PROMINENZA FRASALE E TIPOLOGIA PROSODICA: UN APPROCCIO ACUSTICO

Fabio Tamburini (Bologna)

1. INTRODUZIONE

Ad un esame degli studi sulla prominenza prosodica si riscontrano notevoli problemi di uniformità terminologica, soprattutto a livello dell'identificazione dei fenomeni che fanno parte dell'ambito di indagine. Numerosi studiosi hanno sottolineato più volte come nelle analisi dei fenomeni prosodici si trovi una rilevante eterogeneità tra i termini utilizzati per indicare lo stesso fenomeno, che, da studio a studio e nel tempo, tendono ad assumere differenti connotazioni e riferimenti (Bertinetto, 1981; Jensen, 2004; Spencer, 1996; Taylor, 1992; Wightman, Ostendorf, 1994).

Sembra quindi opportuno, prima di affrontare la descrizione vera e propria del lavoro svolto, definire chiaramente il fenomeno oggetto di questo studio, la prominenza frasale nella lingua parlata, identificandone le proprietà e i tratti fondamentali che contribuiscono ad una sua definizione.

Una delle più note e citate definizioni di prominenza di deve a Terken (1991: 1768):

> *Prominence is the property by which linguistic units are perceived as standing out from their environment*.

Da essa ricaviamo un primo tratteggio del fenomeno in oggetto che risulta essere un fenomeno percettivo in grado di consentire a determinate unità linguistiche di "emergere" rispetto al contesto che le circonda.

D'altra parte Jensen (2004: 27) nella sua definizione

> *The term Prominence […] generally refers to the degree to which something stands out from its surroundings. It may be used about specific properties, such as pitch prominence, […], or more generally, as perceived prominence, about the overall degree of emphasis (or de-emphasis) of a certain item*.

sottolinea anch'esso come la prominenza possa essere vista come un fenomeno percettivo in grado di enfatizzare alcune unità rispetto al contesto nel quale sono inserite.

Mertens (1991: 218) fornisce una descrizione maggiormente dettagliata

> *A syllable is prominent when it stands out from its context due to a local difference for some prosodic parameter. Prominence is continuous (not categorical) and contributions of multiple parameters interact*.

in cui si sottolinea come la prominenza sia intrinsecamente continua nella sua espressione e non categoriale (si veda anche Ladd *et al*. 1994) e come molti parametri prosodici contribuiscano nel supportarla.

Nelle due definizioni seguenti notiamo come i problemi terminologici accennati precedentemente contribuiscano a complicare notevolmente l'analisi degli studi nel settore:

> *What many phoneticians and linguists have called stress, and what most laymen readily understand under this term, refers to nothing more than the fact that in a succession of spoken syllables or words some will be perceived as more salient or prominent than others*. (Couper-Kuhlen, 1986: 19)

> *The term sentence-accent refers to the perceptual salience of some words over others in utterances, […]* (Kohler, 2006: 749).

Kohler si riferisce sostanzialmente al concetto di prominenza denotandolo come *sentence-accent*, mentre Couper-Kuhlen nota come tradizionalmente il termine *stress* sia utilizzato con la stessa accezione di prominenza in numerosi studi, spesso senza fornire un riferimento preciso a *stress* lessicale o *stress* frasale, due concetti sufficientemente distinti, anche se correlati tra loro.

Uniformare la notazione, pur mantenendo un legame forte con i riferimenti bibliografici in lingua inglese, per rendere chiara l'esposizione in lingua italiana ha comportato un certo lavoro: la scelta è stata quella di limitare al minimo indispensabile l'uso della terminologia tratta dall'ambito prosodico, mantenendo per alcuni di tali termini la notazione anglosassone ed evitando in ogni caso l'introduzione di sinonimi o equivalenti traduttivi nel corso dell'esposizione. L'obiettivo globale del lavoro compiuto va nella direzione di una presentazione notazionalmente compatta e uniforme all'interno del contributo, che si concede come deroga l'uso di alcuni termini in lingua inglese sempre evidenziati tipograficamente.

Riassumendo quindi le proprietà del fenomeno in esame potremmo dire che la prominenza è un fenomeno percettivo, di natura continua, che consente di enfatizzare alcune unità linguistiche di tipo segmentale rispetto al contesto che le circonda, ed è supportata da una complessa interazione di parametri di tipo prosodico e fonetico/acustico.

Al fine di poter costruire un modello computazionale del fenomeno in esame è necessario descrivere in modo preciso l'interazione dei parametri in grado di indurre la percezione della prominenza. Dei numerosissimi lavori in questo settore mi sembra opportuno fare riferimento primariamente al lavoro di uno studioso tedesco, Klaus J. Kohler, per la chiarezza, la lucidità e il rigore metodologico con cui descrive i fenomeni coinvolti:

> *The category of sentence accent* (prominenza) *is a separate prosodic level from intonation, controllable independently from rhythm, syllabic and segmental structuring, on a scale from 1 to 3. Although it shares F0 as a physical property with intonation, it is not entirely determined by it, but also depends on syllable and segment duration, intensity, and possibly other features.* (Kohler, 2003: 2930)

> *…, it became clear that beside the accent category that is principally signalled by F0 excursion and may therefore be called pitch accent, another type of accent has to be recognised that is primarily related to non-pitch features, viz. acoustic energy, based on phonatory and articulatory force, and may therefore be called force accent*. (Kohler, 2005: 99)

Dai lavori di Kohler emergono chiaramente due attori precisi, a livello linguistico-prosodico, in grado di supportare il fenomeno della prominenza frasale (o *sentence accent*): i *pitch accent* e i *force accent*. Il primo (*pitch accent*) risulta coincidere pressoché totalmente col concetto omonimo introdotto da Bolinger (1958) ed essenzialmente legato alle variazioni nel profilo della frequenza fondamentale (F0), mentre il secondo (*force accent*) risulta essere un fenomeno completamente indipendente dalla componente intonativa degli enunciati e intimamente legato a fenomeni acustici di altro tipo, per esempio l'intensità e la durata delle unità segmentali.

I due fenomeni sembrano giocare entrambi un ruolo preminente nel supportare la prominenza percepita a livello di enunciato, in linea con ciò che sostengono alcuni studiosi (si veda ad esempio il lavoro di Ladd, 1996), ma anziché in un ottica antagonistica o gerarchica in un'ottica di interazione e rinforzo reciproco.

In questo quadro è possibile tentare una prima, parziale, formalizzazione del fenomeno in esame considerando il diagramma in figura 1 nel quale le due tipologie accentuali suggerite da

Kohler contribuiscono a supportare il fenomeno della prominenza frasale, e che può essere descritta matematicamente con l'equazione

$$\text{Prom}^i = FA^i + PA^i ,$$

dove *FA* e *PA* sono rispettivamente i contributi dei *force accent* e dei *pitch accent* riferiti all'*i*-esima unità segmentale dell'enunciato.

A livello acustico, numerosi studi (Sluijter, van Heuven, 1996; 1997; Anastakos et al. 1995; Bagshaw, 1994; Heldner, 2003; Streefkerk, 1996) suggeriscono, anche in una prospettiva interlinguistica, una dipendenza tra i *force accent* e parametri come la durata e l'enfasi spettrale (*spectral emphasis*, *spectral tilt* o *spectral balance*), mentre i *pitch accent* sarebbero supportati prevalentemente da movimenti nel profilo di F0 e dall'intensità globale all'interno dell'unità segmentale di riferimento. Lo stesso autore ha condotto alcuni esperimenti che hanno suffragato l'esistenza di tali relazioni in riferimento ad alcune lingue (Tamburini, 2003; 2005; 2006).

Nel paragrafo 2 di questo contributo si descriverà brevemente un lavoro sviluppato negli ultimi anni, già presentato in varie sedi (Tamburini 2003; 2005; 2006; Tamburini, Caini, 2005), volto alla costruzione di un sistema per l'identificazione automatica della prominenza frasale; tale studio è stato utilizzato come base per sviluppare un lavoro più ampio del quale presenterò alcuni risultati preliminari ottenuti più recentemente (§3).

2. IDENTIFICAZIONE AUTOMATICA DELLA PROMINENZA FRASALE

Visto l'obiettivo generale dello studio, riteniamo opportuno che tale indagine utilizzi come uniche informazioni i parametri derivabili direttamente, anche se in modo estremamente articolato, dall'espressione sonora dell'enunciato, ovvero dalla sua trasposizione digitale realizzata campionando opportunamente il segnale. Il modello che proporremo e l'algoritmo che implementerà il riconoscimento della prominenza non saranno basati su fonti di informazioni alternative, quali trascrizioni degli enunciati, sia ortografiche che fonetiche, risorse linguistiche etichettate dal punto di vista fonetico, fonologico, prosodico, e nemmeno risorse che contengano informazioni di tipo segmentale sugli enunciati. L'unica informazione fornita all'algoritmo di annotazione sarà la digitalizzazione dell'enunciato (oscillogramma o *waveform*).

Restringere il dominio delle informazioni fruibili in modo così netto elimina, di fatto, tutti quei modelli di analisi parametrica che utilizzano fasi di apprendimento basate su dati autentici: *hidden Markov model* (HMM), reti neurali, alberi di decisione probabilistici, classificatori basati su logiche *fuzzy*, ecc., richiedendo pesanti fasi di apprendimento, che sfruttano appieno le informazioni di vari tipi di risorse linguistiche etichettate, sono modelli teorici esclusi a priori da questo lavoro, nonostante la potenza che hanno dimostrato nelle infinite applicazioni nelle quali sono stati utilizzati.

Il riferimento a risorse linguistiche accuratamente etichettate in modo manuale per poter sviluppare tali modelli ci è sembrato altamente vincolante e limitativo, essendo tali risorse rare, estremamente costose, e, punto non meno rilevante, difficili da realizzare autonomamente. Inoltre, richiedendo nell'ambito del loro sviluppo l'applicazione di modelli teorici o procedurali che potrebbero non corrispondere agli obiettivi del nostro lavoro, potrebbero fornire dati in qualche modo distorti, alterando permanentemente il lavoro di identificazione della prominenza.

Nell'ambito dei modelli teorici che fanno uso di fasi di apprendimento per la determinazione dei parametri del modello stesso, ci sembra importante rilevare che, una volta implementata e realizzata la fase di apprendimento, il sistema risulta permanentemente vincolato alla tipologia di dati e all'intrinseca natura delle risorse utilizzate. Impostare i parametri di un modello utilizzando risorse linguistiche di una specifica lingua limita aprioristicamente ogni possibile conclusione

nell'ambito della lingua utilizzata, non consentendo, prima di tutto dal punto di vista metodologico, alcuna generalizzazione interlinguistica.

Il sistema automatico che presentiamo in questa sezione è sostanzialmente diviso in due macro-blocchi fondamentali: il primo si occupa dell'individuazione delle unità segmentali più idonee a supportare le misure acustiche relative al problema in esame (§2.1), mentre il secondo effettua il calcolo dei parametri acustici necessari all'identificazione della prominenza sfruttando le unità segmentali individuate nella prima fase (§2.2). Nel paragrafo 2.3 prenderemo in esame l'algoritmo vero e proprio per il calcolo del livello di prominenza.

2.1 Segmentazione dell'enunciato

La quasi totalità degli studi nel settore sono sostanzialmente concordi nel basare lo studio della prominenza, dal punto di vista delle unità temporali, su unità di tipo sillabico, che vengono considerate il riferimento per la misurazione di tutti i parametri fonetico-acustici correlati col fenomeno in esame.

Il concetto di sillaba, anche se viene comunemente utilizzato a vari livelli, si presenta tuttavia problematico dal punto di vista fonetico-acustico. Se è possibile definire univocamente la sillaba nella lingua scritta, dal punto di vista della lingua parlata lo scenario cambia radicalmente; la definizione di sillaba infatti pertiene a livelli linguistici più alti di quello considerato in questo lavoro, risultando un concetto derivabile dalle teorie fonologiche di una determinata lingua. La trasposizione di tale unità segmentale a livello fonetico comporta numerosi problemi ed è difficilmente definibile con una chiarezza confrontabile alla definizione che riceve negli altri livelli.

Fenomeni frequenti di resillabificazione, ambisillabicità o deformazioni di tipo fonetico sono tutti fenomeni ai quali vengono sottoposte le sillabe quando effettivamente realizzate nella lingua parlata.

Su queste difficoltà di identificazione dei confini sillabici nel dominio fonetico risulta esserci un sufficiente accordo tra gli studiosi (Kopeček, 1999; Noetzel, 1991; Pfitzinger, *et al*. 1996; Taylor, 1995; Wu, *et al*. 1997), che, come nel caso di Goslin, *et al*. (1999), sottolineano come tali problemi rendano il processo di segmentazione in sillabe degli enunciati una operazione estremamente complessa anche per annotatori umani. Quest'ultimo studio, dopo aver verificato sperimentalmente queste difficoltà sottoponendo annotatori umani a verifiche incrociate, propone l'identificazione di unità sillabiche differenti dal concetto di sillaba, proprio per evitare le interferenze dovute ai fenomeni descritti nell'identificazione dei confini di tali unità.

D'altra parte numerosi studi sull'influenza della prominenza sulle varie componenti sillabiche hanno mostrato come le principali modificazioni siano a carico del nucleo (Greenberg, *et al*. 2003; Jenkin, Scordilis, 1996; Silipo, Greenberg 2000; van Bergem, 1993; van Kuijk, Boves, 1999). Questi studi, saldamente fondati su evidenze sperimentali, hanno correlato in maniera affidabile la presenza di prominenza nelle sillabe con un allungamento della durata della vocale che ne costituisce il nucleo, mostrando inoltre come solo il nucleo sillabico appaia subire queste modificazioni in presenza di fenomeni di prominenza.

L'algoritmo vero e proprio per l'identificazione dei nuclei sillabici è basato su una versione modificata del metodo proposto da Mermelstein (1975): un algoritmo di tipo *convex-hull* che utilizza un profilo energetico particolare (ottenuto moltiplicando i contributi energetici in due specifiche bande di frequenza 800-2000 e 2000-3000 Hz), ristretto nella scelta dei possibili punti di segmentazione a quelli proposti dall'algoritmo di Andre-Obrecht (1988), determina la posizione e i confini dei nuclei sillabici. Queste unità segmentali diventeranno la base per il calcolo dei parametri acustici necessari alla determinazione automatica della prominenza.

2.2 Misurazione dei parametri acustici

La tabella 1 descrive sinteticamente le metodologie utilizzate per la determinazione dei quattro parametri acustici considerati in questo studio: la durata dei nuclei sillabici, l'enfasi spettrale, i parametri che catturano i movimenti nel profilo del *pitch* e l'intensità globale. I dettagli implementativi degli algoritmi sono già stati oggetto d'indagine in Tamburini (2003; 2005; 2006).

E' rilevante sottolineare che, al fine di poter evitare che questo tipo di analisi venga influenzata negativamente da fattori quali la diversità dei locutori o degli stati d'animo degli stessi, ogni grandezza fisica utilizzata in questo studio viene accuratamente normalizzata rispetto alla media e alla varianza di tale grandezza all'interno dell'enunciato (*z-score*).

2.3 Identificazione del livello di prominenza

Gran parte delle definizioni del fenomeno della prominenza proposte nella parte introduttiva sono volte a sottolineare come questo abbia una rilevante componente di confronto sull'asse sintagmatico. Il concetto stesso di prominenza appare quindi intimamente connesso con una analisi del contesto della sillaba in esame: definire come prominente o meno una sillaba esaminando unicamente i valori dei parametri all'interno di essa appare quindi estremamente riduttivo, se non addirittura scorretto. Al contrario, una corretta determinazione della prominenza di una determinata unità segmentale dovrebbe essere basata sull'esame dei parametri fonetico-acustici della sillaba stessa e su un loro confronto coi parametri derivati dal contesto. Appare quindi più corretto identificare una sillaba come prominente (o meno) in funzione dei livelli di prominenza delle sillabe che la circondano.

La metodologia proposta definisce matematicamente una opportuna "funzione di prominenza", legata ai parametri acustici dei nuclei sillabici identificati nella fase di segmentazione, che produca valori su un asse continuo. In base alle considerazioni effettuate precedentemente, riteniamo che un sistema per il riconoscimento automatico della prominenza debba fornire valori e valutazioni su una scala continua piuttosto che una scelta forzatamente dicotomica; definiremo quindi il comportamento del metodo che proponiamo su un codominio continuo di valori e forzeremo la scelta tra i due estremi del *continuum* solo per esigenze di valutazione delle prestazioni.

Abbiamo già discusso precedentemente le relazioni che intercorrono tra i parametri fonetico-acustici e i parametri linguistico-prosodici nella definizione di prominenza, in particolare, riassumendo le considerazioni effettuate, abbiamo già evidenziato che valori alti di durata e di enfasi spettrale possono identificare sillabe che contengono un *force accent* mentre grandi variazioni nel profilo del *pitch* all'interno della sillaba e valori alti di energia globale suggeriscono l'esistenza di un *pitch accent*; la presenza di uno o entrambi i fenomeni di *force accent* e *pitch accent* qualifica la sillaba come prominente.

Queste considerazioni puramente qualitative si possono trasformare in legami quantitativi definendo una funzione di prominenza del tipo

$$\mathrm{Prom}^i = SpEmph^i_{SPLH-SPL} \cdot dur^i + en^i_{ov} \cdot (A^i_{event} \cdot D^i_{event}) \qquad (1)$$

dove $SpEmph_{SPLH\text{-}SPL}$ è l'enfasi spettrale, *dur* è la durata temporale del nucleo, en_{ov} è l'energia globale nel nucleo e A_{event}, D_{event} sono i parametri del modello TILT (Taylor, 2000)[1], riferiti al generico nucleo sillabico *i* all'interno dell'enunciato. La struttura della funzione *Prom*, sebbene sembri scelta arbitrariamente, riflette in realtà le relazioni tra i parametri che abbiamo verificato all'interno di questo lavoro e, in particolare, la somma dei due contributi esprime matematicamente la visione di rinforzo reciproco che attribuiamo alle due tipologie accentuali considerate. Onde evitare interferenze dovute ai differenti intervalli di valori dei due argomenti della somma, questi, prima del calcolo, sono stati entrambi normalizzati rispetto al corrispondente valore massimo all'interno dell'enunciato.

Sulla base della definizione della funzione *Prom*, e considerando l'attribuzione della prominenza come un parametro legato al contesto, l'identificazione delle sillabe prominenti corrisponde alla ricerca dei massimi relativi della funzione *Prom*: in quest'ottica il valore della funzione di prominenza per ogni nucleo viene confrontato coi corrispondenti valori dei nuclei adiacenti e, se rappresenta un massimo, il nucleo sillabico corrispondente (e di conseguenza anche la sillaba ad esso associata) viene etichettato come prominente. Non descriviamo i dettagli implementativi di questo metodo che contiene opportuni correttivi per una adeguata gestione di fenomeni specifici, ma preferiamo rimandare, per ragioni di spazio, a Tamburini (2005).

La figura 2 mostra un grafico della funzione di prominenza per l'enunciato "*Cyclical programs will never compile*" (dr1/fdaw0/sx146) tratto dal *corpus* TIMIT.

Il metodo proposto per la determinazione automatica della prominenza frasale è stato sottoposto ad una attenta valutazione sulla lingua inglese (Tamburini, 2006): un numero rilevante di enunciati è stato estratto da tre *corpora* di lingua inglese, etichettato manualmente rispetto al livello di prominenza percepita ed elaborato dal sistema automatico. Le tabelle 2 e 3 mostrano rispettivamente la composizione dei tre *subcorpora* utilizzati per il test e i risultati ottenuti dal sistema.

Vi è una generale convergenza in letteratura nell'identificare un livello di accordo, tra i diversi annotatori chiamati a giudicare il livello di prominenza prosodica dei medesimi enunciati, attorno all'80-90% dei casi esaminati (Jenkin, Scordilis, 1996; Pickering *et al*. 1996). L'intervallo di variazione è piuttosto ampio, ma la valutazione di questi fenomeni risulta essere estremamente complessa anche per annotatori molto esperti e dipende anche sensibilmente dal numero di livelli di prominenza che si intende inserire.

Il sistema automatico per l'etichettatura della prominenza proposto nell'ambito di questo studio raggiunge prestazioni confrontabili con quelle ottenibili da annotatori umani e può quindi essere considerato una valida soluzione per il processo di etichettatura di questo tipo di fenomeno, anche per quanto riguarda il parlato spontaneo.

3. TIPOLOGIA PROSODICA E PROMINENZA

Nella seconda parte di questo contributo presenteremo alcuni risultati preliminari di un progetto volto all'utilizzazione di metodi automatici per l'identificazione della prominenza frasale nel parlato continuo al fine di investigare il contributo di tale fenomeno nell'ambito dello studio e della classificazione delle lingue in una prospettiva tipologica.

L'interessante lavoro di rassegna presentato da Jun (2005) in questo ambito delinea un modello di tipologia prosodica che considera due differenti aspetti di variazione: la prominenza e l'andamento ritmico degli enunciati. Questa visione è supportata anche dai lavori di altri studiosi, tra i quali Fitzpatrik (2000).

Jun ha analizzato in dettaglio varie lingue all'interno di questo quadro di riferimento considerando i lavori eseguiti da numerosi studiosi nell'ambito dei modelli Autosegmentali Metrici della fonologia intonativa, ha proposto una tassonomia completa e analizzato 21 lingue differenti elaborando i vari parametri delle due dimensioni principali di classificazione.

La prima di queste due dimensioni, la prominenza, è stata studiata nel dettaglio da numerosi studiosi e i risultati, dal punto di vista di una classificazione tipologica sembrano essere sufficientemente attestati e relativamente poco controversi. Le lingue possono essere classificate, dal punto di vista di un modello fonologico lessicale del fenomeno in esame, in quattro categorie:

- *stress-accented*,
- *lexical pitch-accented*,
- *non stress-accented and non lexical pitch-accented*,
- *tonal*,

(utilizzando la stessa terminologia di Jun). Tuttavia, come mostrato da Fitzpatrik (2000), questa suddivisione risulta essere tutt'altro che netta, come spesso accade in questo campo, e anzi sono numerosi i casi di lingue che presentano la covariazione di alcuni di questi tratti contemporaneamente, suggerendo ancora una volta come l'approccio migliore per affrontare questo tipo di problematiche sia quello di considerarle dimensioni di variazione all'interno di un *continuum* multidimensionale (si veda anche Grabe, 2002).

D'altra parte la tradizionale classificazione della dimensione ritmica delle lingue che coinvolge tre categorie – *stress-timed*, *syllable-timed*, e *mora-timed* – è meno accettata e il concetto di isocronia è spesso visto come problematico. Vi sono infatti studi basati su dati sperimentali che tendono a supportare tale visione (Low *et al*. 2001; Ramus *et a*l. 1999) mentre altri studi sperimentali tendono a criticare tale suddivisione e il concetto stesso di isocronia (Pamies Bertran, 1999; Warner, Arai, 2001).

In questo lavoro ci siamo concentrati unicamente sulla dimensione relativa alla prominenza e quindi non assumeremo alcuna posizione nei confronti della dimensione ritmica degli enunciati.

Ci sembra rilevante sottolineare come il modello proposto da Jun sia un modello prevalentemente fonologico e, come sottolinea la stessa Jun (2005: 440-441),

> *Acoustic correlates of postlexical pitch accent are language specific, and this difference will not be captured in a model of prosodic typology where languages are compared based on phonological categories*.

questo modello non sia in grado di catturare le relazioni che intercorrono tra i parametri acustici in grado di supportare il fenomeno della prominenza nelle varie lingue.

Questo lavoro tenta quindi di iniziare un'indagine di carattere tipologico che si basi, anziché su un modello fonologico, su un modello fonetico-acustico basato sulla valutazione "pesata" dei parametri acustici che abbiamo visto essere alla base delle due tipologie di accento frasale (*force accent* e *pitch accent*) che inducono la percezione della prominenza.

3.1 Un approccio fonetico/acustico

La prominenza, nella sua manifestazione acustica, assume differenti connotazioni a seconda della lingua in esame e, benché alcune lingue sembrino mostrare tratti comuni, non risulta finora possibile definire una matrice costante tra di esse.

In prima istanza è pensabile di modificare la funzione di prominenza *Prom* definita nel paragrafo 2.3 nel modo seguente,

$$\begin{aligned} \mathrm{MProm}^i \quad = \quad & W_{FA} \cdot \left\lfloor SpEmph^i_{SPLH-SPL} \cdot dur^i \right\rfloor + \\ & W_{PA} \cdot \left[en^i_{ov} \cdot \left(A^i_{event}(at_M, at_m) \cdot D^i_{event}(at_M, at_m) \right) \right] \end{aligned} \qquad (2)$$

ossia introducendo due parametri W_{FA} e W_{PA} in grado di 'pesare' i contributi dei *force accent* e dei *pitch accent* nel calcolo del livello di prominenza. Per meglio adattare il comportamento della funzione *MProm* alle caratteristiche delle varie lingue è opportuno introdurre due ulteriori dimensioni di variazione, in grado di catturare le differenti tipologie di allineamento tra l'evento intonativo e la sillaba a cui si riferisce; introdurremo quindi due ulteriori parametri di controllo, che indicheremo con at_M e at_m (*alignment type*), in grado di catturare le differenti modalità di allineamento, rispettivamente, degli eventi intonativi legati a un massimo nel profilo del *pitch* e degli eventi intonativi legati a un minimo nel profilo (si veda la figura 3). I due parametri A_{event} e D_{event} sono quindi da considerarsi come funzione delle tipologie di allineamento che si intende utilizzare.

Se raccogliamo tutti i parametri in gioco in un vettore **W** = (W_{FA}, W_{PA}, at_M, at_m), allora l'equazione (2) potrebbe rappresentare una componente universale nella definizione della prominenza, mentre il vettore dei pesi **W** permetterebbe di adattare il comportamento della funzione *MProm* alla specifica lingua, modificando i rapporti tra le varie componenti enfatizzando o riducendo i contributi di alcune di esse, nell'ipotesi che i quattro parametri acustici identificati precedentemente siano tutti e soli quelli universalmente coinvolti nel supportare il fenomeno in esame.

Al fine di valutare tali ipotesi è stato predisposto un insieme di esperimenti completamente differente rispetto a quello presentato nel paragrafo 2.3: in questo caso tutta la parte del sistema dedicata all'identificazione dei nuclei sillabici non è stata utilizzata nelle elaborazioni, ma sono state considerate, dopo opportuna verifica, le segmentazioni degli enunciati fornite a corredo dei *corpora* considerati, in modo da evitare tutte le tipologie di errore potenzialmente introdotte durante la determinazione dei confini dei nuclei sillabici con metodologie automatiche.

Sono state scelte alcune lingue per le quali fossero disponibili *corpora* di lingua parlata sufficientemente estesi e standardizzati da utilizzarsi come basi empiriche per gli esperimenti. La tabella 4 mostra la struttura dei *subcorpora* considerati negli esperimenti, sia dal punto di vista della tipologia del parlato sia nei confronti della effettiva composizione.

L'idea, o l'ipotesi, che sottende tali esperimenti è che la migliore combinazione dei parametri del vettore **W** per una data lingua sia anche la combinazione che consente le migliori performance in termini di classificazione del fenomeno. O, rovesciando l'ipotesi, che la migliore performance, ottenuta mediante una scansione parametrica su tutti i possibili valori dei quattro parametri che compongono il vettore **W**, ci fornisca la definizione del vettore dei parametri per la lingua in esame.

Ogni enunciato dei *subcorpora* considerati è stato accuratamente classificato da parlanti nativi rispetto alla presenza o assenza di prominenza percepita, senza l'utilizzazione di alcuno strumento di analisi fatta eccezione della possibilità di ascoltare e riascoltare gli enunciati, o porzioni di essi, un numero arbitrario di volte.

Il sistema di classificazione della prominenza presentato nelle sezioni precedenti è stato quindi applicato ai vari *subcorpora* al fine di ottenere la miglior combinazione di parametri per le lingue considerate negli esperimenti.

Il grafico in figura 4 mostra alcuni risultati preliminari ottenuti utilizzando tali metodologie; va osservato che, a causa della complessità del fenomeno in esame, dell'effettiva eterogeneità dei *corpora* considerati e della modalità di classificazione della prominenza utilizzata dagli annotatori umani è opportuno considerare i risultati ottenuti unicamente per ricavare indicazioni sulle tendenze delle varie lingue e non su effettivi valori quantitativi precisi dei vari parametri in gioco.

Il grafico mostra in ascissa i valori del rapporto tra i due parametri che pesano i contributi dei *force accent* e dei *pitch accent*; essendo coinvolti unicamente due parametri è sufficiente considerare il loro rapporto, questo consente di mantenere il grafico in due dimensioni e non altera i risultati ottenuti. Per valori del rapporto vicini a 1 avremo una ugual importanza delle due tipologie accentuali, per valori superiori a 1 una predominanza dei *force accent*, mentre per valori minori di 1 una predominanza dei *pitch accent*. A livello interpretativo la predominanza di una tipologia accentuale sull'altra ci porterebbe a concludere che i parlanti nativi di quella lingua tenderebbero a considerare come predominante detta tipologia, e i parametri acustici ad essa associati, nel supportare la prominenza percepita.

In ordinata il grafico mostra la percentuale di accuratezza del sistema nel classificare la prominenza quando confrontata con quella percepita dai parlanti nativi, mentre nella legenda, tra parentesi accanto ai nomi dei *corpora*, vi sono i valori di at_M e at_m che hanno consentito di ottenere i migliori risultati.

Le tre curve relative alla lingua inglese mostrano, seppur con differenti livelli di performance, una sostanziale convergenza nell'ascrivere al *force accent* un ruolo principale nell'identificazione della prominenza, in quanto il sistema automatico fornisce i migliori livelli di classificazione per valori del rapporto $W_{FA}/W_{PA}>1$. Questi risultati sono in linea con quelli ottenuti da alcuni studi

recenti sulla lingua inglese (Silipo, Greenberg, 2000; Kochanski *et al*. 2005) che tendono ad attribuire un ruolo predominante, relativamente a questa lingua, ai correlati acustici legati alla durata e all'intensità nel supportare la prominenza, o alternativamente, mostrano come i parlanti di tale lingua tendano ad utilizzare preferibilmente i *force accent* per rendere una sillaba prominente.

Per quanto riguarda la lingua italiana si registra invece un sostanziale equilibrio tra le due tipologie accentuali in accordo coi lavori di Bertinetto (1981) e D'Imperio (2000) che ascrivono al *pitch* e alle misure di durata ruoli egualmente predominanti.

Contrariamente alle comuni aspettative che vedono la lingua tedesca molto simile a quella inglese, per quanto riguarda i fenomeni considerati, ma in linea con molti lavori tra cui quelli di Portele e Heuft (1997) e Wagner (2005), i dati sul tedesco tendono ad indicare i *pitch accent* come i migliori indicatori di prominenza.

E' opportuno considerare le ultime due curve con ancora maggiore cautela della precedenti, a causa della limitata quantità dei dati e delle modalità di etichettatura manuale alla quale sono stati sottoposti: contrariamente alle altre lingue, e quindi *corpora*, che abbiamo esaminato, queste due curve sono state prodotte utilizzando etichettature eseguite da un unico parlante nativo e che quindi necessitano di ulteriori verifiche. Il dato indicativo che si può ricavare da tali risultati indica anche in questi due casi un predominanza dei *pitch accent* che, nel caso della lingua francese, tendono a trovare riscontri precisi in letteratura (Mertens, 1991).

4. Conclusioni

I risultati descritti, ancorché preliminari, mostrano tendenze sufficientemente in accordo con le conclusioni ottenute da altri studiosi sulle lingue considerate e possono portare a concludere che l'ipotesi alla base di questo tipo di esperimenti, e l'approccio stesso ad un'analisi interlinguistica di questi fenomeni che abbiamo proposto, possano considerarsi sufficientemente validi.

I risultati forniti dagli esperimenti effettuati sono da considerarsi preliminari e suscettibili di ulteriori verifiche; il lavoro che abbiamo presentato non pretende assolutamente di fornire dati conclusivi sui rapporti tra i parametri acustici e il fenomeno della prominenza, ma piuttosto mira ad essere uno studio di fattibilità, un progetto pilota, per verificare le potenzialità di tali approcci a questo problema di analisi prosodica da un punto di vista interlinguistico e tipologico.

Si possono considerare molteplici linee di sviluppo dell'indagine.

Prima di tutto ci sembra necessaria l'utilizzazione di dati maggiormente omogenei tra loro: i *corpora* considerati negli esperimenti presentano diverse modalità di progettazione e costruzione e riguardano differenti tipologie di lingua parlata. Sembra opportuno procedere ad una ripetizione degli esperimenti utilizzando *corpora* omogenei, auspicabilmente riferiti al parlato spontaneo.

In seconda istanza le dimensioni dei *subcorpora* utilizzati negli esperimenti andrebbero notevolmente incrementate, sia dal punto di vista del numero degli enunciati sia per quanto riguarda la composizione del *subcorpus* (numero e varietà dei locutori).

Terzo punto, ma non meno importante dei precedenti, l'annotazione del livello di prominenza percepita da parte dei parlanti nativi della lingua in esame dovrebbe essere condotta in modo più esteso e seguendo protocolli rigorosi e verificabili. Interessante sarebbe anche la possibilità di esplorare tali fenomeni sfruttando *corpora* etichettati non secondo la presenza o l'assenza di prominenza percepita, ma secondo un giudizio sul livello di prominenza espresso sfruttando più valori (vi sono studi che utilizzano ad esempio 4 (Jensen, 2004) o 32 livelli distinti).

Sul fronte più prettamente modellistico-tecnologico, l'annotatore automatico ha mostrato di poter raggiungere risultati confrontabili con quelli ottenibili da annotatori umani e, una volta modificato per utilizzare la funzione di prominenza *MProm* (2), risulta essere sufficientemente flessibile per poter essere adattato con successo a numerose lingue.

NOTE

1 - Si noti che nel caso non sia presente alcun evento intonativo all'interno della sillaba questi due parametri valgono 0.

BIBLIOGRAFIA

Albano Leoni F., 2003, *Tre progetti per l'italiano parlato: AVIP, API, CLIPS*. In: Maraschio N., Poggi Salani T., (a cura di), *Italia Linguistica. Anno mille, anno duemila, Atti del XXXIV Congresso della Società di Linguistica Italiana (SLI)*, Roma, Bulzoni, 675-683.

Anastasakos A., Schwartz R. and Shu H., 1995, *Duration modeling in large vocabulary speech recognition*. In: *Proceedings of ICASSP '95*, 628-631.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R., 1991, *The HCRC Map Task Corpus*. "Language and Speech", 34: 351-366.

Andre-Obrecht R., 1988, *A New Statistical Approach for the Automatic Segmentation of Continuous Speech Signals*, "IEEE Trans. on ASSP", 36: 29-40.

Auran C., Bouzon C. & Hirst D., 2004, *The Aix-MARSEC project: an evolutionary database of spoken British English and automatic tools*. In: *Proceedings of Speech Prosody 2004*, Nara, Giappone, 143-146.

Bagshaw P.C., 1994, *Automatic prosodic analysis for computer-aided pronunciation teaching*, Tesi di dottorato, Università di Edimburgo.

Beckman M.E., 1986, *Stress and non-stress accent*, Dordrecht, Foris.

Bertinetto P.M.,1981, *Strutture prosodiche dell'italiano*, Firenze:Accademia della Crusca.

Bolinger D., 1958, *A theory of pitch-accent in English*, "Word", 14:109-149.

Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B., 2005, *A Database of German Emotional Speech*. In: *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech 2005)*, Lisbona, 1517-1520.

Campione E., Véronis J., 1998, *A multilingual prosodic database*. In: *Proceedings of ICSLP'98*, Sydney, 0845.

Couper-Kuhlen E., 1986, *English prosody*. London: Edward Arnold.

D'Imperio M., 2000, *Acoustical-perceptual correlates of sentence prominence in Italian*. In: *The Ohio State University Working Papers in Linguistics, Columbus OH*, The Ohio State University, 54: 59-79.

Espy-Wilson C.Y., 1994, *A feature-based semivowel recognition system*, "Journal of the Acoustical Society of America", 96: 65-72.

Fant G., Kruckenberg A., Liljencrants, J., 2000, *Acoustic-phonetic Analisys of Prominence in Swedish*. In: Botinis, A. (Ed.), *Intonation*, Kluwer Academic Publisher, 55-86.

Fitzpatrik J., 2000, *On intonational typology*, In: P. Siemund (ed.) *Methodological Issues in Language Typology. Sprachtypologie und Universalienforshung*, 53:88-96.

Garofolo J.S., Lamel L.F., Fisher W.M., Fiscus J.G., Pallett D.S., Dahlgren, N.L., 1993, *The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus Cdrom*. NIST.

Goslin J., Content A., Frauenfelder U.H., 1999, *Syllable segmentation: are humans consistent?* In: *Proceedings of 6th European Conference on Speech Communication and Technology (Eurospeech '99)*, Budapest.

Grabe E., 2002, *Variation Adds to Prosodic Typology*. In: *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 127-132.

Greenberg S., Carvey H., Hitchcock L., Chang S., 2003, *The Phonetic Patterning of Spontaneous American English Discourse*. In: *Proceedings of ISCA/IEEE Workshop on Spontaneous Speech Processing and Recognition*, Tokyo, MAO5.

Heldner M., 2003, *On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish*, "Journal of Phonetics", 31: 39-62.

Jenkin K.L., Scordilis M.S., 1996, *Development and comparison of three syllable stress classifiers*. In: *Proceedings of ICSLP '96*, Philadelphia, 733-736.

Jensen C., 2004, *Stress and Accent. Prominence relations in Southern Standard British English*. Tesi di dottorato, Università di Copenhagen.

Jun S., 2005, *Prosodic Typology*. In: S. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford University Press, 430-458.

Kochanski G., Grabe E., Coleman J., Rosner B., 2005, *Loudness predicts prominence: Fundamental frequency lends little*, "Journal of the Acoustical Society of America", 118: 1038-1054.

Kohler K.J., 2003, *Neglected categories in the modelling of prosody - Pitch timing and non-pitch accents*. In: *Proceedings of the XVth International Congress of Phonetic Sciences (ICPhS'03)*, Barcelona, 2925-2928.

Kohler K.J., 2005, *Form and Function of Non-Pitch Accents*. In: *Prosodic Patterns of German Spontaneous Speech*, AIPUK, 35a: 97-123.

Kohler K.J., 2006, *What is emphasis and how is it coded?* In: *Proceedings of Speech Prosody 2006*, Dresden, 748-751.

Kopeček I., 1999, *Speech recognition and syllable segments*. In: *Proceedings of Workshop on Text, Speech and Dialogue (TSD '99)*, LNAI 1692, 203-208.

Ladd D.R., Verhoeven J., Jacobs K., 1994, *Influence of adjacent pitch accents on each other perceived prominence: two contradictory effects*, "Journal of Phonetics", 22: 87-99.

Ladd D.R., 1996, *Intonational Phonology*. Cambridge:Cambridge University Press.

Low E.L., Grabe E., Nolan F., 2001, *Quantitative characterisation of speech rhythm: Syllable-timing in Singapore English*, "Language and Speech", 43:377-401.

Mermelstein P., 1975, *Automatic segmentation of speech into syllabic units*, "Journal of the Acoustical Society of America", 58: 880-883.

Mertens P.,1991, *Local prominence of acoustic and psychoacoustic functions and perceived stress in French*. In: *Proceedings of the XIIth International Congress of Phonetic Sciences (ICPhS'91)*, Aix-en-Provence, 218-221.

Noetzel A., 1991, *Robust syllable segmentation of continuous speech using neural networks*. In: *Proceedings of IEEE Electro International Conference Record*, New York, 580-585.

Pamies Bertran, A. 1999, *Prosodic Typology: on the Dychotomy between Stress. Timed and Syllable-Timed Languages*, "Language Design", 2:103-130.

Pfitzinger H., Burger S., Hid S., 1996, *Syllable detection in read and spontaneous speech*. In: *Proceeedings of ICSLP '96*, Philadelphia, 1261-1264.

Pickering B., Williams B., Knowles G., 1996, *Analysis of transcriber differences in SEC*. In: G. Knowles, A. Wichmann, P. Alderson (Eds), *Working with speech*, London: Longman, 61-86.

Portele T., Heuft, B., 1997, *Towards a prominence-based syntesis system*. "Speech Communication", 21: 61-72.

Ramus F., Nespor M., Mehler J., 1999, *Correlates of linguistic rhythm in the speech signal*, "Cognition", 73:265-292.

Silipo R., Greenberg S., 2000, *Automatic Detection of Prosodic Stress in American English Discourse*, International Computer Science Institute, TR-00-001.

Sluijter A., van Heuven V., 1996, *Acoustic correlates of linguistic stress and accent in Dutch and American English*. In: *Proceedings of ICSLP' 96*, Philadelphia, 630-633.

Sluijter A., van Heuven V., Pacilly J., 1997, *Spectral balance as a cue in the perception of linguistic stress*. "Journal of the Acoustical Society of America", 101: 503-513.

Spencer A., 1996, *Phonology*, Oxford:Blackwell.

Streefkerk B.M., 1996, *Prominent accent and pitch movements*. "Inst. of Phon. Sciences Proceedings", Università di Amsterdam, 20:111-119.

Tamburini F., 2003, *Automatic Prosodic Prominence Detection in Speech using Acoustic Features: an Unsupervised System*. In: *Proceedings of 8th European Conference on Speech Communication and Technology (Eurospeech 2003)*, Geneva, 129-132.

Tamburini F., 2005, *Fenomeni prosodici e prominenza: un approccio acustico*. Bologna: BUP.

Tamburini F., 2006, *Reliable Prominence Identification in English Spontaneous Speech*. In: *Proceedings of Speech Prosody 2006*, Dresden, PS1-9-19

Tamburini F., Caini C., 2005, *An automatic system for detecting prosodic prominence in American English continuous speech*, "International Journal of Speech Technology", 8: 33-44.

Talkin D., 1995, *A robust algorithm for pitch tracking (RAPT)*. In: W.B. Kleijn, K.K. Paliwal (Eds.), *Speech coding and synthesis*, New York: Elsevier, 495-518.

Taylor P.A., 1992, *A phonetic model of English intonation*, Tesi di dottorato, Università di Edimburgo.

Taylor P.A., 1995, *Using Neural Networks to Locate Pitch Accents*. In: *Proceedings of 4th European Conference on Speech Communication and Technology (Eurospeech '95)*, Madrid, 1345-1348.

Taylor P.A., 2000, *Analysis and Synthesis of Intonation using the Tilt Model*, "Journal of the Acoustical Society of America", 107: 1697-1714.

Terken J., 1991, *Fundamental frequency and perceived prominence*, "Journal of the Acoustical Society of America", 89:1768-1776.

van Bergem, D., 1993, *Acoustic vowel reduction as a function of sentence accent, word stress and word class on the quality of vowels*, "Speech Communication", 12: 1-23.

van Kuijk D., Boves L., 1999. *Acoustic characteristic of lexical stress in continuous telephone speech*. "Speech Communication", 27: 95-111.

van Son R.J.J.H., Binnenpoorte D., van den Heuvel H., Pols, L.C.W., 2001, *The IFA Corpus: a Phonemically Segmented Dutch 'Open Source' Speech Database*. In: *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech* 2001), Aalborg, 2051-2054.

Wagner P, 2005, *Great Expectations – Introspective vs. Perceptual Prominence Ratings and their Acoustic Correlates*. In: *Proceedings of 9th European Conference on Speech Communication and Technology (Interspeech 2005)*, Lisbona, 2381-2384.

Warner N., Arai T. 2001, *Japanese Mora-Timing: A Review*, "Phonetica" 58:1-25.

Wightman C.W., Ostendorf M., 1994, *Automatic Labeling of Prosodic Patterns*, "IEEE Transactions on Speech and Audio Processing", 2: 469-481.

Wu S., Shire M.L., Greenberg S., Morgan N., 1997, *Integrating syllable boundary information into speech recognition*. In: *Proceedings of ICASSP '97*, Munich, 987-990.

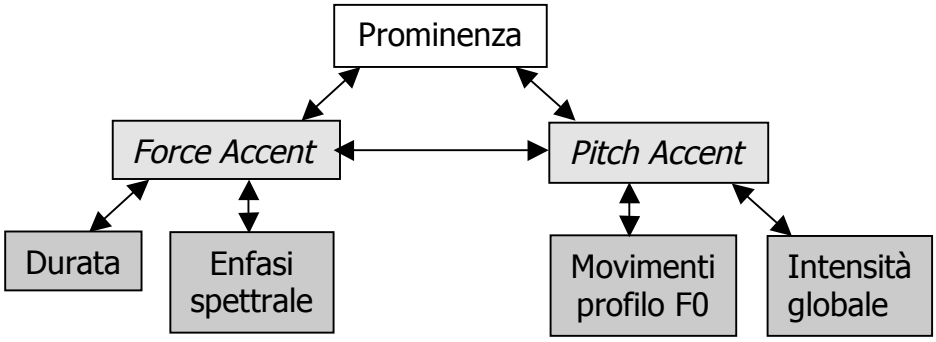

Figura 1: Struttura e relazioni tra il parametro percettivo della prominenza, i parametri linguistico-prosodici e i parametri acustici così come vengono considerarti nel modello proposto.

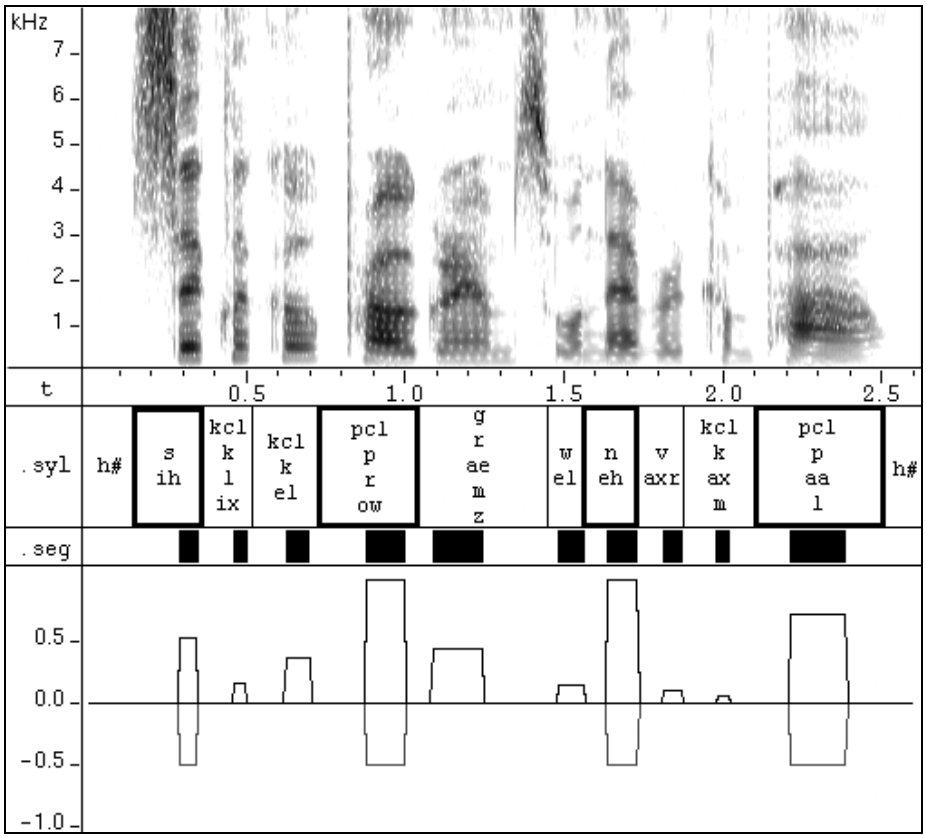

Figura 2: Profilo della funzione di prominenza per l'enunciato "*Cyclical programs will never compile*" (dr1/fdaw0/sx146) tratto dal TIMIT corpus. Dall'alto: lo spettrogramma, la segmentazione in sillabe (mostrata unicamente come riferimento), i nuclei sillabici identificati dal sistema automatico (segnalati da un settore scuro nella traccia "seg"), e infine il valore della funzione di prominenza (*Prom*) per ogni nucleo identificato dalla procedura di segmentazione (sopra l'asse dello 0). I nuclei prominenti identificati dal sistema automatico sono segnalati al di sotto dell'asse dello 0, mentre le sillabe prominenti, classificate manualmente, sono indicate da un rettangolo nero nella traccia relativa alla sillabazione ("syl").

Parametro Acustico	**Descrizione**
Durata del nucleo sillabico	Durata temporale del nucleo sillabico normalizzato considerando la durata media e la varianza delle durate dei nuclei sillabici contenuti nell'enunciato (*z-score*).
Enfasi spettrale	Parametro SPLH-SPL (Fant *et al*. 2000) normalizzato (*z-score*).
Movimenti nel profilo di F0	Rappresentazione dei movimenti nel profilo del *pitch* utilizzando il modello TILT (Taylor, 2000), ottenuti dal programma *ESPS get_f0* (Talkin, 1995).
Intensità globale	Energia RMS calcolata nella banda di frequenza 50-5000 Hz normalizzata (*z-score*).

Tabella 1: Metodologie utilizzate in questo studio per la determinazione dei quattro parametri acustici considerati.

Corpus	**Tipologia**	**Enunciati**	**Sillabe**	**Locutori**
TIMIT	p. letto	382	4780	51 (31m, 20f)
AixMarsec	p. radiofonico	43	704	3 (2m, 1f)
HCRC maptask	p. spontaneo	62	901	10 (5m, 5f)

Tabella 2: Struttura e composizione dei *subcorpora*, relativi alla lingua inglese, utilizzati per la valutazione del sistema di identificazione automatica della prominenza frasale.

Corpus	**Errore**	**Inserimenti**	**Cancellazioni**
TIMIT	18.64%	9.52%	9.12%
AixMarsec	18.89%	10.37%	8.52%
HCRC maptask	20.75%	8.99%	11.76%

Tabella 3: Prestazioni ottenute dal sistema di identificazione automatica della prominenza frasale sulla lingua inglese.

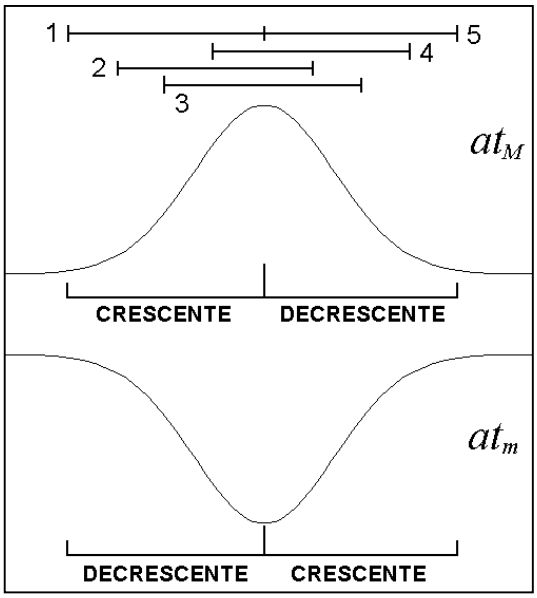

Figura 3: Tipologie di allineamento tra l'evento intonativo e il nucleo sillabico.

Lingua / *Corpus*	**Tipologia**	**Enunciati**	**Sillabe**	**Locutori**
Inglese Americano TIMIT (Garofolo *et al*. 1993)	p. letto	382	4780	51 (20f, 31m)
Inglese Britannico AixMARSEC (Auran *et al*. 2004)	p. radiofonico	114	1641	8 (3f, 5m)
Inglese Britannico HCRC (Anderson *et al*. 1991)	p. spontaneo (MapTask)	62	901	10 (5f, 5m)
Italiano CLIPS (Albano Leoni, 2003)	p. radio-TV	84	2455	19 (5f, 14m)
Tedesco EmoDB (Burkhardt *et al*. 2005)	p. letto emotivo	77 (en.neutri)	935	10 (5f, 5m)
Olandese IFA (van Son *et al*. 2003)	p. letto	103	2006	7 (4f, 3m)
Francese MULTEXT(Campione, Veronis, 1998)	p. letto	80	1456	8 (4f, 4m)

Tabella 4: Tipologia e composizione dei *subcorpora* utilizzati negli esperimenti.

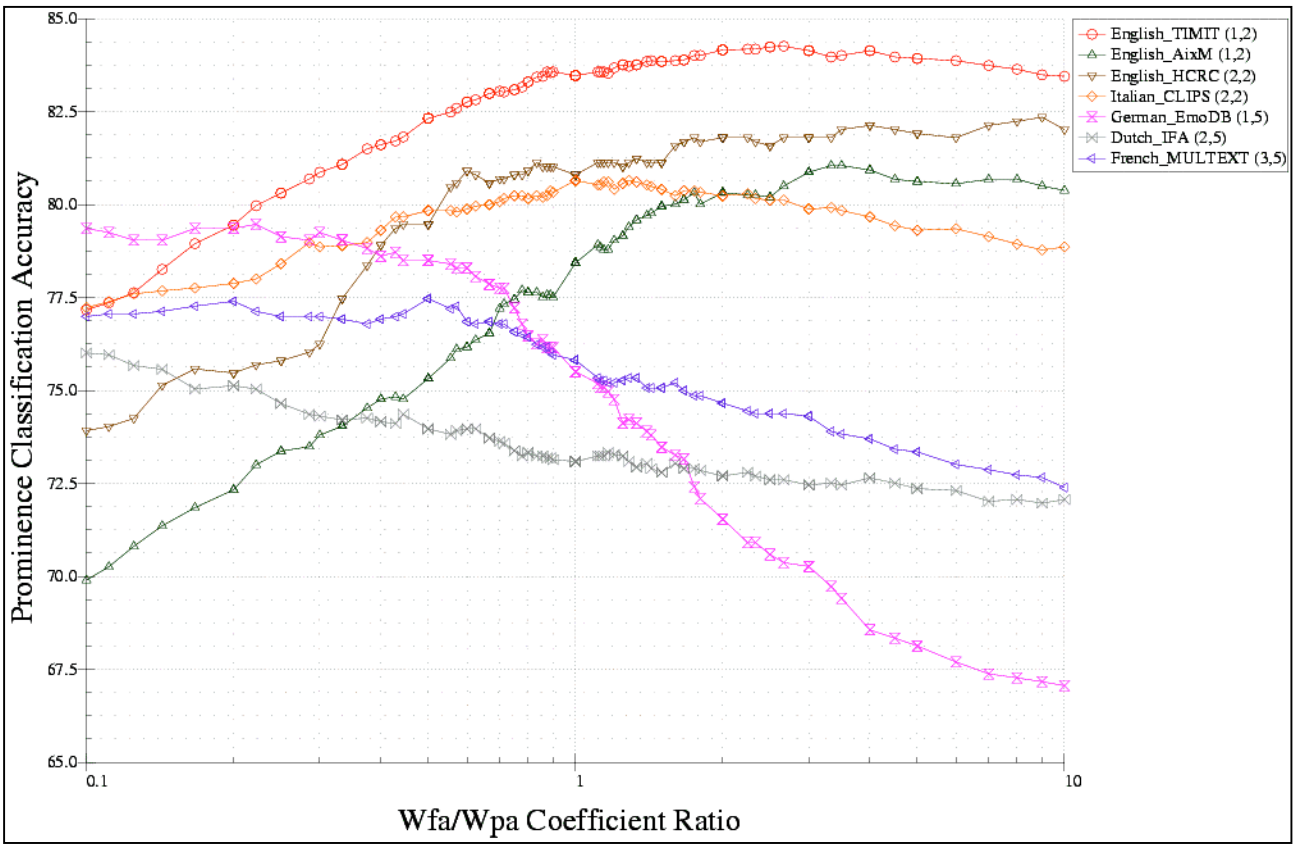

Figura 4: Risultati di accuratezza ottenuti dal classificatore automatico al variare del rapporto W_{FA}/W_{PA}.

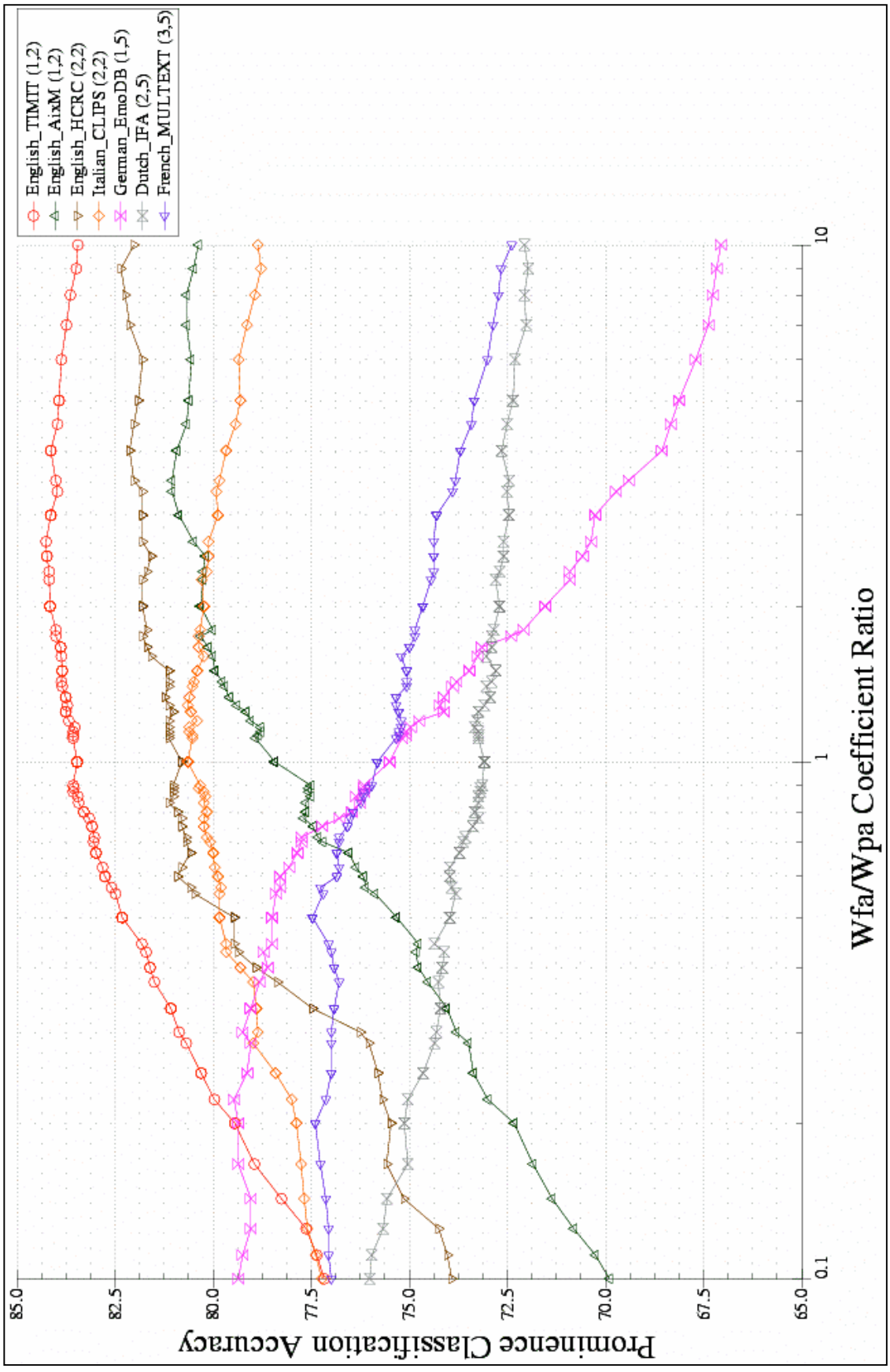

Figura 4: Risultati di accuratezza ottenuti dal classificatore automatico al variare del rapporto W_{FA}/W_{PA}.