). Le tre macro, eseguibili sia dalla linea di comando di Multi-Speech usando il comando MACRO RUN sia dalla voce Run del menù Macros, usano il file MASTER.INI come file di configurazione per impostare sia lo spazio di lavoro sia i parametri per l'analisi. La prima macro (detta master) consente all'utente di caricare un file di parlato (LOAD), ascoltarlo (SPEAK) e selezionarne una porzione su cui potrà eseguire in successione le due macro slaves (slave1 e slave2).

```
* Macro slaves:
* slave1
```

```
* slave2
SET FROM.NOSAVE C:\Kaylab\ini\user\MASTER.INI
CLOSE ALL
WINDOW RESET WORKAREA
USE A
MESSAGE Carica un file di speech
LOAD
SOURCE =
INFO
SPEAK A 0 *
TITLE Speech
MESSAGE Ora scegli una porzione di segnale e poi esegui la macro successiva quando hai selezionato la porzione desiderata
SET FROM.NOSAVE INI
```

La seconda macro (slave1, vedi qui di seguito) copia in una nuova window la porzione di segnale selezionata dall'utente (USE B e COPY A SS SE) e su di essa esegue il calcolo dello spettrogramma, (SPG B 0 *), l'estrazione della F0 (PITCH B 0 *) e sia il calcolo sia il tracciamento (in rosso, SET WINDOW.PEN RED) delle formanti (FMT B 0 *).

```
* Macro slave1
SET FROM.NOSAVE
C:\Kaylab\ini\user\MASTER.INI
USE B
COPY A SS SE
SPEAK B 0 *
TITLE Speech Selection
USE C
SPG B 0 *
USE D
PITCH B 0 *
USE B
LINK C
LINK D
USE C
SET WINDOW.PEN RED
FMT B 0 *
SET FROM.NOSAVE INI
```

```
* Macro slave2
SET FROM.NOSAVE
C:\Kaylab\ini\user\MASTER.INI
SOURCE B
WINDOW
LPC ! =
RESULT
SET FROM.NOSAVE INI
```

La terza macro (slave2), infine, esegue un'analisi LPC (LPC ! =) sulla porzione di file selezionata (SOURCE B) e visualizza il risultato su una nuova finestra (WINDOW) oltre a visualizzare (RESULT) i risultati di tutte le analisi svolte in una finestra di dialogo che ne consente il salvataggio in un file di testo, utile per elaborazioni successive.

Da un esame anche sommario di tali macro risulta evidente che la gestione della esecuzione è a totale carico dell'utente che deve localizzare e visualizzare un file di parlato, selezionarne la porzione desiderata e poi passare alla esecuzione delle macro slave1 e slave2 nell'ordine, dal momento che la macro slave2 fa uso della finestra in cui la macro slave1 copia la sezione del file di parlato, sezione su cui viene eseguita l'analisi LPC (vedi i comandi SOURCE B e LPC ! =).

5 CASE STUDY 2: ESPS

Nel caso di ESPS[iii] ([2]) l'utente ha a disposizione tecniche di scripting sia sincrone sia asincrone. Le prime ([7]) prevedono che l'utente esegua singoli comandi o script attraverso il pannello di controllo principale di xwaves, comandi che possono essere diretti sia al programma (comandi globali) sia ad un oggetto visualizzato dal programma (comandi puntuali). Le seconde prevedono l'uso di script immessi sulla linea di comando di una shell di Unix.

Esempi di tecniche del primo tipo sono i seguenti ([7]):

waves make file DgmtA01F.wav	comando globale che carica il file indicato e lo visualizza in una window di segnale;
. activate file DgmtA01F.wav.out op set id F0	comando puntuale che consente la sola visualizzazione del grafico della F0

Un esempio delle tecniche del secondo tipo è dato dallo script seguente ([4]):

```
#!/bin/csh
foreach f (*.wav)
    get_f0 $f $f.f0
end
```

Tale script può essere eseguito come un qualunque comando di Unix e consente l'estrazione del pitch (comando get_f0) da un certo numero di file di parlato contenuti nella directory di lavoro e i cui nomi sono individuati da $f. I file prodotti dallo script hanno nomi $f.f0. Una caratteristica di tale script è quella di essere completamente autonomo rispetto a xwaves ed EnSig pur facendo uso di un comando di ESPS (get_f0).

Volendo far interagire tale script con xwaves si può fare uso del comando "send_xwaves" che implementa l'interazione con xwaves sia dalla linea di comando (ad esempio è possibile far visualizzare da xwaves un file di parlato con il comando "send_xwaves make name speech file speech.wav") sia dall'interno di script di Unix.

La versione seguente dello script precedente ([4]) calcola il pitch per ciascun file e poi visualizza in una finestra il file di parlato e in un'altra il pitch e altre grandezze (ovvero l'energia RMS, la curva di probabilità sordo/sonoro e il valore di picco dell'autocorrelazione):

```
#!/bin/csh
foreach f (*.wav)
    get_f0 $f $f.f0
    send_xwaves make name speech file $f
    send_xwaves make name speech file $f.f0
    send_xwaves pause
    if ( $status != 0 ) then
        exit
    endif
```

```
    send_xwaves kill
end
```

L'esecuzione è resa sincrona (ovvero dipendente da azioni dell'utente dall'interno di xwaves) dalla presenza del comando "send_xwaves pause": tale comando permette all'utente di analizzare con calma il contenuto di una coppia di finestre per poi passare alla successiva premendo il pulsante CONTINUE presente sul pannello di controllo di xwaves. Scopo del comando "send_xwaves kill" è quello di rimuovere le finestre non più utili per impedire un inutile affollamento dello schermo.

Come ultimo esempio illustrativo delle potenzialità di ESPS si riporta qui di seguito il testo di uno script tratto da [7] e il cui compito principale è quello di semiautomatizzare l'etichettatura di un gruppo di file di parlato dei quali provvede a calcolare anche lo spettrogramma.

```
#!/bin/csh –f
send_xwaves set middle_op 'blow up; function'
send_xwaves set ref_size 3 ref_step 2.5 ref_start 0.0
send_xwaves attach function xlabel
foreach file ($*)
    set bn=`basename $file`
    send_xwaves make name $bn file $file loc_x 0 loc_y 150 height 200
    send_xwaves $bn spectrogram loc_x 0 loc_y 375
    send_xwaves send make name $bn file file.1.lab color 125
    send_xwaves send make name $bn file file.2.lab color 118
    send_xwaves send activate name $bn fields 1 3 5
    send_xwaves pause
    if ( $status != 0 ) then
        exit
    endif
    send_xwaves kill
end
```

Lo script suddetto permette l'elaborazione di un certo numero di file i cui nomi sono passati sulla linea di comando come parametri. Di ciascun file viene visualizzata la forma d'onda (send_xwaves make name $bn file $file loc_x 0 loc_y 150 height 200) e viene calcolato e visualizzato lo spettrogramma (send_xwaves $bn spectrogram loc_x 0 loc_y 375). Fatto questo, lo script associa a ciascun file di parlato (e all'associato spettrogramma) due file di etichette (con i comandi send_xwaves send make name $bn file file.1.lab color 125 e send_xwaves send make name $bn file file.2.lab color 118) sfruttando il fatto che prima di iniziare a lavorare sui singoli file l'utente ha aperto il programma di etichettatura (send_xwaves attach function xlabel) e si mette in attesa di un'azione dell'utente (comando send_xwaves pause) prima di ripulire lo schermo e passare al file successivo. In questo modo l'utente può lavorare a suo piacimento su ciascun file per passare al successivo solo quando ritiene di aver terminato, avendo la possibilità di riprendere il lavoro di etichettatura in un secondo tempo, semplicemente rieseguendo lo script.

Lo script illustra altre due caratteristiche molto utili di ESPS: la possibilità di aggiungere comandi personalizzati ai menù di xwaves e la possibilità di interagire con programmi ausiliari, i cosiddetti "attachments".

La prima possibilità è illustrata dai comandi "send_xwaves set middle_op 'blow up; function'" e "send_xwaves set ref_size 3 ref_step 2.5 ref_start 0.0" che consentono all'utente di definirsi un comando attivabile con il pulsante centrale del mouse per ingrandire una porzione del segnale attorno alla posizione corrente del cursore. Tale comando rende più agevole il posizionamento di marker etichettati sulla forma d'onda del segnale.

La seconda possibilità è illustrata dal comando "send_xwaves attach function xlabel" con il quale viene richiamato il programma "xlabel" ([7]) che consente di visualizzare ed editare fino a 32 file di etichette contemporaneamente. Altri attachment predefiniti sono "xspectrum" (che consente di eseguire sui file visualizzati da xwaves una power spectrum analysis parametrica), xmarks (che consente di assegnare temporizzazioni a sequenze predefinite di etichette) e "xchart" (una estensione di "xlabel") ma l'utente può scrivere attachment personalizzati usando il linguaggio Tcl/Tk e alcune funzioni ad hoc per interagire con xwaves. Per ulteriori dettagli si rimanda ai manuali citati in Bibliografia, in particolare a [7].

CONCLUSIONI

Il presente lavoro contiene una breve introduzione a problematiche vaste e complesse su cui ho in programma di tornare in un futuro prossimo con la stesura sia di macro per Multi-Speech sia di script per ESPS.

RINGRAZIAMENTI

Vorrei ringraziare Barbara Gili Fivela e Silvia Calamai per avermi dato, a vario titolo e in tempi diversi, l'occasione di avvicinarmi a queste problematiche e Franco Ferrero per una epifania avuta leggendo il suo prezioso articolo ([3]).

BIBLIOGRAFIA

[1] Calamai S., Vocali atone e toniche a Pisa, *Atti delle XII Giornate di Studio del Gruppo di Fonetica Sperimentale,* Macerata, 13-15 Dicembre 2001, in corso di stampa.
[2] AAVV, *ESPS Programmer's Manual,* Entropic Inc., 1998.
[3] Ferrero F., User's Manual of C:\CSL50\MACROS\FORMANTI.FRA [-> ALT-R], *Quaderni del Centro di Studio per le Ricerche di Fonetica*, Volume XIV, 1995:395-402.
[4] Gili Fivela B., *Script di ESPS,* strumenti di lavoro, Laboratorio di Linguistica, Scuola Normale Superiore.
[5] AAVV, *Introducing esps/waves+,* Entropic Inc., 1998.
[6] AAVV, *Multi-Speech, Software Instruction Manual,* Kay Elemetrics Corp, 1999.
[7] AAVV, *waves+ Manual,* Entropic Inc., 1998.

---

i Indicheremo nel seguito con la dizione di "tecniche di scripting" le tecniche necessarie per la stesura di script e di macro. La distinzione sarà chiarita, di volta in volta, dal contesto.

ii EnSig è, in pratica, una interfaccia tramite la quale l'utente può accedere sia a waves+ sia a ESPS sia ai programmi per la gestione della scheda audio sul sistema ospite mentre xwaves è il programma che fornisce l'interfaccia grafica per l'accesso a waves+.

iii Simbologie e convenzioni tipiche del Sistema Operativo Unix usate nel paragrafo sono date per familiari al lettore.