

LA VELOCITÀ DI ARTICOLAZIONE COME PARAMETRO IDENTIFICATIVO: POTENZIALITÀ E LIMITI PER LA CARATTERIZZAZIONE DEL PARLANTE¹

Antonio Federico*, Laura Mori**, Andrea Paoloni ***

*Enea; **Università di Roma 'La Sapienza'; ** *Fondazione Ugo Bordoni
federico@casaccia.enea.it; laura.mori@unitus.it; pao@fub.it

1. INTRODUZIONE

Nell'ambito della caratterizzazione del parlante oltre ai più noti parametri di tipo spettrale, direttamente connessi con la fisiologia dell'apparato fonatorio (F0 e frequenze formantiche), è possibile avvalersi anche di informazioni derivanti da parametri di altra natura, ad esempio i parametri temporali. Tra questi la velocità di articolazione (VA) viene descritta in letteratura come il meno soggetto alla variabilità individuale.

Numerosi lavori confermano la validità di questo assunto in varie lingue europee: Goldman-Eisler (1968); Grosjean-Deschamps (1975); Miller-Grosjean-Lomanto (1984); Magno-Caldognetto-Vagges (1991); Magno Caldognetto-Ferrero-Zmarich (1996); Künzel (1997).

Per velocità di articolazione si intende la durata media della sillaba, calcolata dividendo il numero di sillabe fonetiche² individuate nella catena fonetica per la durata della stessa. Inoltre si precisa che, in questa sede, al fine di utilizzare la stima della velocità di articolazione a scopo di identificazione del parlante, si è ritenuto funzionale allo studio prendere in considerazione le realizzazioni in cui il parametro risultasse più stabile³.

Per quanto concerne l'utilizzazione della velocità di articolazione nell'identificazione del parlante in ambito forense, uno studio sperimentale condotto sulla validità di alcuni parametri prosodici del tedesco ha dimostrato che la velocità di articolazione è sufficientemente stabile a livello individuale, varia nei diversi parlanti e non è influenzata in maniera significativa dalla modalità comunicativa e quindi può essere considerata un parametro utile (cfr. Künzel, 1997).

Al fine di testare la stabilità della velocità di articolazione in italiano, precedentemente (Mori e Paoloni, 2004) ci siamo concentrati su esperimenti che ci hanno consentito di verificare la sensibilità della velocità di articolazione alla variazione diafasica, in

¹ Il testo è frutto del lavoro comune dei tre autori, tuttavia a fini accademici si precisa che i paragrafi sono stati redatti come segue: § 1 (Laura Mori), § 2-4-5 (Antonio Federico), § 3-6 (Andrea Paoloni).

² Per determinare il numero delle sillabe fonetiche ci si basa sul profilo dell'energia, con i relativi picchi di intensità, e si effettua una segmentazione basata sull'analisi spettro-acustica (forma d'onda e sonogramma) e sul riscontro uditivo.

³ Come già specificato in precedenza (Mori e Paoloni, 2004), considerato l'obiettivo dello studio di tale parametro, finalizzato ad individuare l'abitudine articolatoria caratteristica del parlante, la velocità di articolazione è stata misurata solo su catene foniche composte da un minimo di sei sillabe comprese tra due pause silenti, senza pause vocalizzate e allungamenti segmentali.

riferimento alla modalità di parlato (letto, spontaneo, ripetuto) e al registro (formale vs. informale).

I risultati ottenuti (Federico *et al.*, in stampa) ci consentono di affermare che la velocità di articolazione può contribuire all'identificazione del parlante in ambito forense, sebbene si riconoscano ancora dei limiti connessi con la variabilità del parametro in relazione alla modalità d'eloquio e alla struttura delle catene fonetiche⁴. In questa sede presentiamo una tecnica di compensazione che consente di ridurre l'incidenza della composizione fonetica, migliorando la discriminazione anche quando il materiale disponibile sia limitato.

2. LA VELOCITÀ DI ARTICOLAZIONE COME PARAMETRO IDENTIFICATIVO

Al fine di utilizzare la velocità di articolazione nell'identificazione del parlante è necessario analizzare a fondo le proprietà statistiche, in particolare la stabilità e la convergenza, di tale parametro.

Il primo modello statistico della velocità di articolazione è stato sviluppato nel 2004 sulla base dei dati della FUB e del RACIS. L'elemento statistico modellato è la singola sillaba. In prima istanza sono stati calcolati i momenti delle distribuzioni probabilistiche dei dati campionari, segnatamente il centroide della popolazione, cioè la media campionaria della velocità di articolazione sillabica calcolata su tutti i parlanti, e le varianze intra-parlante ed inter-parlante. Il modello è stato formulato per i soli dati campionari attribuiti ad un discorso di tipo spontaneo.

È stata sottoposta a test l'ipotesi di eventuale normalità delle distribuzioni probabilistiche. Poiché il modello è stato inserito nel sistema SPREAD/AR, derivato dal sistema SPREAD (Federico; 2005) che esegue il riconoscimento del parlante a partire dalle formanti vocaliche, che opera con un approccio generale non-parametrico, l'ipotesi di normalità non è necessaria per la distribuzione di probabilità inter-parlante. SPREAD-AR esegue infatti la modellazione non-parametrica di tale distribuzione. Viene invece ritenuta l'ipotesi di normalità per la distribuzione delle medie della velocità di articolazione per campioni al di sopra di dieci sillabe.

È stata sottoposta a test la proprietà di convergenza del campione, indispensabile per la applicazione di tale parametro a procedimenti di riconoscimento del parlante.

L'ipotesi corrisponde all'assunto che il processo di articolazione sillabica sia aleatorio, stazionario ed ergodico e che quindi, tra l'altro, le varianze delle medie campionarie convergano a zero come n^{-1} .

Nessuna delle tre condizioni è nei fatti soddisfatta pienamente. La più temibile delle eccezioni deriva dalla probabile sussistenza di una forma deterministica, la cui legge non è al momento nota, di dipendenza tra velocità di articolazione e modalità di eloquio che violerebbe, in misura difficile da determinare, la ipotesi di ergodicità ed in sub-ordine di stazionarietà. Una prova di convergenza è stata fatta per campioni del parlato spontaneo.

Le proprietà di convergenza sono rappresentate in Fig. 1 dove il dato campionario (*samples*) è confrontato con n^{-1} (*nominal*) e con la curva di regressione campionaria (linea continua). La proprietà di convergenza è rispettata.

⁴ La velocità di articolazione varia in funzione del numero di fonemi presenti nella sillaba, a seconda che si tratti di sillaba aperta del tipo CV o di sillabe più complesse, e anche in funzione dei fonemi che la compongono (ad. es. presenza di consonanti geminate o rafforzate).

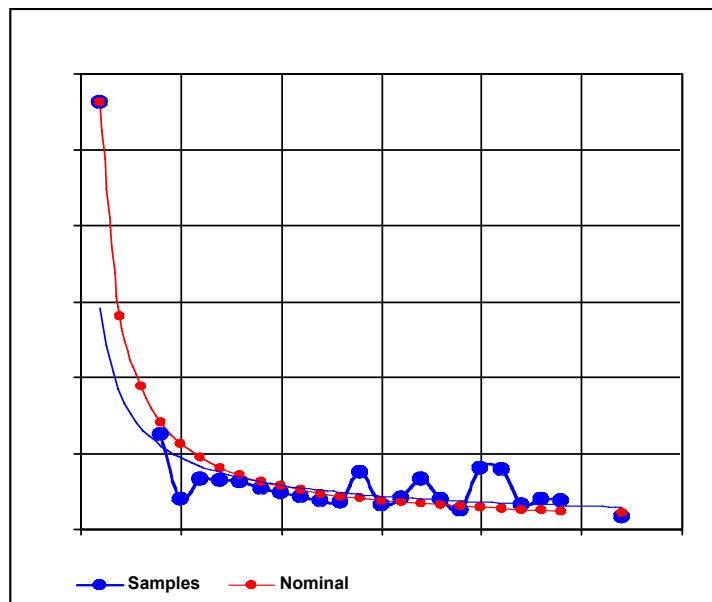


Fig. 1 Correlazione tra la velocità di articolazione (AR: Articulation Rate) e i parametri spettrali frequenza fondamentale e frequenze formanti delle vocali

3. LA VELOCITÀ DI ARTICOLAZIONE : POTENZIALITÀ E LIMITI

Studi precedentemente condotti al fine di valutare l'eventuale impiego della velocità di articolazione come per le procedure identificative a fini forensi, hanno messo in evidenza che la velocità di articolazione non è correlata con i valori delle formanti utilizzati per il riconoscimento e la verifica del parlante (Federico *et al.*, in stampa).

Il test di correlazione, effettuato su materiale ricavato dal database del *Raggruppamento Investigazioni Scientifiche dell'Arma dei Carabinieri*⁵ (Tabella 1), dimostra che non esiste una correlazione tra questo parametro di tipo temporale e parametri spettrali valori della estratti dallo stesso segnale.

La velocità di articolazione è soggetta ad una variazione locale (Giannini, 2000; Pettorino, 2004) che in una certa misura è connessa con il testo del messaggio: la struttura sillabica e la classe di fonemi presenti nella catena fonica incidono sulla durata della sequenza e quindi sulla velocità di articolazione. Se il numero di stringhe considerate per il calcolo della velocità di articolazione è considerevole, si riuscirà a compensare l'incidenza del testo e si otterrà un valore di riferimento medio della una velocità di articolazione del parlante. Nel caso di materiale ridotto, condizione frequente per perizie forensi, la stima della velocità media di articolazione può venire influenzata in maniera significativa dalla tipologia delle sillabe presenti nelle catene fonetiche utilizzate per misurare la velocità di articolazione. Per compensare la componente testo-dipendente si è pensato di utilizzare una procedura basata sulla realizzazione di una medesima composizione sillabica eseguita un sintetizzatore vocale standard.

⁵ A questo proposito si ringrazia per la collaborazione il Capitano Davide Zavattaro del RACIS di Roma.

La velocità di articolazione di una voce sintetica progettata per l'italiano presenta una variabilità dipendente solamente dalla struttura fonologica delle catene, ma insensibile alla variabilità inter-parlante. È pertanto possibile compensare il valore della velocità di articolazione calcolata su un determinato testo, utilizzando la velocità di articolazione dello stesso testo pronunciato da un sistema di sintesi della voce.

	AR	aF0	aF1	aF2	aF3	eF0	eF1	eF2	eF3
AR	1	0,06	0,04	-0,03	0,03	0,02	0,13	-0,17	0,04
		iF0	iF1	iF2	iF3	oF0	oF1	oF2	oF3
		-0,07	0,23	-0,38	-0,02	0,03	0,18	0,07	0,06

Fig. 2 Correlazione tra la velocità di articolazione (AR: Articulation Rate) e i parametri spettrali frequenza fondamentale e frequenze formanti delle vocali.

4. LA PROCEDURA DI COMPENSAZIONE

La nuova variabile compensata ($VA_c = VA - VA_{sint}$) ha caratteristiche statistiche differenziate rispetto alla velocità di articolazione.

Il modello statistico della nuova variabile si ottiene da un'analisi della varianza del dato (ANOVA) che ne scompone la variabilità nella quota assegnata al sintetizzatore, dipendente unicamente dal testo, e nella quota assegnata al parlante che è la parte propriamente aleatoria della velocità di articolazione.

La varianza del parametro compensato va dunque assunta come effettiva misura della variabilità intra-parlante. Con i dati attualmente disponibili la quota di varianza attribuita al parlante indipendentemente dal testo (%W) risulta pari al 71% della variabilità totale riferita alla singola sillaba ($W = 5,88$; $W_c = \%W * W = 4,19$). La varianza inter-parlante non si deve viceversa ritenere modificata, poiché il suo valore atteso è determinato dal valore atteso delle medie dei singoli parlanti che, a differenza delle stime campionarie, non è sensibile al testo.

Con il tipo di sintetizzatore utilizzato in questo lavoro il valore medio campionario della velocità di articolazione della voce sintetica (Media VA_{sint}) è pari a 6,92 sillabe/s mentre il valore medio dello scarto tra voce naturale e voce sintetica è stimato pari a $\Delta = +0,44$ sillabe/s. Si ricorda che il valore campionario del centroide della VA ricavato dall'attuale database del discorso naturale è pari a \bar{VA} di 7,17 sillabe/s.

Per l'analisi della varianza è significativo lo studio della correlazione lineare del parametro normalizzato. La matrice di Tabella 2 dimostra innanzitutto che la correlazione tra variabile compensata e variabile sintetica è statisticamente insignificante ($\rho = -0,21$). La prima, secondo il nostro modello è sensibile solo al parlante, la seconda solo al testo. Il 39% di informazione contenuta nella VA naturale risulta legata a VA_{sint} , quindi al testo. La nuova variabile compensata perde del tutto la correlazione con l'informazione legata al testo e migliora la sua capacità di interpretare la dinamica intra-parlante ($\rho = 0,82$).

Per effetto della varianza intra-parlante ridotta, viene ad aumentare la precisione di questo parametro nel processo di identificazione statistica del parlante.

Correlazioni	VA	VA _c	VA _{sint}
VA	1		
VA _c	.82	1	
VA _{sint}	.39	-0,21	1

Fig. 3 Correlazione lineare del parametro normalizzato.

5. RISULTATI

Al fine di verificare il contributo che il sistema di identificazione proposto fornisce al riconoscimento del parlante, si è proceduto alla definizione di un sistema di analisi statistica dei dati ed è stato effettuato un test di identificazione.

Il sistema SPREAD/AR riceve in input i dati delle misurazioni di VA (valore medio e numero delle sillabe), oppure, se è disponibile un sintetizzatore, le misurazioni di VA_c, relative ad un parlante incognito (voce anonima) ed a un parlante di riferimento (voce saggio), come mostrato nel *display* di comando del programma (figura 4).

Il test di identificazione del parlante viene eseguito dal sistema con tutti i principali algoritmi sviluppati in letteratura e di maggior utilizzo nella pratica di laboratorio.

Il programma esegue i test di identificazione a partire dal modello statistico qui presentato e dai dati che costituiscono il database di riferimento per la popolazione adulta maschile italiana. Come mostra la figura 4 i test possibili sono quattro, il Test di Fisher, il test di Rossi; il test di Bayes di minimo rischio; e il test di verosimiglianza.

Tutti i test forniscono una stima del Rapporto di verosimiglianza (*Likelihood Ratio*).

TEST GUIDATO DI IDENTIFICAZIONE			
VARIABILE: ARTICULATION RATE			
Anonimo:	Saggio:		
VARIABILI DI INPUT			
Centroide della voce Anonima	6.30	Fattore α del test di Fisher %	1%
Numero delle sillabe della voce Anonima	144	Fattore Q del test di Bayes	1.0
Centroide del Saggio fonico	6.65	Probabilità di colpevolezza a priori	50%
Numero delle sillabe del Saggio	119	Likelihood ratio del test sulle formanti	1.00
Varianza campionaria del Saggio	4.00		

Figura 4 Display del programma SPREAD AR (input).

Il rapporto di verosimiglianza consente, mediante la formula di Bayes, di esprimere, nota la probabilità a priori che l'anonimo e il parlante del saggio siano la medesima persona, la probabilità a test eseguito (a-posteriori), per semplice moltiplicazione.

Se altri test sono stati eseguiti, come ad esempio quello sulle formanti, si dovrà procedere alla moltiplicazione di tutti i rapporti di verosimiglianza, posto che i relativi test possano considerarsi statisticamente indipendenti:

$$P_{post} = LRAR * LR1 * LR2 * \dots * LRK * P_{prior}$$

Le figura 5 mostra il display del programma SPREAD AR per la parte relativa ai dati di output dei quattro test disponibili.

RISULTATI DEL TEST DI FISHER: $D < D(\alpha)$			
TEST DI PAOLONI	SI	DISTANZA A-S	1.47
LIKELIHOOD RATIO DEL TEST	2.64	Probabilità di falsa identificazione	0.3750
PROBABILITY ODDS A POSTERIORI	2.64	Probabilità di falsa ricezione	0.01
PROBABILITA' DI COLPEVOLEZZA A POSTERIORI	72.53%	Potere risolutivo	2.67
RISULTATI DEL TEST DI ROSSI (LIKELIHOOD: $LR > 1$)			
TEST di ROSSI	SI	DISTANZA A-S	2.00
LIKELIHOOD RATIO DEL TEST	3.36	Distanza di intersezione ($LR=1$)	3.28
PROBABILITY ODDS A POSTERIORI	3.36	Probabilità di falsa identificazione	0.2614
PROBABILITA' DI COLPEVOLEZZA A POSTERIORI	77.05%	Probabilità di falsa ricezione	0.12
RISULTATI DEL TEST DI BAYES (MINIMUM RISK: $D < D(Q)$)			
TEST di FEDERICO	SI	DISTANZA A-S	1.47
LIKELIHOOD RATIO DEL TEST	4.08	$D(R)$	1.33
PROBABILITY ODDS A POSTERIORI	4.08	Probabilità di falsa identificazione	0.1742
PROBABILITA' DI COLPEVOLEZZA A POSTERIORI	80.30%	Probabilità di falsa ricezione	0.29
RISULTATI DELL'ANALISI DEL DATO (LIKELIHOOD RATIO)			
IDENTIFICAZIONE	SI	DISTANZA A-S	1.47
LIKELIHOOD RATIO	2.67		
PROBABILITY ODDS A POSTERIORI	2.67		
PROBABILITA' DI COLPEVOLEZZA A POSTERIORI	72.73%		

Fig. 5 Display del programma SPREAD AR (output).

La figura 6 mostra le funzioni di verosimiglianza stimate dal programma rispettivamente per l'appartenenza dell'anonimo al saggio ovvero per l'appartenenza alla popolazione di riferimento.

Il programma consente l'esecuzione del test anche sulle velocità di articolazione compensate. Il valore medio atteso, centroide della popolazione, ed il valore atteso della variabilità inter-parlante non cambiano con la compensazione. Tuttavia il sintetizzatore ha una sua velocità di articolazione, evidentemente diversa da quella della popolazione di riferimento. Per la correzione di questo *bias* si procede al modo seguente:

- si eseguono le misure, stringa per stringa, della velocità di articolazione differenziale ΔVA tra voce naturale e sintetizzatore per ricavare, tanto per la voce anonima che per la voce del saggio, i differenziali:

$$\Delta VA = VA - VA_{sint}$$

- si inseriscono nel programma i valori corretti delle velocità di articolazione:

$$VA' \text{ (anonimo)} = \text{Media} (VA_{sint} + \Delta VA \text{ anonimo})$$

$$VA' \text{ (saggio)} = \text{Media} (VA_{sint} + \Delta VA \text{ saggio})$$

- Per ottenere il dovuto effetto di riduzione delle varianze si modificano i parametri di input al seguente modo:

$$\text{Numero delle sillabe della voce anonima } na' = na / \%W$$

$$\text{Numero delle sillabe della voce del saggio } ns' = ns / \%W$$

$$\text{Varianza campionaria del saggio } Ws' = Ws * \%W$$

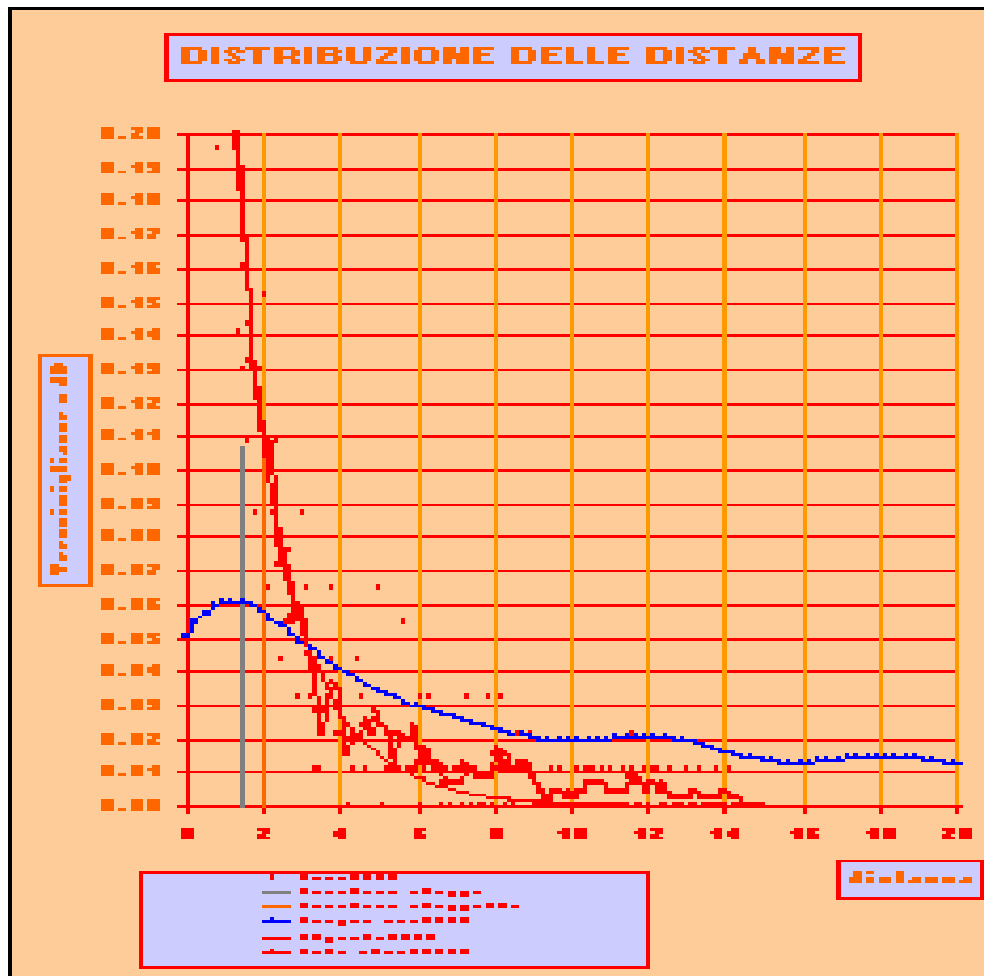


Fig. 6 SPREAD AR le funzioni di verosimiglianza.

6. CONCLUSIONI

Nel presente lavoro, dopo una premessa sull'impiego del parametro velocità di articolazione (VA) nella caratterizzazione del parlante viene illustrata una procedura di compensazione sillabica volta ad eliminare la variabilità della VA relativa ai fattori linguistici, ossia alla composizione delle catene sillabiche. Obiettivo della riduzione della variabilità intraindividuale ottenibile con la compensazione è rendere il parametro più efficace nella caratterizzazione del parlante.

La tecnica di compensazione utilizzata fa uso di un sintetizzatore in grado di fornire una velocità di articolazione costante in modo da normalizzare, stringa per stringa, le durate sulla base della composizione lineare della VA prodotta dal parlante con quella prodotta dal sintetizzatore.

I risultati ottenuti dimostrano la validità della tecnica utilizzata e consentono un significativo miglioramento delle prestazioni ottenibili con questo parametro.

BIBLIOGRAFIA

Cutugno, F. Passaro, G. Petrillo, M. (1999) Sillabificazione fonologica e fonetica, *Atti del XXIII Congresso della Società di Linguistica Italiana*, Napoli.

Federico, A.; Mori, L., Paoloni, A., Zavattaro, D. (in stampa), *The articulation rate and its applications in the field of forensic speaker recognition*.

Federico A., 2005, "SPREAD AR personal identification procedure"; Titolo ultimo Rapporto a Paoloni

Giannini, A. (2000) Range di variabilità della velocità di articolazione in italiano, *Atti del XXVIII Convegno Nazionale della Associazione italiana di acustica*, Trani

Künzel, H.J. (1997) Some general phonetic and forensic aspects of speaking tempo, *Forensic linguistics. The international journal of speech, language and the law*, vol. 4, n. 1

Magno Caldognetto, E., Vagges, K. (1991a) Le pause quali indici diagnostici per lo stile del parlato spontaneo, *Atti delle 2e giornate del Gruppo di Fonetica Sperimentale*, Calabria

Magno Caldognetto, E. Vagges, K. (1991b) Indici di fluenza, tipologia e distribuzione delle pause nel parlato spontaneo, *Atti del XIX Convegno nazionale della Associazione Italiana di Acustica*, Napoli

Magno Caldognetto, E., Ferrero, F., Zmarich, C. (1996) Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza, *Atti delle 7e giornate del Gruppo di Fonetica Sperimentale*, Napoli

Mori, L. & Paoloni, A. (2004) *Sulla sociolinguistica forense: la costituzione di corpora vocali per l'analisi della velocità di articolazione*. In Costituzione, gestione e restauro di corpora vocali (De Dominicis, A./Mori, L./Stefani, M.), Roma: Esagrafica