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
Loquendo S.p.A., Torino
{andrea.panizza, tini.brunozzi}@guest.telecomitalia.it; enrico.zovato@loquendo.com

1. SOMMARIO

Questo lavoro si inserisce nell'ambito di un obbiettivo più vasto rivolto all'individuazione dei correlati acustico-prosodici tra stili di testo e stili di lettura per un sistema di sintesi vocale *text-to-speech* (TTS)[1].

In particolare, si fa qui riferimento a uno strumento per l'annotazione e la modellizzazione prosodica di enunciati sintatticamente marcati (Avesani *et al*., 2000) per la riproduzione acustica di determinate funzioni pragmatiche in relazione allo sviluppo di applicazioni che utilizzino la sintesi vocale nell'interazione uomo-macchina.

La necessità specifica di migliorare la lettura di frasi interrogative, senza compromettere la naturalezza timbrica raggiunta con la tecnica *Unit Selection* (Balestri *et al*., 1999) per la sintesi di frasi dichiarative con stile di lettura neutro, ha orientato il nostro lavoro all'analisi prosodica di segnali acustici di frasi sintatticamente marcate come domanda (in italiano, le domande di tipo *wh*-).

Tale analisi necessita di un apposito progetto di frasi interrogative che non sono quindi più progettate solamente per realizzare una base dati acustica specializzata quanto, piuttosto, per comporre un repertorio di frasi con funzione pragmatica di 'atto di domanda', cioè azioni linguistiche (nel nostro caso, domande) che, nel contesto del dialogo uomo-macchina, elicitano altre azioni linguistiche (nel nostro caso, risposte).

2. IL PROGETTO DEL *CORPUS*

Il progetto più generale dei testi per le frasi interrogative, a differenza di quello per le dichiarative, non si fonda quindi sul criterio di copertura statistica condotto su un ampio *database* testuale ricavato da testi scritti (articoli di giornale), la cui modellizzazione non sarebbe significativa poiché si tratta spesso di domande retoriche e di periodi lunghi con fenomeni di incassatura sintattica che ne renderebbero difficile le lettura e una adeguata resa prosodica, bensì sull'identificazione di un insieme ridotto ma significativo di enunciati interrogativi, definiti come frasi a un solo sintagma, con funzione pragmatica. Sono stati quindi progettati tre repertori di frasi interrogative derivanti dalle due tipologie di interrogative, dirette di tipo "x" o "*wh*-", e dirette alternative o polari (*yes/no* e disgiuntive).

[1] Per quanto si tratti di un lavoro collettivo, per fini specifici si intendono così assegnati i vari paragrafi: 1, 2, 7 Tini Brunozzi; 3, 4, 7 Panizza; 5, 6, 8 Zovato.

Interrogativa diretta tipo "x"	*wh-*	*Che giorno torni?*	-wh
Interrogativa diretta alternativa (o polare)	*yes/no*	*Ti serve qualcosa?*	y/n
	disgiuntiva	*Pubblico o privato?*	disgiuntiva

Tabella 1: Tipologie di frasi interrogative considerate nel progetto del *corpus*

Il progetto del *corpus* per l'etichettatura prosodica e la modellizzazione di frasi interrogative, nello specifico dei costrutti sintattici di tipo *wh-*, ha comportato la compilazione di una lista di frasi in cui la classe delle teste di sintagma interrogativo (chi, come, dove, quando, perché, ecc.) fosse combinata con la classe delle preposizioni (ad es., a > *a chi?*) allo scopo di ottenere un *set* di frasi interrogative definito (una lista "quasi" chiusa). Questo *set* di frasi è stato ulteriormente espanso in modo tale che la particella *wh-* non coincidesse esclusivamente con la coda interrogativa finale di sintagma e con la relativa risalita prosodica (cfr. es. 1 e 2 tabella 2), ma che fosse un nucleo prosodico con *focus* interrogativo cui poter giustapporre, a confine destro, materiale testuale (> acustico) di varia lunghezza (> durata) (cfr. es. 3 e 4 tabella 2).

1	*Chi?*
2	*A chi?*
3	*A chi* chiedere?
4	*A chi* chiedere informazioni sul traffico autostradale?

Tabella 2: Esempi di frasi interrogative di tipo *-wh*

Sono stati fatti, quindi, alcuni esperimenti di *collage* (cfr. tabella 5) congiungendo segnali interrogativi (la parte della testa *wh-*) con segnali dichiarativi (code deenfatiche, neutre) allo scopo di convalidare percettivamente il valore illocutivo della domanda così ottenuta.

A lato, va detto che il progetto delle frasi prevede che alla parte *wh-* sia collegata una parte dichiarativa neutra introdotta da plosiva sorda (cfr. es. 3 tabella 2) in modo da poter consentire al motore di sintesi una migliore concatenazione acustica tra fonemi, dato che i criteri di taglio delle unità acustiche incoraggia i punti di minore sonorità, come lo è di fatto l'occlusione prima della plosione di una occlusiva sorda (/p/, /t/, /k/).

La lista di frasi interrogative viene fatta leggere da uno *speaker* professionista e viene registrata in due modalità, sia con intonazione interrogativa, sia con intonazione dichiarativa, per predisporre due basi dati acustiche che condividono la stessa base dati testuale, allo scopo di poterle predisporre in parallelo all'analisi contrastiva per valutarne i modelli prosodici corrispondenti (Firenzuoli, 2001) (cfr. figura 1).

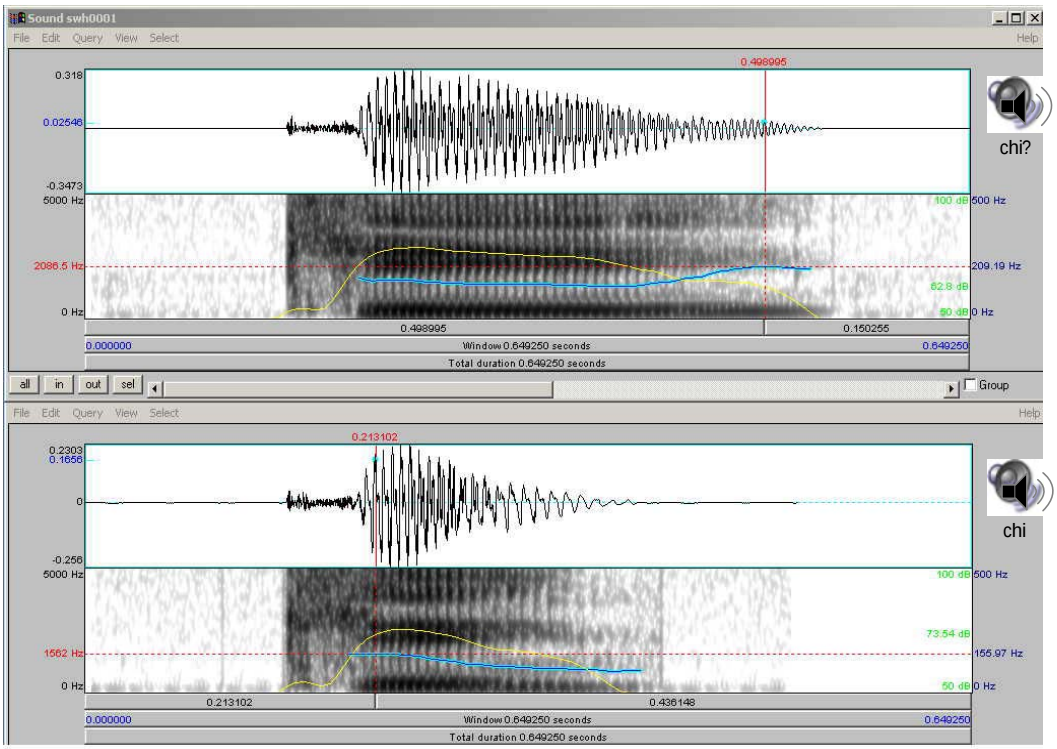


Figura 1: Esempio di frase letta in modalità interrogativa ("*chi?*", sopra)
e dichiarativa ("*chi.*", sotto)

3. IL PROGETTO DEL *TOOL* "SYBILLA" PER L'ANNOTAZIONE PROSODICA

Al fine di poter visualizzare e annotare l'andamento di f_0 delle frasi del *corpus* predisposto per questo lavoro, è stata progettata un'interfaccia grafica di annotazione.

A partire dal *software* Xwaves Entropic, normalmente utilizzato per il controllo dell'allineamento fonetico nell'ambito delle attività del laboratorio di Sintesi Vocale di Loquendo S.p.A., si è dunque scelto di progettare un *tool*, denominato "Sybilla"[2], in grado di visualizzare le curve di f_0 e di energia di un enunciato per eseguire annotazioni e rilevamenti utili alla modellizzazione di *pattern* intonativi.

L'unità di base che è stata scelta al fine delle nostre osservazioni, e che meglio si presta ad essere analizzata a livello prosodico, è la sillaba (intesa come sillaba prosodica), poiché è

[2] Il nome Sybilla è stato scelto sia per consonanza con 'sillaba', che è l'unità di analisi del *tool*, sia per l'idea che lo strumento possa fornire risposte utili all'astrazione di modelli.

l’elemento che porta informazioni fonetiche e intonative non solo a livello di parola ma anche di enunciato.

La sillaba è dunque considerata come unità minima dotata di nucleo vocalico che prescinde dal confine di parola nel *continuum* dell’enunciato.

Le sillabe vengono estratte automaticamente, mediante utilizzo della scala di sonorità, individuando i confini sillabici in prossimità di minimi relativi in termini di sonorità, come nell’esempio seguente (cfr. tabella 3):

State scherzando
s t `a - t e s - k e r - ts `a n - d o

Tabella 3: Esempio di scomposizione in unità sillabiche

Il *software* Sybilla, realizzato ai nostri scopi specifici, si presenta all’utilizzatore con quattro visualizzazioni principali (cfr. figura 2):

- la rappresentazione in forma d’onda del segnale vocale analizzato;
- la rappresentazione spettrografica del segnale vocale analizzato;
- la visualizzazione dell’andamento della frequenza f_0 e la rappresentazione della curva di energia;
- una serie di sottocampi, ciascuno in grado di fornire indicazione di alcune etichette e parametri acustici.

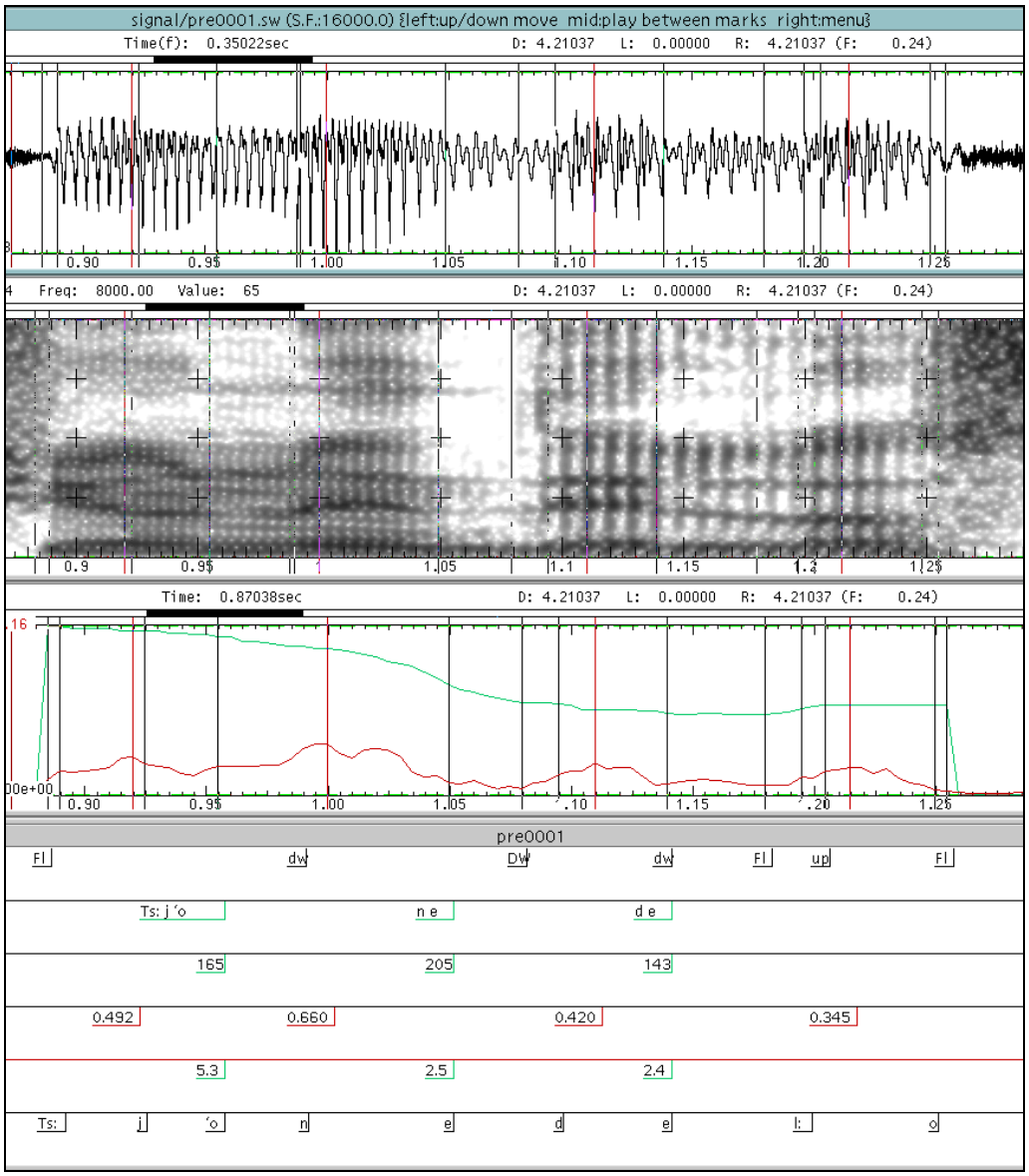


Figura 2: Esempio di schermata fornita dallo strumento “Sybilla”

4. I LIVELLI DI ANNOTAZIONE E LE ETICHETTE MORFOLOGICHE DI SYBILLA

I livelli di annotazione di Sybilla sono complessivamente sei:

1) Andamento di f_0:

per la modellizzazione manuale dell'andamento della curva di frequenza f_0 all'interno di ciascuna sillaba, mediante una serie di etichette che indicano le caratteristiche morfologiche della curva (ascendente, discendente, piatta). L'etichettatura degli andamenti comprende i seguenti casi:

- crescita moderata: up
- crescita rapida: UP
- decrescita moderata: dw
- decrescita rapida: DW
- andamento piatto: Fl
- tratto non vocalizzato: Nv
- silenzio: Sl

2) Unità sillabiche:

con la rappresentazione lineare della segmentazione automatica dei confini dei gruppi vocalici a partire dall'allineamento fonematico;

3) Valore di f_0:

con la rappresentazione lineare dell'allineamento automatico al confine destro di sillaba, che indica il valore medio in Hz della frequenza f_0 nella sillaba in oggetto;

4) Energia normalizzata:

per l'annotazione automatica dei punti di picco energetico per ciascuna sillaba, e l'indicazione automatica del relativo valore normalizzato, rispetto al punto di energia massima dell'intera frase, posto =1 per convenzione;

5) Durata percentuale:

contenente l'indicazione automatica della durata percentuale di ciascuna sillaba, considerando la durata complessiva della frase =100;

6) Segmentazione fonetica:

estrazione automatica dei simboli fonetici dal *database* testuale con possibilità di controllo manuale dell'allineamento al segnale. La modifica manuale dei confini dei fonemi, comporta un conseguente riaggiustamento automatico di tutti i campi precedentemente descritti.

L'annotazione condotta per il presente lavoro è stata compiuta inserendo l'etichetta morfologica laddove viene rilevato un cambio di pendenza nella visualizzazione della curva di f_0. Tale criterio di etichettatura manuale sulla curva di f_0, ha suggerito l'utilizzo 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, per garantire la maggior uniformità possibile tra i diversi annotatori.

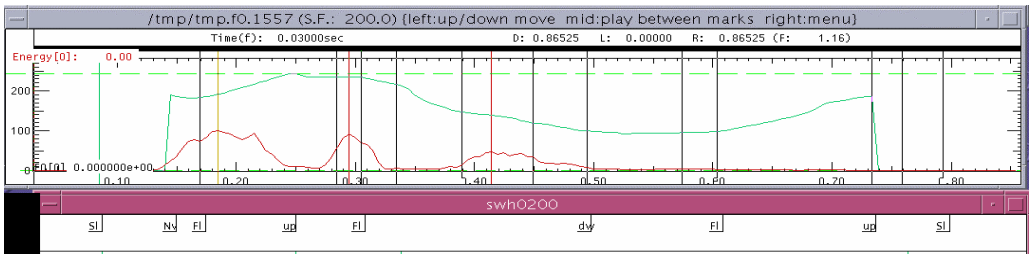


Figura 3: Visualizzazione della curva di f_0 e di energia tramite lo strumento "Sybilla"

Laddove la sola percezione visiva non si rivela efficace a discriminare il tipo di andamento da annotare, è stata applicata una formula per calcolare una soglia di valori utile a disambiguare il lavoro di annotazione:
Dato:

$$R = 100 \cdot \frac{|F0\_fine - F0\_inizio|}{(t\_fine - t\_inizio)} \qquad (1)$$

- per R<1,5 → andamento Fl
- per 1,5<R<20 → andamento up/dw
- per R>20 → andamento UP/DW

I motivi che hanno suggerito l'introduzione di questo formalismo per l'annotazione dei femomeni intonativi vanno ricondotti all'esigenza di fornire una stilizzazione dei movimenti intonativi a livello soprasegmentale sufficientemente precisa ma semplice. In tal modo si dovrebbe favorire la modellizzazione dei fenomeni prosodici presi in considerazione, per selezionare in modo efficace le unita' acustiche del sistema di sintesi vocale precedentemente citato.

Ovviamente non si ha in tal caso la completezza di annotazione di un sistema multi-livello come ad esempio ToBI (Avesani *et al.*, 2004) in cui vengono descritti in modo esaustivo diversi fenomeni prosodici, prevedendo infatti l'annotazione dei toni relavi alla sillaba, al sintagma e al confine di sintagma ed una specifica annotazione delle giunture delle unita' tonali. Un formalismo di questo tipo comporterebbe, per i nostri scopi, una complessa gestione dei dati e richiederebbe molta cura da parte degli annotatori al fine di evitare interpretazioni soggettive.

Il sistema qui proposto risulta essere piu' simile a INTSINT (Hirst, 1994) che prevede l'annotazione di tutti i punti fonologicamente significativi della curva di f_0 ed assegna in modo coerente una etichettatura di tono assoluto o relativo, dove nel primo caso si fa riferimento al *range* dei valori di f_0 dello *speaker*, mentre nel secondo si tiene conto della variazione rispetto ai punti significativi adiacenti. Inoltre nella trascrizione possono essere conservati i valori assoluti di f_0 in corrispondenza dei punti significativi. Nel nostro caso, invece, si mantiene traccia del valore medio di f_0 a livello di sillaba.

5. ANNOTAZIONE PROSODICA

In questo lavoro, l'annotazione prosodica prevede una fase di analisi automatica seguita da una fase manuale che consiste nel controllo delle annotazioni automatiche e nell'integrazione di ulteriori etichette. In modo automatico vengono calcolati i confini fonetici e sillabici e i valori di f_0 e di energia. Le operazioni manuali consistono nel controllo delle etichette fonetiche e sillabiche ma soprattutto nell'inserimento delle annotazioni morfologiche che descrivono l'andamento della frequenza fondamentale. In pratica si opera una forma di astrazione in base alla quale la curva di f_0 viene vista come una sequenza di linee spezzate a ciascuna delle quali viene assegnata una etichetta morfologica che descrive la pendenza delle stesse.

5.1 Analisi acustica automatica

L'analisi acustica è piuttosto articolata e prevede l'utilizzo di diversi moduli. Innanzi tutto viene usato un trascrittore fonetico per convertire i testi in sequenze di simboli fonetici. Questo strumento si basa su regole di trascrizione e su lessici per le eccezioni di pronuncia. Successivamente si procede con l'allineamento dei fonemi alla forma d'onda. A tal scopo viene utilizzato un allineatore automatico basato su modelli di Markov. I modelli sono stati precedentemente addestrati con materiale vocale appartenente allo stesso *speaker* oggetto di indagine. I segnali sono stati acquisiti in studio in forma digitale con frequenza di campionamento di 44.1 kHz, anche se per l'analisi acustica si è fatto uso della versione sottocampionata a 16 kHz. Per ogni fonema sono si fa riscorso a modelli a 3 o 5 stati, e per quanto riguarda i parametri acustici sono stati considerati 8 coefficienti mel-cepstrali e un coefficiente di energia, insieme alle derivate prime e seconde. Ottenuta la segmentazione fonetica, si determinano i confini di sillaba. Questa operazione viene effettuata a partire dai confini fonetici mediante utilizzo della scala di sonorità, individuando i confini sillabici in prossimità di minimi relativi in termini di sonorità (Cutugno *et al*., 2002).

La seconda fase consiste nell'estrapolazione dei valori della frequenza fondamentale e di energia. La frequenza fondamentale viene calcolata mediante la procedura *get_ f_0* di ESPS facendo ricorso a finestre di analisi il cui passo di avanzamento è di 5 millisecondi (80 campioni nel caso di frequenza di campionamento 16000 Hz). L'energia viene calcolata su finestre di analisi traslate della stessa quantità utilizzata per il calcolo di f_0.

A partire da questi dati per ciascuna unità sillabica è possibile calcolare i seguenti valori:

- i valori medi di f_0
- i valori di picco di energia (normalizzati rispetto al picco di frase)
- le durate percentuali, escludendo le pause

Tutte queste informazioni vengono fornite all'interfaccia grafica che viene usata per il controllo e l'annotazione manuale.

5.2 Annotazione manuale

Manualmente è possibile correggere i confini fonetici ed eventualmente modificare e/o inserire etichette fonetiche. Inoltre vengono inserite le etichette morfologiche che descrivono l'andamento della frequenza fondamentale. Al termine della fase manuale vengono ricalcolati i parametri acustici, in quanto eventuali correzioni dei confini sillabici renderebbero non più

corretti i dati numerici. Pertanto, a ogni successiva analisi, i confini delle unità sillabiche vengono riallineati ai confini fonetici mentre le etichette prosodiche sono svincolate da qualsiasi tipo di segmentazione, e i parametri di f_0, picchi di energia e durate percentuali vengono ricalcolati, tenendo conto della nuova situazione segmentale mentre, in questa fase, le etichette morfologiche e il loro posizionamento non subiscono alcuna modifica.

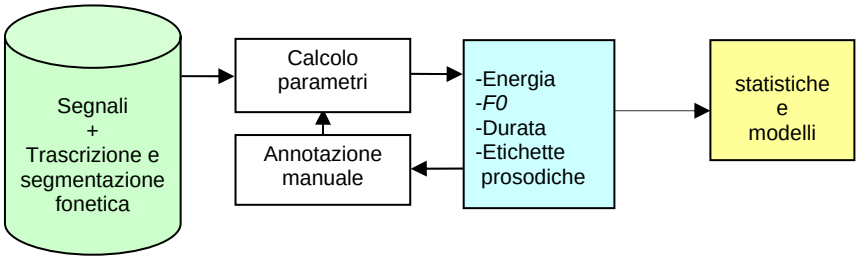


Figura 4: Schema del processo di analisi e di elaborazione dei dati

6. ELABORAZIONE DEI DATI

Tutte le annotazioni manuali e i dati acustici calcolati automaticamente sono stati elaborati in modo da poter identificare, dove possibile, delle regolarità nelle tipologie di frase considerate. L'obiettivo principale è in particolare quello di ottenere un modello qualitativo degli andamenti di f_0 e verificarne la coerenza con esperimenti di manipolazione del segnale.

Le sillabe sono state classificate in base alla loro posizione e distinte in toniche e atone, intendendo nel primo caso sillabe in cui fosse presente un accento lessicale.

Al fine di facilitare il confronto, si è fatto ricorso a unità di tempo normalizzato, corrispondenti ad un decimo dell'intera durata della parte vocalizzata della sillaba. In tal modo tutte le sillabe hanno durata "1" e per ogni unità di tempo è stata associata una etichetta morfologica in base alle annotazioni manuali. In definitiva ogni sillaba è descritta da 10 etichette morfologiche. Se ad esempio per una sillaba il profilo intonativo prevedeva un andamento "dw" per il 50% della durata seguito da un andamento "up" per la restante metà, allora in termini di unità di tempo normalizzato si avrebbero 5 intervalli "dw" seguiti da 5 intervalli "up".

Per quanto riguarda la distribuzione delle etichette, bisognerebbe distinguere le stesse a livello di sillaba in termini di posizione all'interno della sillaba, presenza o meno di accento lessicale, contesto ovvero tipo di sillaba che precede e segue, ecc. In questa fase iniziale del nostro lavoro il materiale annotato non ha completa rappresentatività in questo senso, tuttavia le analisi in termini di distribuzioni ci hanno permesso di identificare delle significative regolarità nella parte iniziale e finale delle frasi interrogative, a prescindere dalla durata e struttura sillabica delle frasi interrogative analizzate. Queste regolarità sono peraltro avvallate anche dagli esperimenti di risintesi. In particolare sono stati considerati quattro casi che riguardano la parte iniziale e terminale delle frasi interrogative:

- sillaba atona che precede la prima tonica

- prima sillaba tonica
- ultima sillaba tonica
- sillaba atona che segue l'ultima tonica

Nei grafici della figura che segue (fig. 5) vengono visualizzate le distribuzioni delle etichette morfologiche inserite manualmente nei quattro casi sopra specificati distingiendo tra frasi interrogative *wh-* e *yes/no*. Nelle frasi *wh-* è possibile notare che la prima sillaba tonica è caratterizzata da un andamento che è principalmente discendente. In pratica questa risulta essere una regolarità nelle frasi interrogative di questo tipo, cosa non altrettanto vera nel caso di sillabe atone che precedono la prima tonica in cui, come si nota, gli andamenti sono più uniformemente distribuiti. Anche la parte terminale di queste frasi presenta degli andamenti comuni, infatti l'ultima sillaba tonica è caratterizzata da profili discendenti o piatti ed è per lo più seguita da sillabe atone con profili decisamente crescenti.

Anche nel caso di frasi interrogative di tpo *yes/no* si hanno dei riscontri interessanti in queste quattro tipologie di sillabe. Per quanto riguarda la parte iniziale non si ha un andamento dominante, anche nella prima tonica si ha una leggera prevalenza di profili ascendenti, ma non così marcata come nel caso di profili discendenti della prima sillaba tonica delle frasi di tipo *wh-*. Molto più significativa è la parte finale in cui si ha una assai netta prevalenza di andamenti discendenti nell'ultima tonica seguita da profili ascendendi nelle ultime sillabe atone. Le interrogative *yes/no* sembrano quindi essere caratterizzate, dal punto di vista intonativo, soprattutto nella parte finale.

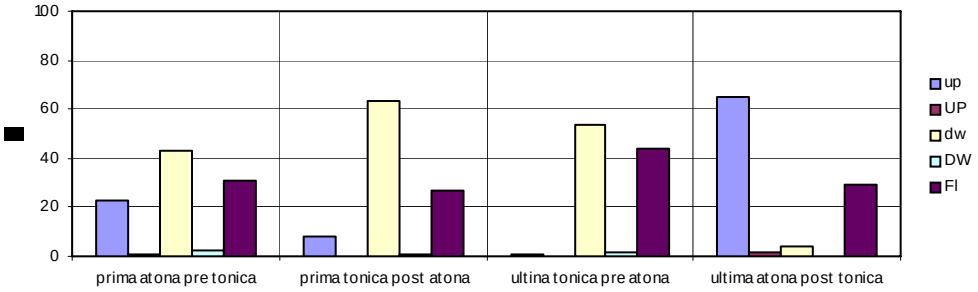


Figura 5: Distribuzioni delle etichette morfologiche in quattro tipologie di sillabe in frasi interrogative di tipo *wh-*

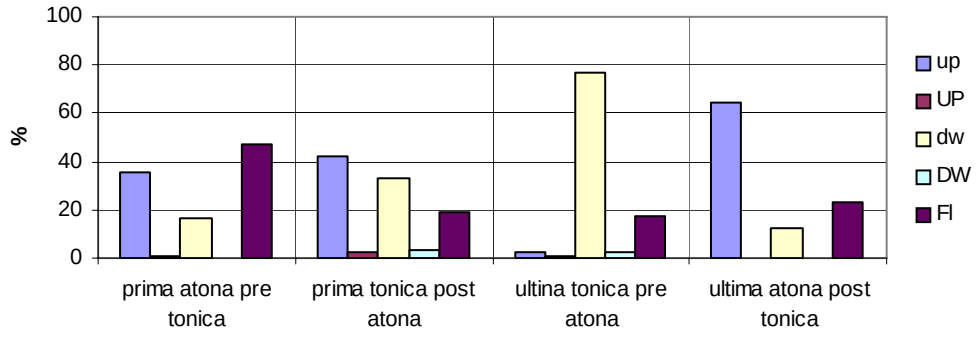


Figura 6: Distribuzioni delle etichette morfologiche in quattro tipologie di sillabe in frasi interrogative di tipo *yes/no*

Sono altrettanto interessanti le indicazioni numeriche fornite dai parametri acustici calcolati in modo automatico. Nella tabella che segue (tab. 4) vengono riportate le variazioni medie di f_0, del picco di energia e della durata, delle quattro tipologie di sillaba prese in considerazione. Le variazioni sono calcolate rispetto ai valori medi, a livello di enunciato, ovvero rispetto alla media su tutte le sillabe dell'enunciato dei parametri interessati.

		atona *pre* prima tonica	prima tonica *post* atona	ultina tonica *pre* atona	atona *post* ultima tonica
Variazione f_0	**wh**	1,21	1,03	0,69	0,90
	yes/no	1,23	1,22	0,72	0,95
Variazione picco	**wh**	1,30	2,41	0,44	0,17
	yes/no	1,32	1,48	1,18	0,46
Variazione durata	**wh**	0,74	1,28	1,29	0,96
	yes/no	0,68	0,96	1,32	1,07

Tabella 4: Variazioni dei parametri acustici in quattro tipologie di sillabe in frasi interrogative di tipo *yes/no* e *wh-*

Per quanto concerne la frequenza fondamentale, non si hanno grosse differenze tra frasi di tipo *wh-* e frasi di tipo *yes/no*. Entrambe presentano valori superiori alla media nella parte iniziale della frase, che decrescono andando verso la parte terminale in cui anche numericamente si riscontra l'andamento decrescente-crescente. Analizzando i picchi di energia, emerge che nelle frasi *wh-* si ha una variazione positiva significativa nella prima sillaba tonica e un netto calo nella parte terminale della frase, mentre nelle frasi *yes/no* sembra essere marcata anche l'ultima tonica. Per quanto riguarda le durate, le sillabe toniche presentano sostanzialmente degli aumenti rispetto ai valori medi, mentre le atone iniziali sono più brevi e quelle finali rientrano nei valori medi.

Nelle figure che seguono vengono riportate le distribuzioni di etichette in funzione del tempo normalizzato. In particolare si fa riferimento alla parte terminale delle frasi interrogative nel caso in cui le ultime due sillabe siano nell'ordine tonica e atona, caso piuttosto frequente nel campione di frasi analizzate. Ciascuna figura riporta 20 unità di tempo di cui dieci relative alla penultima sillaba e le seguenti 10 all'ultima. In entrambe i casi è apprezzabile una prevalenza di etichette "dw" nella penultima sillaba e "up" nell'ultima. L'unica differenza risiede nel fatto che mentre nel caso di frasi interrogative *wh-* la transizione dall'andamento decrescente a quello crescente sembra graduale, testimoniata da una significativa percentuale di andamenti "Fl" in corrispondenza del decimo *frame*, nelle interrogative *yes/no* questa transizione risulta essere meno graduale. Infatti la concentrazione di etichette "Fl" si verifica per un intervallo di tempo minore. È da notare inoltre come nelle interrogative *yes/no* le concentrazioni di andamenti "dw" nella penultima sillaba e "up" nell'ultima atona sono più alte rispetto alle interrogative *wh*.

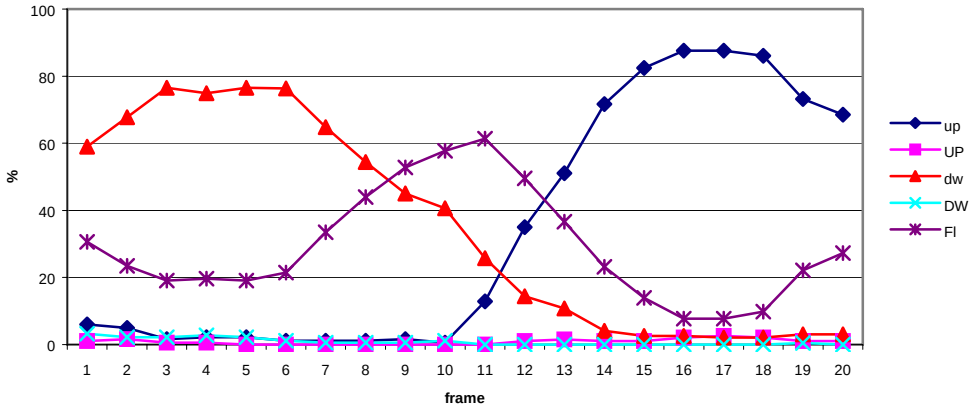


Figura 7: Distribuzioni delle etichette morfologiche nel caso di penultima sillaba tonica e ultima sillaba atona in frasi interrogative di tipo *wh-*

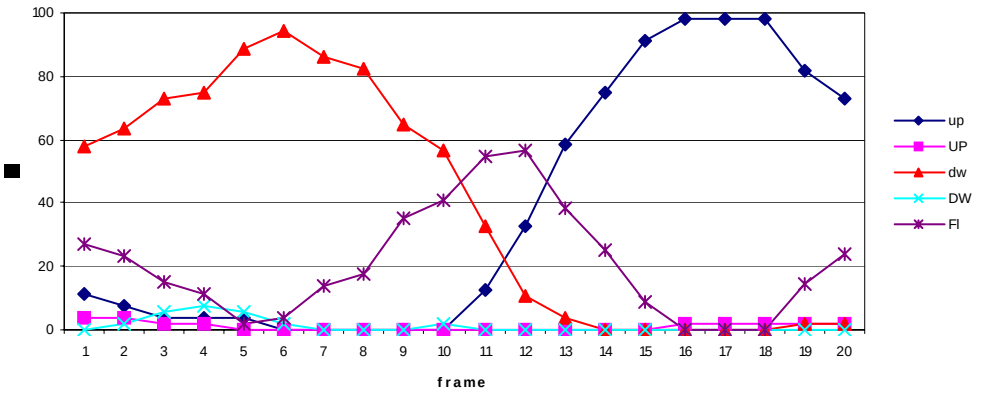


Figura 8: Distribuzioni delle etichette morfologiche nel caso di penultima sillaba tonica e ultima sillaba atona in frasi interrogative di tipo *yes/no*

Nello specifico di questo lavoro, i risultati riportati si riferiscono a casi particolari relativi all'analisi del materiale registrato da un solo *speaker*, tuttavia come si è visto forniscono delle indicazioni interessanti, in accordo con la pratica sperimentale. Gli sviluppi dovranno quindi consistere nell'ampliamento del *corpus*, in modo da raggiungere una copertura statistica più rappresentativa, e l'acquisizione di materiale da più locutori. Di conseguenza, con materiale di maggiori dimensioni si dovranno usare metodi di classificazione e analisi più sofisticati come ad esempio CART (*Classification and Regression Tree*) che permettono di ottenere delle modelizzazioni utili in vista di un impiego di questa simbologia in un sistema di sintesi vocale da testo scritto, in cui l'inventario delle unità acustiche possa essere usato proficuamente nella selezione del profilo intonativo più adeguato al contesto e alla tipologia della frase da sintetizzare (Cosi *et al*., 2003).

7. ESPERIMENTI

A partire dalla analisi condotta in modo sistematico delle frasi interrogative mediante lo strumento Sybilla, sono stati condotti degli esperimenti di sintesi al fine di verificare la validità delle modellizzazioni riguardanti gli andamenti intonativi tramite strumenti empirici:

- Esperimenti di sintesi mediante "*collage*"
- Esperimenti di *prosody transplantation*
- Esperimenti di risintesi

Definiamo *collage* la possibilità di ottenere un segnale di frase interrogativa di tipo *wh*-, con la semplice giunzione della testa interrogativa con una coda, selezionata da frasi dichiarative-neutre con andamento intonativo sospensivo e in certi casi anche conclusivo.

db Voce	testa wh-	coda dichiarativa neutra		audio
Matteo	*A che ora*	Sosp.	*tornare al menù iniziale*	Matteo 1
	Per quanto tempo	Concl.	*non è stato riconosciuto*	Matteo 2
Luca	*Chi vuole*	Sosp.	*dopo il segnale acustico*	Luca 1
	Perché	Concl.	*Trentino Alto Adige*	Luca 2
Paola	*A che ora*	Sosp.	*desidera parlare con un operatore*	Paola 1
	Chi ha	Concl.	*testi scritti in alcune lingue straniere*	Paola 2

Tabella 5: Esperimenti di "*collage*"

Per quanto riguarda gli esperimenti di *prosody transplantation* l'annotazione del materiale interrogativo con Sybilla, e la successiva analisi compiuta su di esso, hanno suggerito la possibilità di ottenere segnali acustici di domande direttamente con la manipolazione dei parametri prosodici delle corrispondenti frasi dichiarative.

Tramite il programma Praat (Boersma & Weenink, 2005) viene quindi imposto manualmente a una frase dichiarativa l'andamento di f_0 dell'identica frase interrogativa, annotato con Sybilla.

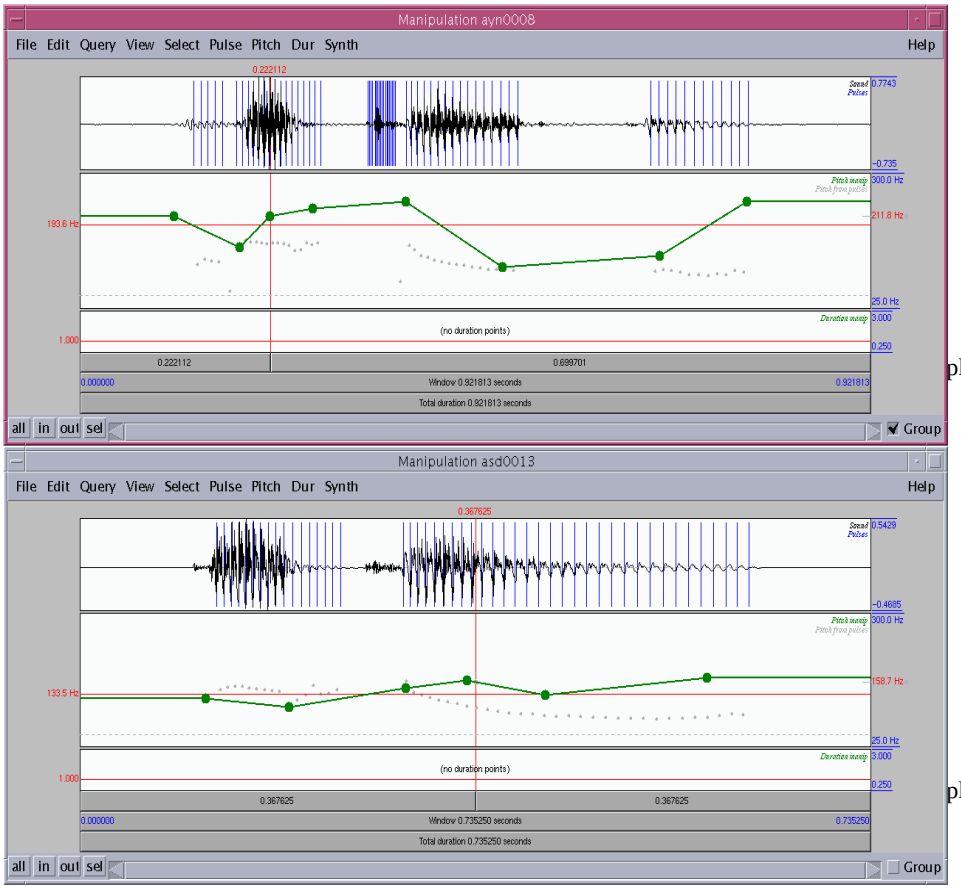
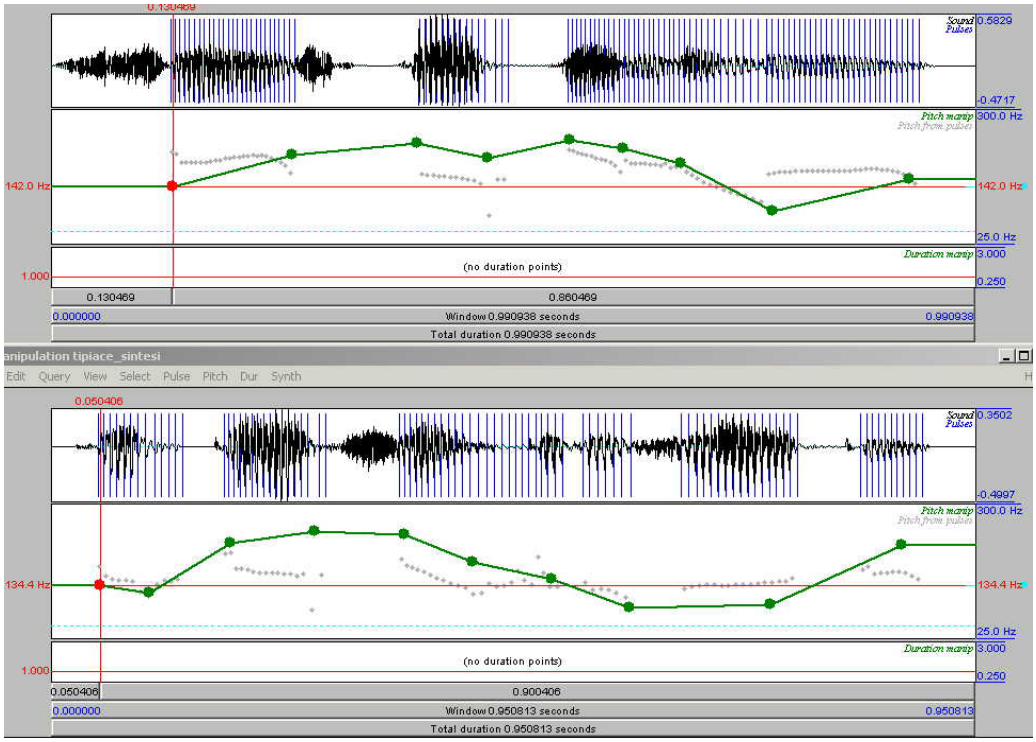


Figura 9: Esperimenti *di prosody transplantation*

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 dell'accento primario.

In questa fase del lavoro, quindi, non si opera più su materiale preregistrato e non si utilizzano modelli interrogativi "copia" ma si cerca, invece, di rendere applicabile la tecnica di manipolazione di f_0 con modelli ottenuti tramite comparazione di annotazioni.



Matteo 1 Matteo 2

Figura 10: Esperimenti di risintesi

8. CONCLUSIONI

L'obiettivo dell'annotazione prosodica compiuta con Sybilla, unito agli esperimenti di *prosody transplantation* e risintesi è quello di verificare la possibilità di astrarre un modello

generale per la prosodia delle frasi interrogative e di riuscire ad imporlo alle frasi prodotte in sintesi.

Allo stesso modo, l'obbiettivo degli esperimenti di *collage* e la successiva modifica del modulo di selezione di unità acustiche del TTS, è quello di verificare che le frasi interrogative di tipo *wh-* possano essere sintetizzate a partire da un progetto mirato dei testi e con manipolazioni prosodiche limitate alle sole code di sintagma.

Il *focus* interrogativo sulla testa del sintagma unito sia alla riduzione del *range* di frequenza della frase nella sua parte centrale sia all'elevazione del *pitch* nella parte terminale, consentono la percezione dell'atto di domanda.

9. BIBLIOGRAFIA

Avesani, C.; Vayra, M., 2000. Costruzioni marcate e non-marcate in italiano: il ruolo dell'intonazione. *X Giornate di Studio del Gruppo di Fonetica Sperimentale*, Napoli, 1-15.

Avesani, C.; Cosi, P.; Fauri, E.; Gretter, R.; Mana, N.; Rocchi, S.; Rossi, F.; Tesser F., 2004. Definizione ed annotazione prosodica di un database di parlato-letto usando il formalismo ToBI. In F. Albano Leoni, F. Cutugno, M. Pettorino, R. Savy (a c. d.) *Il parlato italiano. Atti del Convegno Nazionale*, Napoli: M. D'Auria Editore, CD- ROM M01.

Balestri, M.; Pacchiotti, A.; Quazza, S.; Salza, P.; Sandri S., 1999. Choose the Best to Modify the Least: a New Generation Concatenative Synthesis System. In *Proceedings of Eurospeech 1999*, Budapest, 2291-2294.

Boersma, P.; Weenink, D., 2005. *Praat: doing phonetics by computer*, http://www.praat.org/

Cosi, P.; Avesani, C.; Tesser, F.; Gretter, R.; Pianesi, F., 2003. On the Use of Cart-Tree for Prosodic Predictions in the Italian Festival TTS. In P. Cosi, E. Magno Caldognetto, A. Zamboni (a c. d.) *Voce, Canto, Parlato – Studi in onore di Franco Ferrero*, Padova: Unipress, 73-81.

Cutugno, F.; D'Anna, L.; Petrillo, M.; Zovato, E., 2002. APA: Towards an automatic *tool* for prosodic analysis. *Speech Prosody 2002, International Conference*, Aix-en-Provence, France, April 11-13, 2002.

Firenzuoli, V., 2001. Verso un nuovo approccio allo studio dell'intonazione a partire da corpora di parlato: esempi di profili intonativi di valore illocutivo dell'italiano. In N. Maraschio, T. Poggi Salani (a c. d.) *Atti del XXXIV Congresso internazionale degli Studi della Società di Linguistica Italiana*, Firenze, 19-21 ottobre 2000, Roma: Bulzoni.

Hirst, D. J., 1994. The Symbolic Coding of Fundamental Frequency Curves: from Acoustics to Phonology. In *Proceedings of the International Symposium on Prosody*, Yokohama, Japan.