



Loquendo
VOCAL TECHNOLOGY AND SERVICES

**Uno strumento per l'annotazione
e la modellizzazione prosodica
di enunciati marcati
per un sistema di sintesi vocale**

**Andrea Panizza
Francesca Tini Brunozzi
Enrico Zovato**

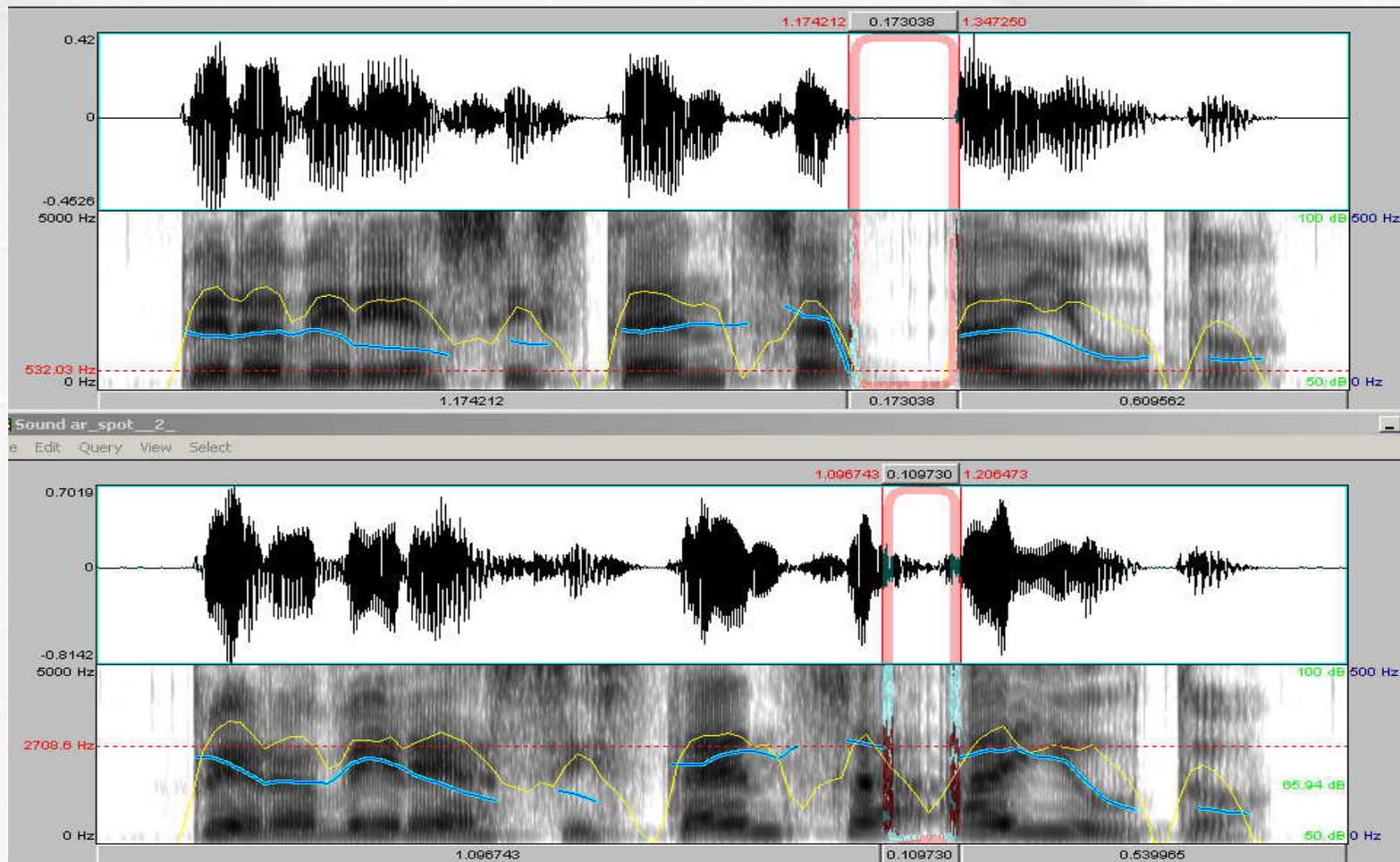
Fisciano (SA), 30 Novembre – 2 Dicembre 2005

Premesse del lavoro

- **Ambito generale: ricerca di correlati acustico-prosodici tra stili di testo e stili di lettura per un sistema di sintesi vocale da testo scritto (TTS).**
 - progettazione di un tool per l'annotazione e l'analisi prosodica di enunciati marcati (focus contrastivi, atti illocutivi, parlato emozionale, ecc.).
- **Ambito specifico: modellizzazione dei parametri acustici e degli andamenti prosodici di enunciati sintatticamente marcati.**
 - migliorare la resa acustica della sintesi di frasi interrogative di tipo wh-.

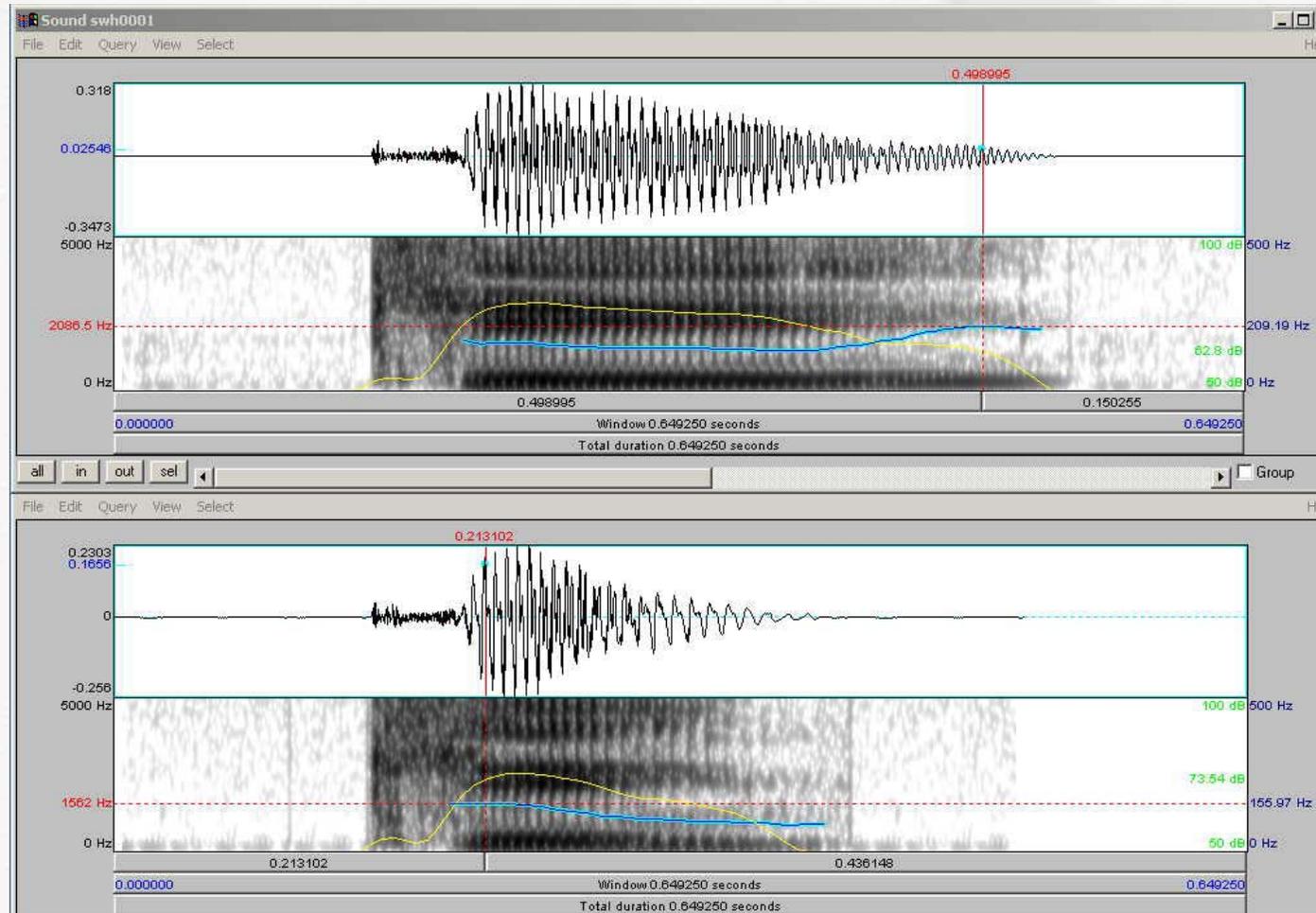
Ambito generale

La presenza di *focus* in enunciati marcati ha messo in evidenza particolari correlazioni tra il livello pragmatico e il livello acustico.



Ambito specifico

Confronto tra gli andamenti di f_0 di un enunciato interrogativo e dello stesso enunciato letto con intonazione dichiarativa.



Il tool *Sybilla* per l'annotazione prosodica

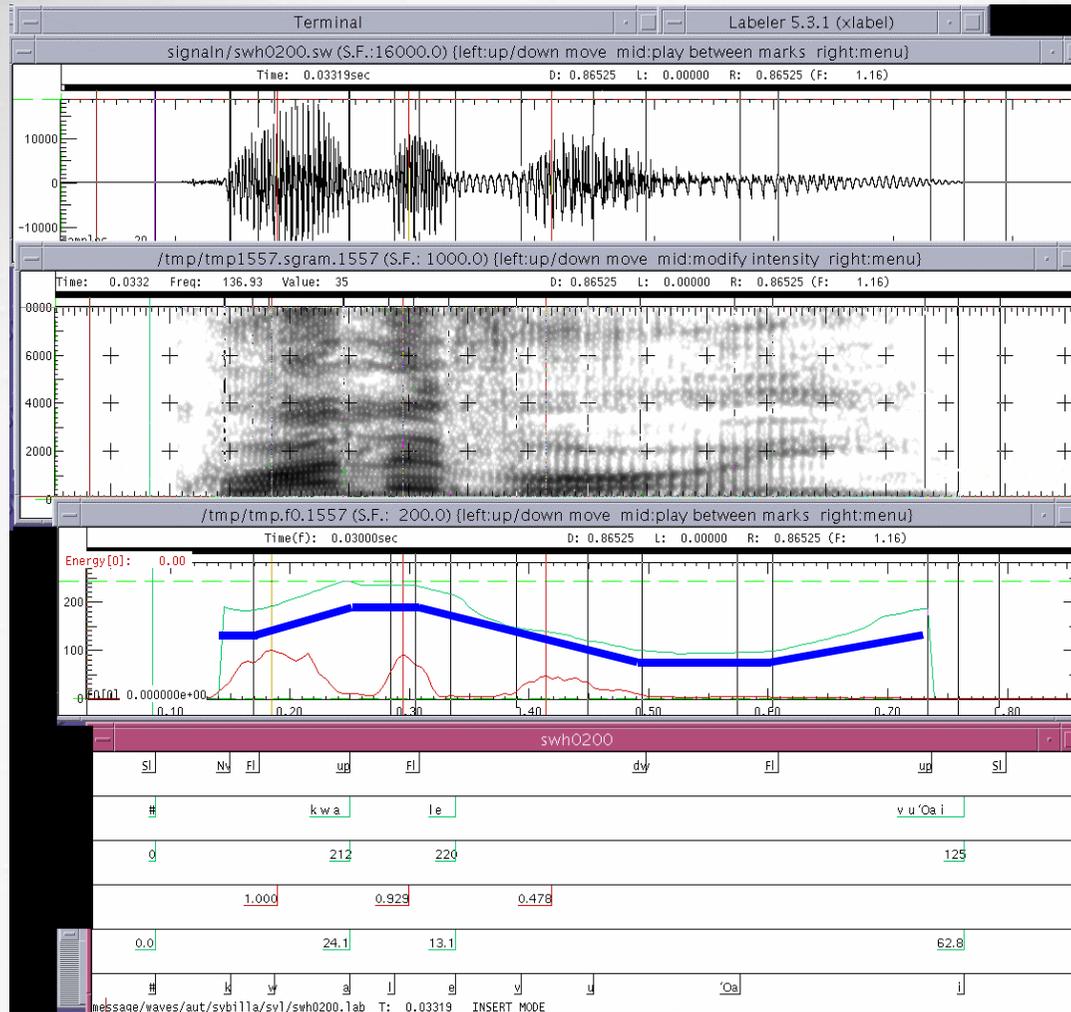
- Sviluppo di un tool (con XWaves Entropic) per l'annotazione prosodica che permette di:
 - descrivere qualitativamente e morfologicamente gli andamenti di f_0 .
 - fornire misure circa i parametri di energia e durata.
- Le informazioni vengono fornite dal tool a livello di segmentazione in sillabe, intese come unità acustiche e non grammaticali.

Tutto a posto?

t`u - t: o a - p`o s - t o

- Il tool è in grado di fornire in modo automatico, per ogni sillaba acustica, i valori di energia e durata, il valore medio di f_0 , e la visualizzazione dei confini dei fonemi.
- L'annotazione manuale prevede invece l'inserimento di etichette morfologiche atte a rappresentare l'andamento della curva di f_0 .

Interfaccia e livelli di annotazione



•Curve di **f0** e **Energia**
(con esempio di **stilizzazione**)

•Etichette prosodiche

•Unità sillabiche

•Valori di f0

•Energia normalizzata

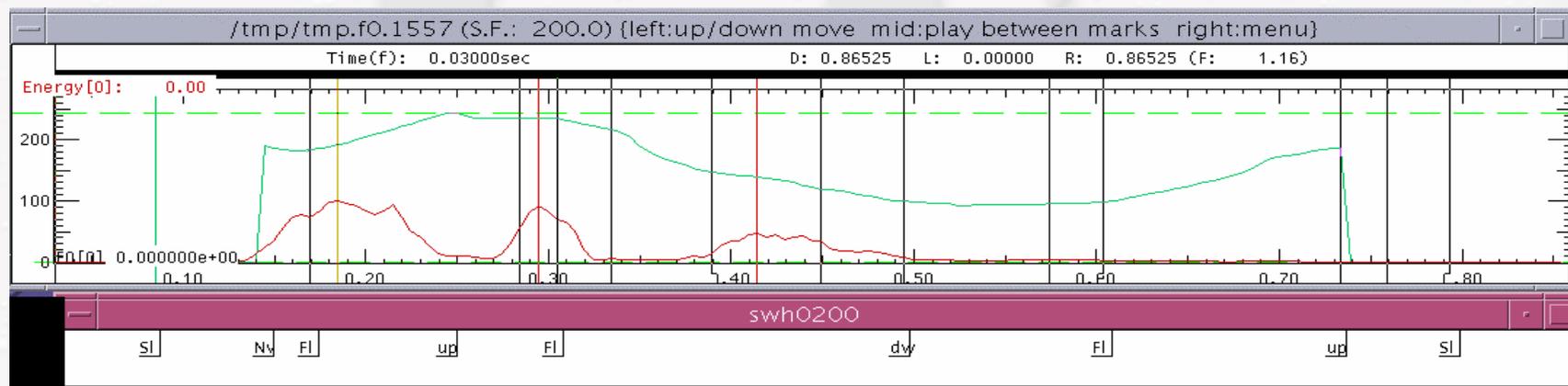
•Durata percentuale

•Segmentazione fonetica

Le etichette morfologiche per l'annotazione

Al fine di descrivere le caratteristiche della curva di f_0 , è stato definito un alfabeto di etichette morfologiche, che sono state inserite ad ogni cambiamento significativo della pendenza della curva intonativa:

- crescita moderata **up**
- crescita rapida **UP**
- decrescita moderata **dw**
- decrescita rapida **DW**
- andamento piatto **FI**
- non vocalizzato **Nv**
- silenzio **SI**



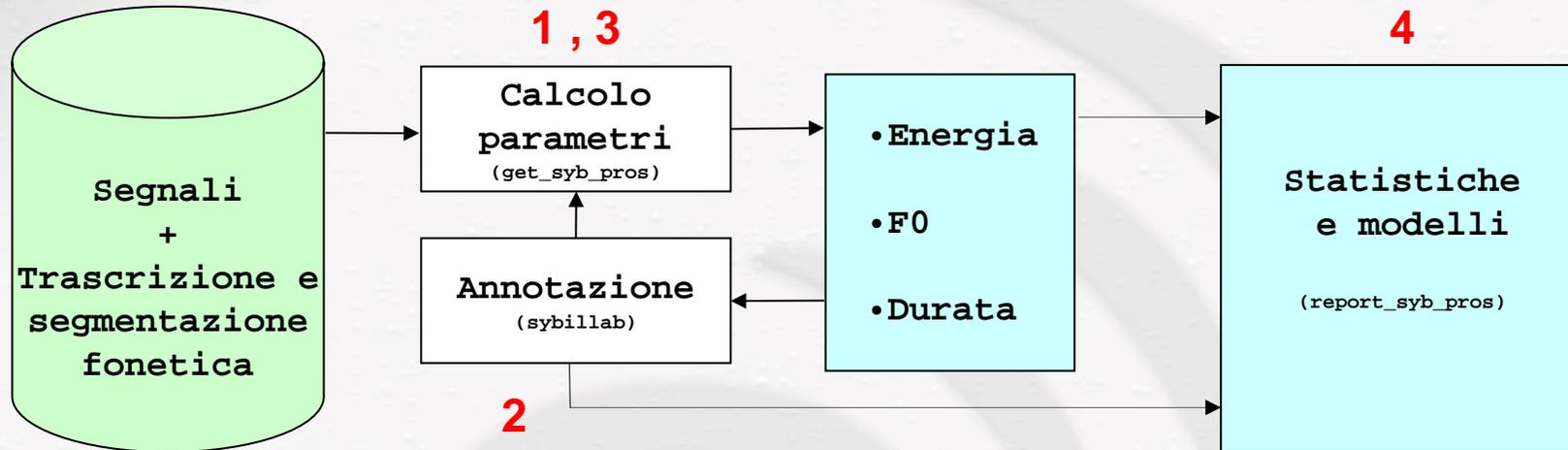
Le soglie di valore per l'annotazione

- Il lavoro di annotazione prosodica ha suggerito la scelta di tre intervalli utili a discriminare gli andamenti in casi ambigui alla sola percezione visiva.
- Dato $R=100*\text{abs}(f0_fine-f0_inizio)/(t_fine-t_inizio)$ [Hz/sec.]
 - per $R<1,5$ → andamento **FI**
 - per $1,5<R<20$ → andamento **up/dw**
 - per $R>20$ → andamento **UP/DW**
- Il lavoro di etichettatura manuale della curva di f_0 ha comportato la scelta di una finestra di analisi costante per la visualizzazione del segnale (circa 50 ms), allo scopo di uniformare la visualizzazione di segnali di durata diversa, al fine di garantire la maggior uniformità possibile tra i diversi annotatori.

Le caratteristiche tecniche del tool *Sybilla*

- Segmentazione automatica in unità fonetiche (HMM) e individuazione dei confini di sillaba in coincidenza dei fonemi che sono minimi relativi di sonorità.
- Calcolo della frequenza fondamentale: finestre di analisi di 10 ms e passo di avanzamento di 5 ms.
- Calcolo dell'energia RMS effettuato ad intervalli di tempo pari a 5 ms.
- Calcolo della durata percentuale a partire dalla segmentazione in sillabe.

Protocollo Sybilla

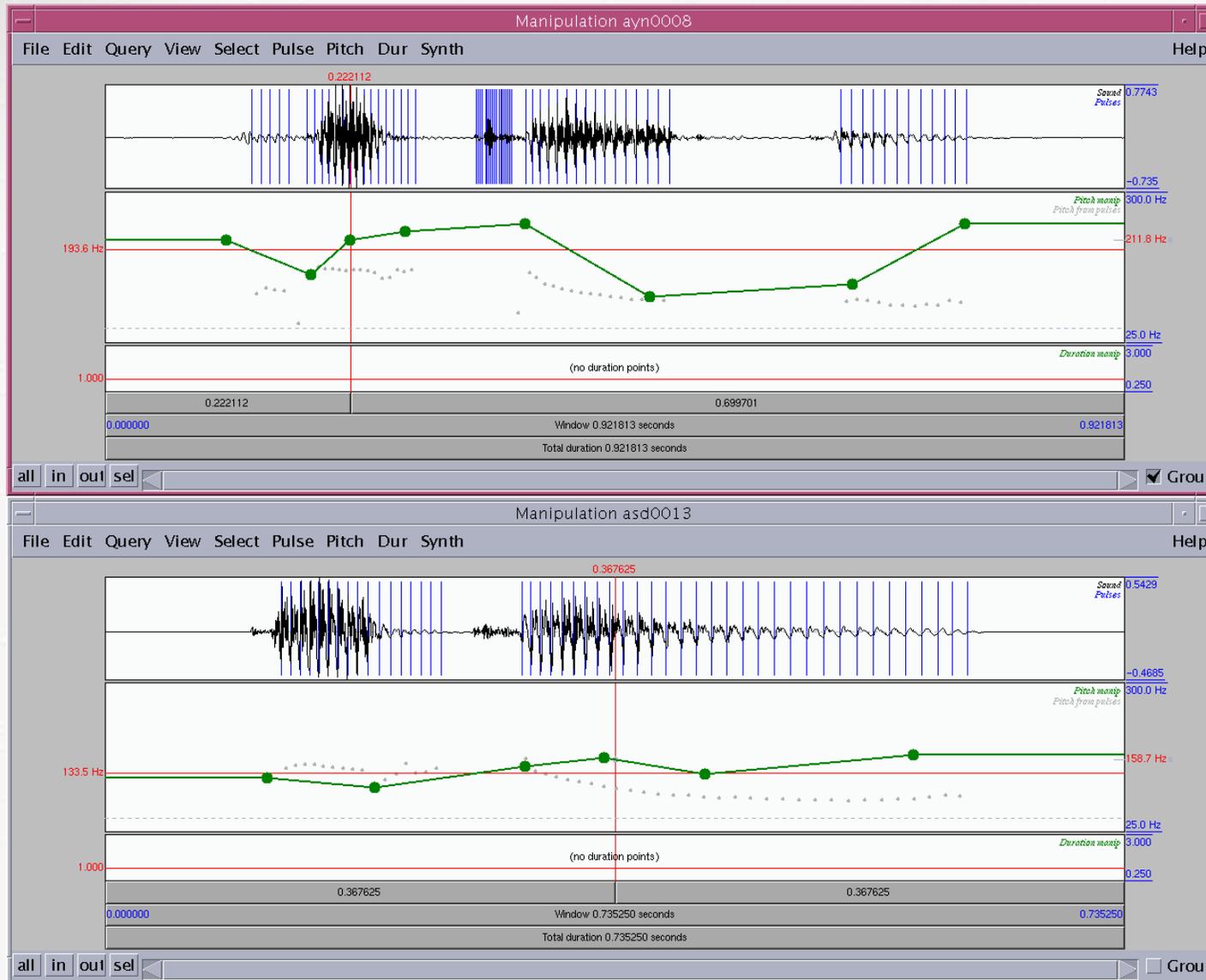


1. Analisi automatica e aggiornamento dati.
2. Intervento manuale: controllo fonetico (trascrizione e segmentazione automatica) ed etichettatura prosodica.
3. Aggiornamento dati in base a correzioni e annotazioni.
4. Elaborazione dei dati e delle etichette, modellizzazione.

Esperimenti di *prosody transplantation*

- L'annotazione degli enunciati interrogativi con il tool, e la successiva analisi compiuta su di essi, hanno fornito i modelli utili alla produzione di frasi interrogative mediante la manipolazione dei parametri prosodici delle corrispondenti frasi dichiarative.
- Tramite il programma *Praat*, a scopo di studio e verifica, viene imposto a una frase dichiarativa un profilo di f0 stilizzato linearmente e manipolato coerentemente con le annotazioni prosodiche della corrispondente frase interrogativa (annotazioni Sybilla). La modifica di f0 avviene tramite tecniche di tipo PSOLA.

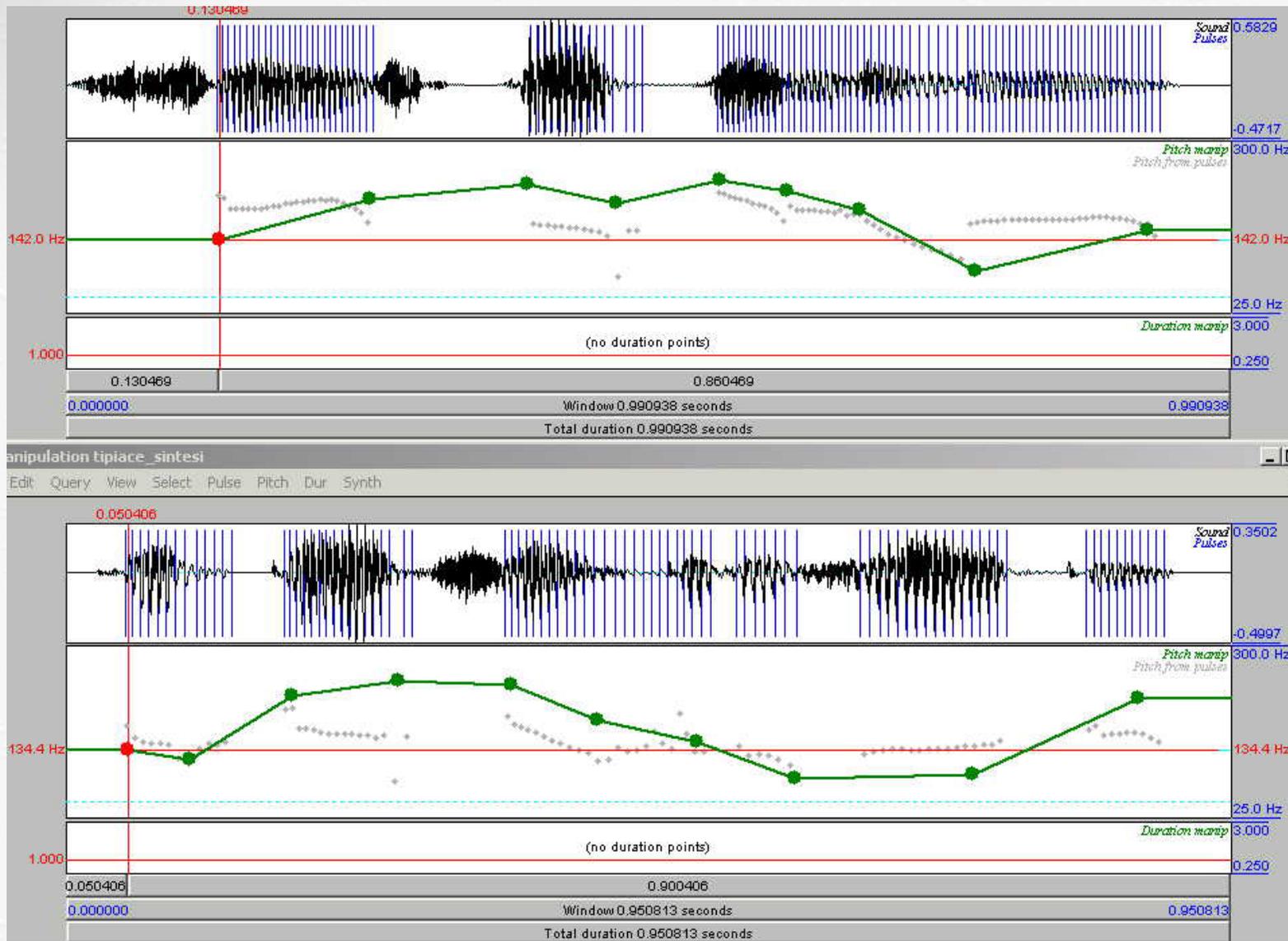
Esempi di *prosody transplantation*



La manipolazione di enunciati sintetici

- La generalizzazione della *prosody transplantation* è costituita dagli esperimenti di 'risintesi'.
- Definiamo 'risintesi' la possibilità di ottenere una frase interrogativa a partire da sintesi dichiarativa-neutra.
- A tale scopo, al segnale ottenuto in sintesi è stato imposto l'andamento di f_0 di una frase del repertorio di frasi interrogative con stesso numero di sillabe e stessa posizione degli accenti.
- In questa fase del lavoro, quindi, si cerca di rendere applicabile alle frasi sintetizzate la tecnica di manipolazione di f_0 , con modelli ottenuti tramite comparazione di annotazioni.

Esempi di risintesi

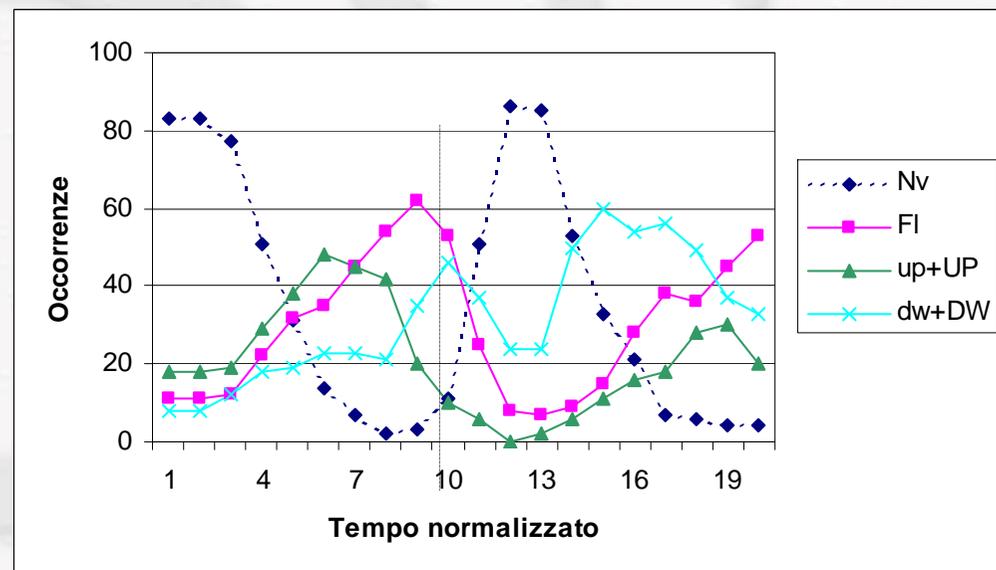


Elaborazione dei dati e sviluppi futuri

- Lo scopo finale di questo lavoro è l'utilizzo delle annotazioni per produrre modelli che predicano in modo semplice la struttura prosodica delle unità sillabiche, in modo da pilotare efficacemente la selezione di unità acustiche in un sistema di sintesi vocale da testo scritto.
- In prospettiva, l'annotazione potrebbe essere automatizzata e potrebbero essere usati strumenti di classificazione (ad esempio *CART*) per la generazione di modelli prosodici.
- Semplici aggregazioni hanno comunque evidenziato, nel caso specifico delle frasi interrogative "wh-", strutture comuni a sillabe aventi stesse caratteristiche posizionali e accentuali.

Esempio:

distribuzione etichette nel caso in cui la prima sillaba sia atona e la seconda tonica, in frasi *wh-*



Conclusioni

- E' stato presentato uno strumento per l'annotazione prosodica di enunciati, mediante la possibilità, sia di correggere la trascrizione e la segmentazione sillabica fornite automaticamente, sia di inserire etichette atte a descrivere la morfologia della curva intonativa.
- Nonostante la sinteticità delle etichette utilizzate, la "simulazione" di andamenti intonativi dedotti dall'annotazione, ha fornito risultati interessanti e credibili a livello percettivo.
- Il naturale sviluppo di questo lavoro è il miglioramento dell'annotazione prosodica per un suo efficace utilizzo in un sistema di sintesi vocale.