L'INSIEME È LA SOMMA DELLE PARTI CONSIDERAZIONI PRELIMINARI SU DATI ARTICOLATORI DELLA LINGUA OTTENUTI TRAMITE ULTRASUONI ED ARTICULOGRAFO ELETTROMAGNETICO 3D

Mirko Grimaldi
Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento (Lecce)
mirko.grimaldi@ateneo.unile.it; grimaldi.mirko@alice.it

1 SOMMARIO

Lo studio articolatorio del linguaggio è fortemente condizionato dalle caratteristiche fisiologiche degli organi dell'apparato fonatorio da un lato, e dalle caratteristiche tecniche della strumentazione attualmente disponibile dall'altro. Tuttavia negli ultimi anni i dati ottenuti grazie ad alcune di queste tecnologie hanno consentito di rivoluzionare il modo di vedere il tratto vocale. Informazioni che prima si potevano solo inferire dal segnale acustico sono ora accessibili direttamente.

Più in particolare le tecniche di immagine in uso per monitorare i movimenti della lingua hanno portato ad ipotizzare una deformazione più complessa di questo organo. Ricorrendo anche a ricostruzioni tridimensionali della superficie della lingua, sono state avanzate alcune ipotesi che vedono la lingua come un insieme di segmenti coronali controllati in modo indipendente: i movimenti della lingua sarebbero cioè controllati dalla coordinazione sinergica di *segmenti muscolari funzionali*. Tali segmenti sembrerebbero essere organizzati ortogonalmente lungo l'asse longitudinale del tratto vocale e sarebbero composti da un sistema muscolare multiplo.

In questo lavoro, partendo da alcuni cenni di anatomia e fisiologia della lingua, discuteremo di due delle tecnologie più promettenti per lo studio articolatorio del linguaggio: gli Ultrasuoni (US) e l'Articulografo Elettromagnetico (EMA). Queste due tecniche, utilizzabili presso il CRIL, sono state sincronizzate, insieme ad un Elettroglottografo, per le esigenze sperimentali del progetto europeo CONTACT permettendoci di ottenere dati articolatori della lingua diversificati. Benché gli obiettivi del progetto CONTACT non fossero direttamente indirizzati allo studio articolatorio di determinate caratteristiche fonetico-fonologiche, i dati acquisiti ci offrono ugualmente la possibilità di fare riflessioni preliminari su possibili protocolli specificamente dedicati allo studio di segmenti o gruppi di segmenti funzionali all'interno delle lingue naturali. Oltre alle immagini ottenute con gli US, e alle posizioni dei sensori EMA sulla lingua, faremo ricorso alla ricostruzione tridimensionale della sua superficie, cercando di correlare le diverse informazioni ottenute.

Benché le potenzialità di queste due tecnologie possono essere sfruttate singolarmente, il perfezionamento del processo di acquisizione sincronizzata di dati EMA ed US offrirà opportunità di analisi articolatorie più potenti indirizzate, per esempio, a tutta la vasta gamma dei processi coarticolatori che caratterizza le varietà romanze di area italiana, già rivelatisi di enorme interesse per le teorie fonologiche contemporanee, e che ora – insieme ad altri livelli del linguaggio ancora poco esplorati – potrebbero essere guardati sotto una nuova luce e meglio interpretati anche in un prospettiva multidisciplinare.

2 LO STUDIO ARTICOLATORIO DEL LINGUAGGIO: PROBLEMI E PROSPETTIVE MULTIDISCIPLINARI

Mentre il segnale acustico prodotto dai movimenti dell'apparato fonatorio assume la forma di un'onda sinusoidale continua, permettendo così un'agevole analisi delle caratteristiche acustiche del linguaggio, lo studio degli articolatori che tale segnale producono si presenta irto di innumerevoli difficoltà. Gli articolatori che compongono il tratto vocale, infatti, non si muovono tutti alla stessa velocità, sono strutturati in modo differenziato, e presentano dimensioni, forma e capacità di movimento diverse a seconda se siano formati da tessuti molli (lingua, labbra, velo) o da tessuti duri (mandibola, palato). Le peculiarità fisiologiche degli articolatori, peraltro, incidono sulla velocità con cui essi si muovono, per cui uno strumento adeguato a catturare le risposte motorie della mandibola non si rivela utile per analizzare i movimenti della lingua. Allo stesso modo articolatori che sono visibili ad una analisi superficiale, come le labbra, sono molto più facili da monitorare rispetto a strutture collocate in profondità nella cavità orale, come per esempio il velo.

Ma più di ogni altro fattore, è l'interazione fra gli articolatori a generare serie complicazioni per lo studio articolatorio del linguaggio. Alcuni gesti articolatori sembrano molto più correlati di altri, per cui diventa complicato distinguere il contributo di un articolatore rispetto agli altri. Come osserva Stone (1997), se per esempio il movimento della mandibola è correlato con i movimenti della parte anteriore della lingua (in particolare della punta), tale correlazione viene progressivamente a mancare con i movimenti che avvengono posteriormente.

Un altro caso esemplificativo è quello della /u/ dell'inglese americano studiato da Perkell *et alii*, 1993 con l'Articulografia Elettromagnetica: è stato infatti notato che nella produzione ripetuta di questo fonema – caratterizzato da una leggera restrizione prodotta dall'innalzamento della lingua nella regione velo-palatale e dall'arrotondando delle labbra – non persiste una correlazione sistematica fra i movimenti della lingua e quelli delle labbra. Gli articolatori in questione, quindi, sembrerebbero muoversi attraverso uno scambio motorio reciproco. Gli autori ipotizzano che, data la reciproca variazione di due articolatori controllabili indipendentemente, alla base della realizzazione della /u/ vi sia un goal acustico/uditivo piuttosto che un goal gestuale (motorio).[1]

Ci sono poi anche problematiche legate agli strumenti attualmente a disposizione, per cui se l'elettropalatografo permette di determinare in modo eccellente i contatti della lingua con il palato, allo stato attuale non abbiamo nessuno strumento che, per quanto flessibile, individui i punti della lingua dove questi contatti avvengono (anche perché uno strumento del genere limiterebbe l'espansione, la contrazione, la flessione e la torsione della superficie della lingua). Infine – benché negli ultimi anni siano stati fatti alcuni passi avanti – risulta ancora molto problematico individuare la correlazione fra i complessi movimenti dei muscoli orofacciali e la produzione dei suoni delle lingua naturali (cf. Hardcastle, 1999; Jou *et alii*, 2006).

Per questi motivi gli strumenti per studiare il livello articolatorio del linguaggio devono essere non invasivi, in grado di adattarsi a una determinata superficie (come l'elettropalatografo), avere dimensioni ridottissime per essere posizionati su superfici molli (come i sensori per l'articulografo elettromagnetico e i sensori per l'elettromiografia), o capaci di monitorare i movimenti della lingua dall'esterno (come le diverse tecniche di immagine). A causa di tutte queste difficoltà per la maggior parte degli studiosi risulta più

[1] Dati equivalenti sono stati ottenuti per la /r/ da Guenther *et alii*, 1999.

agevole analizzare il livello acustico del linguaggio e inferire dai dati ottenuti gli eventi fisiologici ad essi correlati.

Tuttavia negli ultimi anni gli studi articolatori sono stati protagonisti di uno sviluppo importante, producendo ricadute significative anche a livello dei modelli teorici. Alcune tecniche di immagine a disposizione hanno rivoluzionato il modo di vedere il tratto vocale e hanno permesso di avere informazioni anche su strutture profonde come quelle della laringe. Nello stesso tempo le tecniche articulografiche si sono sempre più evolute e hanno trasformato le idee circa un fenomeno universale come quello della coarticolazione, mostrando relazioni fra gli articolatori a cui prima era riservata solo una speculazione teorica (cf. Hoole *et alii* 1993; Hardcastle *et alii*, 1996; Hardcastle & Nigel, 1999; Hoole & Nguyen, in press; Cho, 2002; Recanses 2002).

Fra le tecniche d'immagine più diffuse bisogna menzionare i *Raggi-X*, la *Xeroradiografia*, la *Tomografia*, la *Risonanza Magnetica*, e gli *Ultrasuoni* (US), mentre fra le tecniche articulografiche, vanno citati l'*Articulografia Elettromagnetica* (EMA) e i vari sistemi *Optotrak* (per approfondimenti si veda Stone, 1997). In questa costellazione di tecniche strumentali vanno ovviamente inseriti le già menzionate *Elettropalatografia* ed *Elettromiografia*, mentre un discorso a parte meriterebbero le tecniche di indagine dell'attività laringea, come *Elettroglottografia* e l'*Analisi Stroboscopica* (cf. Hirose, 1997).

Grazie a queste nuove possibilità tecnologiche anno visto la luce una gran quantità di ricerche sul controllo motorio del parlato e sui meccanismi alla base della produzione dei suoni delle lingue naturali: scopo principale è l'identificazione delle invarianti motorie, ovvero di quegli elementi che stanno alla base della programmazione motoria del parlato. Per fare ciò è diventato essenziale rispondere alla seguente domanda, così formulata da Perkell (2007: 6): "What is the *task space,* or the domain of the fundamental control parameters?"

Uno degli approcci più incisivi a riguardo si fonda sulla constatazione, ben nota, che ogni parlante produce stringhe sonore sotto forma di segnale acustico e che tali stringhe, attraverso il sistema uditivo, possono essere trasformate in strutture intelligibili da parte di un ascoltatore: "These acoustic cues consist mainly of time-varying patterns of formant frequencies for vowels and glides, and noise bursts, silent intervals, aspiration and frication noises, and rapid formant transitions for consonants" (Perkell 2007: 6). Le proprietà di tale segnale sono determinate da parametri che possono essere osservati in diversi domini, come per esempio:

- nei livelli della tensione muscolare
- nei movimenti degli articolatori
- nei cambiamenti funzionali che avvengono all'interno del tratto vocale
- negli eventi aerodinamici.

Ogni possibile combinazione di questi parametri può ipoteticamente rappresentare una variabile importante del controllo motorio del linguaggio.

Tale approccio combinato con metodologie di ricerca neurocognitive ha prodotto numerose evidenze a supporto dell'idea che nella programmazione articolatoria fonemica siano pienamente coinvolti i domini uditivi e somatosensoriali. Proprio da questo filone di ricerca è nato uno dei più promettenti modelli neurocognitivi della pianificazione motoria denominato DIVA. Si tratta di un sistema basato sulle reti neurali che impara a controllare i movimenti di un apparato vocale virtuale per produrre e percepire i suoni delle lingue naturali. In questo modello i goal fonemici sono codificati sotto forma di proiezioni (mappature) neurali dalla corteccia premotoria alla corteccia sensoriale: tali mappature

delineano regioni corticali e sottocorticali (come il cervelletto) all'interno di aree uditivo-temporali e somatosensoriali-temporali. Il modello utilizza la codifica degli errori del feedback uditivo per apprendere e sviluppare corrette proiezioni motorie. Una volta che il sistema linguistico acquisito diventa fluente, suoni, sillabe e parole possono essere decodificate in modo autonomo senza ricorre al feedback uditivo (cf. Guenther *et alii* 2006).

Un tentativo ancora più pionieristico di combinare dati articolatori con approcci neurocognitivi è quello condotto all'interno del progetto europeo CONTACT. Obiettivo del progetto CONTACT è l'identificazioni delle invarianti motorie e neurali che sono contemporaneamente coinvolte durante lo sviluppo della percezione e produzione del linguaggio e che, secondo la teoria dei neuroni specchio, sarebbero le stesse interessate nei processi di manipolazione (cf. Rizzolatti & Craighero 2004). L'esistenza di tali invarianti neurali verrà testata attraverso un sistema di apprendimento artificiale (basato sul modello delle reti neurali), ovvero un baby-robot, che cercherà, partendo dall'elaborazione di dati articolatori grezzi, di dati motori e dati audio-visivi, di sviluppare autonomamente capacità sensorimotorie diversificate.

Per l'acquisizione dei dati articolatori il progetto CONTACT ha fatto ricorso al sistema US ed EMA 3D, insieme all'Elettroglottografo presenti al CRIL. Questi strumenti (vedi 5) sono stati sincronizzati, consentendo l'acquisizione simultanea di informazioni relative al contorno sagittale della lingua e al tracciato prodotto da un set di sensori collocati sulla superficie della stessa.[2]

In questa sede ci concentreremo esclusivamente su due di queste tecniche strumentali: gli US e l'EMA 3D. Trattandosi di protocolli sperimentali pionieristici e non propriamente finalizzati allo studio diretto delle caratteristiche articolatorie del linguaggio, i dati che discuteremo oltre hanno solo la funzione di illustrare le potenzialità di questi strumenti per lo studio articolatorio del linguaggio. Tale limitazione è anche dovuta alla scelta, ipoteticamente plausibile per il progetto CONTACT, di acquisire i dati US senza stabilizzare la testa dei parlanti. Dal momento che, come vedremo più avanti in 5.1, gli US non riproducono le immagini delle strutture ossee o di altre strutture fisse, è assolutamente necessaria la stabilizzazione della testa del parlante e della sonda perché si abbiano dei parametri di riferimento fissi del tracciato sagittale della lingua da analizzare in studi di tipo articolatorio (cf. Stone 2005)[3].

3 LA LINGUA E IL SUO CONTROLLO MUSCOLARE: IPOTESI RECENTI

L'osservazione della complessa deformazione della lingua durante la produzione del parlato, dovuta (fra le altre cose) al già menzionato fenomeno della coarticolazione, solleva un problema rilevante: quali sono i meccanismi che controllano un tessuto muscolare altamente deformabile per produrre gesti motori così fini, essenziali per la realizzazione dei suoni delle lingue naturali?

I classici lavori sulla coarticolazione (Ohman, 1967, Mermelstein, 1973, Hardcastle, 1976, Hardcastle, 1985) hanno portato ad ipotizzare una suddivisione fra la punta e il corpo della lingua che produrrebbero dei movimenti quasi indipendenti in posizione avanzata o

[2] Per maggiori dettagli e informazioni inerenti anche i dati elettroglottografici si veda Grimaldi et *alii*, 2007, Grimaldi et *alii*, 2008.

[3] Al CRIL è stato di recente realizzato un sistema in legno, che non provocando perturbazioni del campo magnetico del cubo EMA (cf. 5.2), immobilizza la testa dei soggetti durante l'acquisizione sincronizzata di dati US ed EMA.

arretrata, e verso l'alto o verso il basso. Questa suddivisione è stata anche assunta sia dai modelli geometrici della lingua (Coker & Fujimura, 1965; Mermelstein, 1973), che da quelli della sintesi articolatoria (Rubin, 1981). I modelli basati sull'analisi fattoriale hanno invece individuato due o tre fattori basilari che, combinati fra loro, permettono di catturare la forma completa del contorno della lingua, in special modo per quanto concerne le vocali (cf. Harshman, Ladefoged & Goldstein, 1977; Maeda, 1989; Stone, Goldstein & Zhang, 1997; Slud, Stone, Smith & Goldstein, 2002).

Tuttavia, grazie alle recenti tecniche di immagine – che hanno permesso di visualizzare piani multipli della lingua – è possibile ipotizzare che la deformazione di questo organo è sicuramente più complessa. Questi dati sono poi stati già usati per ricostruzioni tridimensionali della superficie della lingua, che mostrano deformazioni molto più complesse di quelle presupposte da un modello basato esclusivamente sulle funzionalità del corpo e della punta della lingua (cf. Wilhelms-Tricarico, 1995; Stone & Lundberg, 1996; Yang & Stone, 2002).

Ipotesi ancora più recenti tendono a considerare la lingua come un insieme di segmenti coronali controllati in modo indipendente (Stone, Epstein & Iskarous, 2004): i movimenti della lingua sarebbero cioè controllati dalla coordinazione sinergica di *segmenti muscolari funzionali*. Tali segmenti sono organizzati ortogonalmente lungo l'asse longitudinale del tratto vocale e sono composti da un sistema muscolare multiplo.

A supporto di questa ipotesi gli stessi autori evidenziano che alcune ricerche nel campo della neuroanatomia hanno dimostrato che la lingua è caratterizzata da un sistema di innervazione molto complesso. Alcuni studi hanno messo in evidenza che il nucleo ipoglosso (il complesso di nervi che irradia la lingua) è composto di circa 13.000 motoneuroni (Wozniak & Young, 1969; Atsumi & Miyatake, 1987; O'Kusky & Norman, 1995). Per esempio ogni fibra del muscolo genioglosso contiene un proprio motoneurone (Mu & Sanders, 1999; Takemoto, 2001): dal momento che ogni motoneurone è potenzialmente un'unità di controllo indipendente, il genioglosso potrebbe avere segmenti di controllo multipli (cf. Miyawaki, 1975). Per queste ragioni la neuroanatomia della lingua fornisce evidenze per ipotizzare un controllo selettivo della lingua, opposto a una visione della lingua come un blocco unico.

Partendo da questi presupposti alcuni gruppi di ricerca stanno attualmente verificando la possibilità che la lingua sia suddivisa in segmenti funzionali correlati alla sua conformazione fisiologica (e in parte controllati anche in modo indipendente), piuttosto che vedere l'organo suddiviso in blocchi muscolari che si muovono nel tratto vocale lungo gli assi verticale ed orizzontale. L'idea di base è che questi segmenti sarebbero controllati o in modo indipendente o si aggregherebbero in unità più ampie per creare determinate strutture articolatorie funzionali all'interno di una determinata lingua naturale (cf. Stone, Epstein, Iskarous, 2004).

4 CENNI DI ANATOMIA E FISIOLOGIA DELLA LINGUA

Caratteristica principale della lingua è quella di essere un organo muscolare che non ha struttura cartilaginea od ossea, e per questo motivo riesce a mantenere un volume costante quando i suoi muscoli si contraggono. Attraverso una selettiva contrazione di determinati muscoli la lingua può creare un supporto rigido che permette la contrazione di certe fibre muscolari, generando così movimenti selettivi mirati (Takemoto, 2001).

I muscoli della lingua sono suddivisi in *intrinseci* ed *estrinseci* (vd. Figure 1-3). I muscoli intrinseci hanno origine e sono inseriti all'interno della lingua stessa, formando una matrice tridimensionale di fibre motorie così suddivisa (cf. Kent 1997: 171-203):

- *muscolo superiore longitudinale* (SL): si colloca immediatamente sotto la superficie della lingua, e consente il suo accorciamento e la curvatura verso l'alto della lamina.
- *muscolo inferiore longitudinale* (IL): si trova al di sotto del muscolo superiore longitudinale e consente di abbassare la punta della lingua.
- *muscolo trasverso e verticale* (TV): il primo va da lato a lato della lingua, mentre il secondo si estende inferolateralmente dal dorso della lingua: entrambe consentono l'abbassamento e l'appiattimento della superficie linguale.

In sintesi, questo complesso muscolare è responsabile dei movimenti fini della lingua ed è coinvolto nei più delicati adattamenti necessari alla lingua per assumere forme diverse.

I muscoli estrinseci connettono la lingua alle strutture vicine, come la mandibola, l'osso ioide, il palato, l'osso temporale e la faringe e si possono schematicamente descrivere come segue:

- *muscolo genioglosso* (GG): ha origine nella spina mentale della mandibola e s'inserisce nell'osso ioide. Si tratta di un complesso differenziato di fibre che comprende la maggior parte della massa della lingua e, a seconda di quali fibre si contraggono, produce la protrusione della punta della lingua, oppure movimenti in avanti e verso l'alto.
- *muscolo stiloglosso* (SG): ha forma allungata, origina dall'apice del processo stiloideo (osso temporale) e dalla parte alta del legamento stilomandibolare e da qui si porta in basso, in avanti e medialmente per inserirsi nell'apice e nel setto linguale. Contraendosi muovono la lingua superiormente e dorsalmente.
- *muscolo ioglosso* (IG): origina dalla parte laterale della faccia superiore dell'osso ioide e dal grande corno, da qui si porta in alto e medialmente per inserirsi sulle facce laterali del setto linguale fino all'apice. Contraendosi ritrae e abbassa il corpo della lingua.
- *muscolo genioioide* e *mioioide* (GI) e (MI): originano dall'osso ioide e si estendono verso la parte anteriore della mandibola. Si trovano alla base della bocca e permettono l'elevamento dell'osso ioide e la lingua.

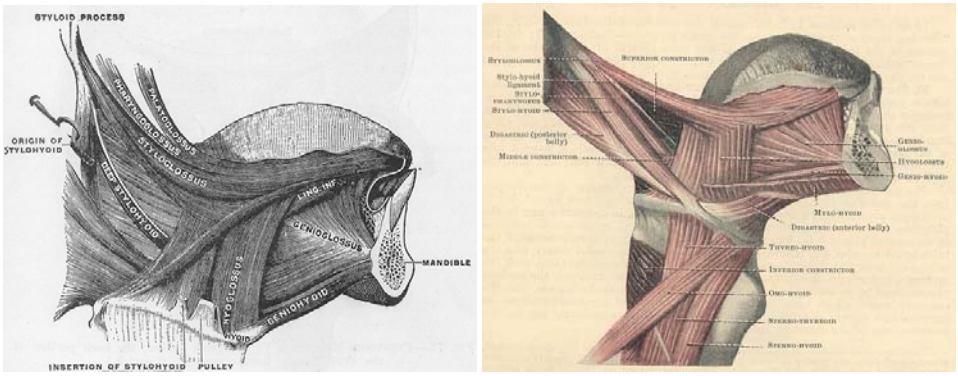

Figure 1-2: muscoli intrinseci ed estrinseci della lingua. Immagini adattate da Gray (1918).

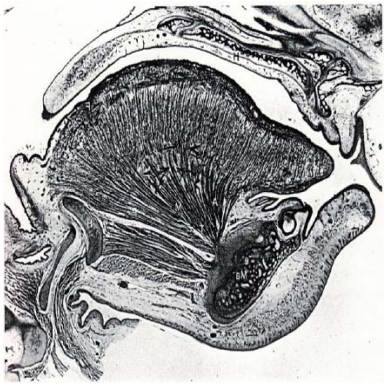

Figura 3: immagine sagittale della lingua (adattata da Stone, Epstein, Iskarous 2004).

5 IL SISTEMA AD ULTRASUONI E L'ARTICULOGRAFO ELETTROMAGNETICO (EMA)

5.1 Il sistema ad ultrasuoni

Gli ultrasuoni per monitorare e studiare i movimenti della lingua stanno diventando un sistema sempre più accessibile e diffuso (cf. Stone 2005). Benché l'utilizzo dei classici ecografi in un laboratorio di linguistica sia ancora limitato dal costo eccessivo, sono da poco stati immessi sul mercato sistemi portatili che consentono ugualmente buone analisi ecografiche. Aspetti della produzione del linguaggio, precedentemente dedotti da dati parziali o indiretti, possono ora essere osservati direttamente. Dal momento che gli ultrasuoni non sono invasivi, forniscono un segnale abbastanza agevolmente interpretabile, e sono per giunta diventati portatili, si prestano molto bene sia per l'uso prettamente clinico che per vari aspetti della ricerca sperimentale. Negli ultimi anni si è perciò iniziato a riflettere su come impiegarli anche nelle indagini geolinguistiche sul campo, o in settori teorico-applicativi come lo studio articolatorio dei processi di acquisizione della seconda lingua (cf. Gick *et alii*, 2002, Gick, 2008).

Un sistema ad ultrasuoni emette onde sonore ad alta frequenza (2-14 MHz) attraverso un trasduttore, o sonda (in genere convessa), che contiene cristalli piezoelettrici stimolati da tensione elettrica. Quando la sonda è collocata contro la pelle del mento, gli ultrasuoni attraverso i tessuti molli (come quelli della lingua), e vengono riflessi quando raggiungono un alto gradiente di impedenza acustica (tipicamente quella generata dall'interfaccia lingua-aria o lingua-palato). I cristalli piezoelettrici emettono in modo sequenziale nel tempo le onde sonore e ricevono gli echi riflessi: in questo modo è possibile catturare sezioni di tessuto piuttosto che un singolo punto. Gli echi sono quindi processati da un computer per essere trasmessi in forma di immagini continue sul sistema video dell'ecografo, oppure possono essere registrati per ulteriori analisi (Hedrick *et alii* 1995).

Siccome, come abbiamo detto, gli ultrasuoni sono riflessi dall'interfaccia lingua-aria o lingua-palato è possibile visualizzare solo la lingua e non, per esempio, il palato, la mandibola, o le pareti posteriori della faringe. A seconda se la sonda venga collocata in modo perpendicolare od orizzontale sotto il mento si può, rispettivamente, ottenere l'immagine sagittale e coronale della lingua (vd. Figure 4 e 5). L'area più luminosa rappresenta la lingua del soggetto: più luminoso è il dettaglio, maggiore è il gradiente di densità del tessuto e, di conseguenza, l'eco acquisita. La lamina della lingua è chiaramente

visibile in quanto presenta un alto gradiente di densità. Tuttavia la mandibola e l'osso ioide creano un'"ombra acustica" in entrambe i margini dell'immagine, perché rifrangono l'ultrasuono prima che raggiunga la superficie della lingua: in questo modo possono essere oscurate parti della punta (soprattutto quando si trova in posizione elevata) o della radice della lingua.

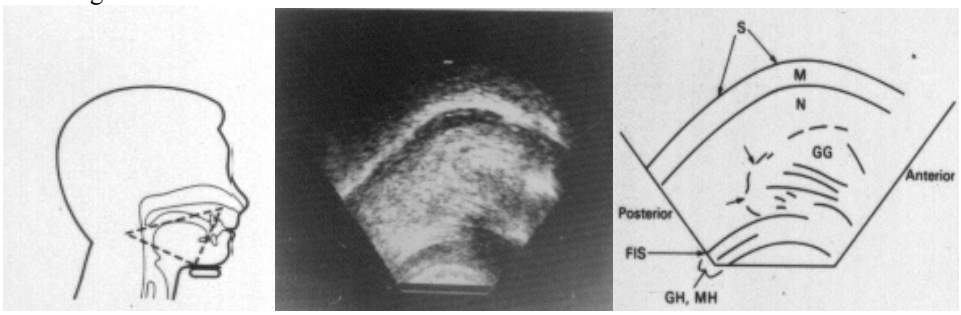

Figura 4: collocazione della sonda a sinistra, immagine sagittale della lingua al centro e sua rappresentazione schematica a destra adattata da Stone (2005): S = Tongue Surface; M = mucosa; N = Superior, Inferiorior logitudinalis, and Vertical, Horizontal network of fibres; GG = Genioglossus; FIS = Floor Intermuscular septum; GH = Geniohioid; MH: Miohioid.

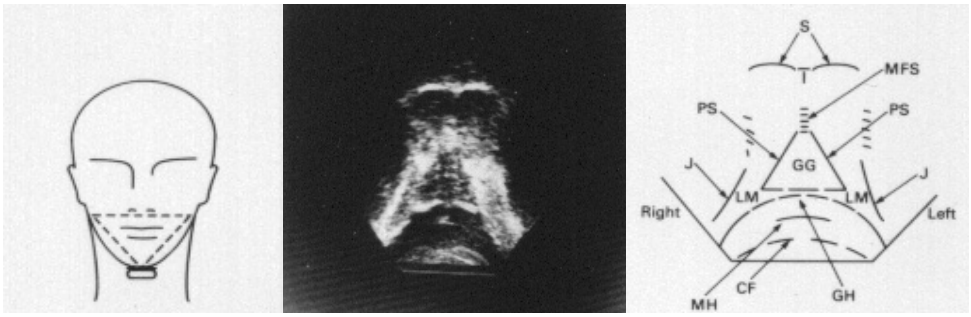

Figura 5: collocazione della sonda a sinistra, immagine coronale della lingua al centro e sua rappresentazione schematica a destra adattata da Stone (2005): S = Tongue Surface; M = mucosa; N = Superior, Inferiorior logitudinalis, and Vertical, Horizontal network of fibres; GG = Genioglossus; FIS = Floor Intermuscular septum; GH = Geniohioid; MH: Miohioid. MFS = Median Fibrous Septum, LM = Lateral Muscles, J = Jaw Inner Aspect, PS = Paramedian Septum, CF = Cervical Fascia.

I dati discussi in questo lavoro sono stati acquisiti con un sistema Aplio XV, costruito da Toshiba Medical System corp., che consente anche l'esportazione delle immagini ecografiche sotto forma di un flusso video continuo, ad una frequenza di 25 immagini al secondo, grazie all'apposito uscita S-Video. Il flusso video è stato acquisito in maniera sincrona con il segnale acustico per mezzo di una scheda esterna di acquisizione, e quindi registrato in tempo reale su un PC dedicato allo scopo (per maggiori dettagli cf. Grimaldi *et alii* 2007).

5.2 L'articulografo elettromagnetico 3D

Come il sistema ad ultrasuoni, anche l'articulografo elettromagnetico (EMA, ElectroMagnetic Articulograph) permette di ottenere importantissime informazioni cinematiche della lingua. Tuttavia sino a poco tempo fa questa tecnica era limitata da severe restrizioni imposte ai movimenti dei soggetti analizzati, compromettendo così la naturalezza delle loro produzioni. L'introduzione di un sistema di rilevamento tridimensionale all'interno di un cubo ha permesso di risolvere questi problemi. In particolare, il modello AG500, costruito da Carstens Medizinelektronik GmbH (Germania), consente di tracciare la posizione di un set di sensori. Tali sensori possono essere incollati su lingua, labbra e denti incisivi, nonché su altri punti presi come riferimento al fine di compensare gli eventuali movimenti della testa compiuti dal soggetto durante l'esperimento.

L'EMA è in grado di determinare le coordinate cartesiane x, y e z, nonché azimuth ed elevazione di un massimo di 12 sensori di campo magnetico, ad una frequenza di campionamento pari a 200 Hz. I sensori di misura sono costituiti da piccole bobine a singolo asse: altre 6 bobine, posizionate in modo da formare un riferimento tridimensionale, emettono altrettanti campi magnetici a differenti e ben note frequenze, tra i 7500 e i 13750 Hz. L'idea principale che sta dietro il processo di determinazione delle coordinate spaziali dei sensori è che la distanza fra i sensori e le bobine di riferimento sia determinata dai valori assoluti delle correnti indotte. Il moto di un sensore e la direzione vettoriale di un sensore può essere così calcolato partendo dalle sei bobine di riferimento usando un metodo iterativo (Zierdt & Carstens, 1995).

Un limite dell'EMA è rappresentato dalla sua sensibilità ad oggetti che permettono la conduzione di elettricità o di campi magnetici. Se uno o più di questi oggetti sono collocati nelle vicinanze delle bobine, il segnale registrato viene distorto. L'EMA deve essere calibrato prima di ogni registrazione, in modo da sintonizzare le sei bobine di riferimento con i dodici sensori (o meno, a seconda del numero di dati richiesti dal protocollo sperimentale). I sensori sono collocati in un apposito caricatore installato su un disco (Circal) in alto al centro del cubo EMA e mosso da un motore elettrico. Il motore, grazie a un software sviluppato da Carstens Medizinelektronik GmbH, muove circolarmente ogni sensore sino a che non ne viene individuata la posizione precisa all'interno del campo magnetico e le corrispondenti ampiezze di corrente elettrica non sono state acquisite (cf. Tavella 2007).

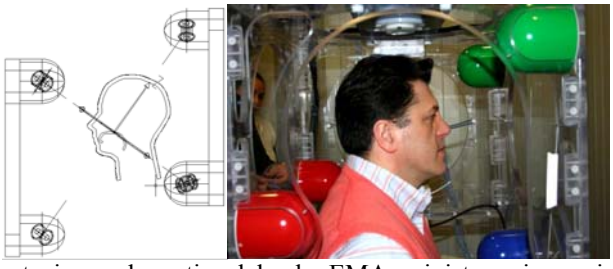

Figura 6: rappresentazione schematica del cubo EMA a sinistra, e immagine di una sessione di prova al CRIL a destra. Sono visibili cinque delle sei bobine di riferimento: due rosse, due verdi e una blu (la sesta, blu, si trova proprio dietro la testa del soggetto in cima la cubo). In alto sulla testa del soggetto è visibile il disco (bianco) su cui, nella fase della calibratura, è inserito il caricatore che contiene i sensori. Immagini adattate da Zierdt *et ali*, 1999 e da Tavella, 2007.

Figura 7: immagine in primo piano del disco e del caricatore con quattro sensori a sinistra, e di cinque sensori EMA incollati sulla lingua di un soggetto prima di un esperimento a destra. Evidenziata dall'ellissi in rosso la collocazione dei sensori da cui sono stati tratti i dati discussi nel paragrafo in 6 (cf. Figure 6-14). Immagini adattate da Tavella, 2007.

Durante la sessione di registrazione, le correnti alternate indotte nei sensori dal campo magnetico generato dalle bobine di riferimento sono filtrate in frequenza, digitalizzate e trasmesse in tempo reale ad un PC di controllo attraverso una connessione Ethernet. Nel PC un software apposito, incluso in dotazione con l'AG500, registra su hard disk i valori ricevuti, rendendoli disponibili per la successiva fase di calcolo delle traiettorie seguite dai sensori mobili. Infine, un meccanismo di sincronizzazione (Sybox Opto-4) è connesso con l'EMA, permettendo la registrazione parallela di dati articolatori e del segnale vocale corrispondente (cf. Tavella 2007: 49-52).

6 PRELIMINARI PER UN'ANALISI ARTICOLATORIA DEI FONEMI DELL'ITALIANO

Pur con i limiti evidenziati in 2, i dati raccolti all'interno del progetto europeo CONTACT consentono di fare alcune riflessioni preliminari sulle potenzialità che gli strumenti sofisticati descritti in 5 offrono per lo studio articolatorio delle lingue naturali.[4] Tali limiti ci obbligano a prendere in considerazione le produzioni di un solo parlante dell'italiano salentino, visto che la scelta di non immobilizzare la testa dei soggetti registrati per il progetto CONTACT non rende possibile il confronto fra le stesse produzioni di parlanti diversi.[5] Tuttavia i dati di cui discuteremo fra poco, considerati alla stregua di

[4] Un altro limite è rappresentato dal fatto che i contesti segmentali del corpus non sono controllati, come in genere si richiede per l'osservazione sistematica di parametri sia acustici sia articolatori rilevanti.

[5] Sono stati acquisti i dati articolatori di parlato letto da 9 soggetti (6 di sesso maschile e 3 di sesso femminile), di età compresa fra i 23 e i 30 anni, tutti parlanti della varietà leccese di italiano. I soggetti hanno letto per tre volte 74 parole e 68 pseudo-parole, realizzate sia in forma dichiarativa che interrogativa. Parole e pseudo-parole sono state scelte in modo da rappresentare tutti i fonemi consonantici e vocalici attestati in italiano (esclusa l'opposizione /e/-/ɛ/ ed /o/-/ɔ/) in condizioni prosodiche paragonabili. Per questo motivo, le parole bersaglio hanno principalmente la prima sillaba accentata (ad esempio, /ˊmatto, ˊnome, ˊstrada/), il cui attacco è rappresentato da una delle consonanti bersaglio ed è seguito da diverse vocali (ad esempio, /ˊmatto, ˊmuffa, ˊmoro/). Inoltre, per una maggiore rappresentatività del corpus, sono state anche inserite parole caratterizzate da una diversa posizione dell'accento (ad esempio, /matˊtone, paˊpa/). Le pseudo-parole, invece, sono monosillabi scelti in modo da rappresentare tutte le consonanti in posizione iniziale, seguite dalle vocali /a, u, i/ (ad esempio, /ˊna, ˊnu, ˊni, ˊʎa, ˊʎu, ˊʎi/).

quelli provenienti da un esperimento pilota, si possono rilevare una buona base di partenza per calibrare in futuro protocolli meglio mirati allo studio di segmenti o gruppi di segmenti funzionali all'interno dei sistemi linguistici.

Per questo scopo mi pare utile provare ad analizzare le parti funzionali della lingua principalmente coinvolte nella realizzazione di un segmento fonologico complesso. Un buon esempio può essere rappresentato da un segmento fonologico affricato seguito da una vocale anteriore alta, come /tʃi/ (vedi Figura 8), prodotto da un parlante di sesso femminile. Come è abbondantemente noto:

- i suoni affricati sono realizzati da una prima fase occlusiva seguita da una breve fase fricativa (con un restringimento del tratto vocale).
- la /i/ è realizzata da un innalzamento ed avanzamento della lingua.

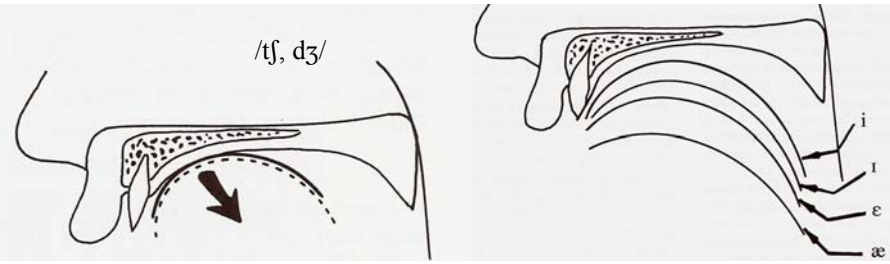

Figura 8: sulla sinistra rappresentazione schematica dell'articolazione di un segmento affricato. La freccia rappresenta il movimento che va dall'ostruzione (fase occlusiva) al restringimento del tratto vocale (fase fricativa). Sulla destra rappresentazione schematica della configurazione articolatoria relativa ad alcune vocali anteriori, fra cui /i/. Immagini adattate da Kent, 1997.

Utilizzando l'acquisizione sincronizzata di dati US e dati EMA, possiamo osservare le diverse fasi articolatorie attraverso cui la lingua dà forma a questo gruppo di fonemi, tentando, per quanto possibile, di focalizzare l'attenzione anche sui segmenti anatomici funzionali coinvolti.

Per visualizzare i risultati delle procedure di acquisizione dei dati, è stato implementato un agevole software di analisi. Il software è attualmente in fase di sviluppo da Francesco Sigona al CRIL e si basa sull'idea di integrare in un unico ambiente altri software già esistenti, specializzati ed ottimizzati nel compiere operazioni specifiche. Al momento l'ambiente si compone di PRAAT (realizzato da Boersma and Weenink, University of Amsterdaam), Edgetrak (Li, *et alii*, 2005), il noto software open source Mplayer ed una interfaccia grafica (Graphic user interface), basata su Matlab e battezzata provvisoriamente soltanto GUI, in grado di interagire, entro certi limiti, con i sopra citati software, anche grazie alle funzionalità messe a disposizione dal linguaggio di scripting del sistema operativo e all'interfaccia a linea di comando di PRAAT e Mplayer (Cf. Grimaldi *et alii* 2007, Grimaldi *et alii* 2008 per maggiori dettagli).

Grazie a un algoritmo, anch'esso in fase di sperimentazione e perfezionamento presso il CRIL, i dati US a nostra disposizione possono essere correlati ai tre sensori EMA collocati sul piano sagittale della lingua e associati al suo contorno estratto dalle immagini US. come rappresentato nelle Figure 6-14 (la punta della lingua si trova nella parte sinistra di ogni figura, mentre la radice si trova sulla parte destra):

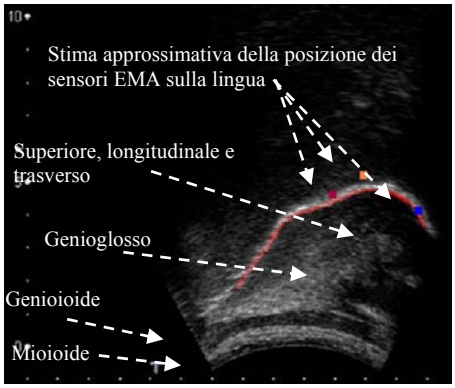

Figura 6: immagine sagittale del frame corrispondente alla fase di occlusione del segmento /tʃ/. La linea rossa individua il contorno della lingua. I quadrati colorati bordò, arancione e blu indicano la posizione approssimativa dei sensori EMA sulla lingua, rispettivamente sul dorso, sulla lamina e sulla punta.

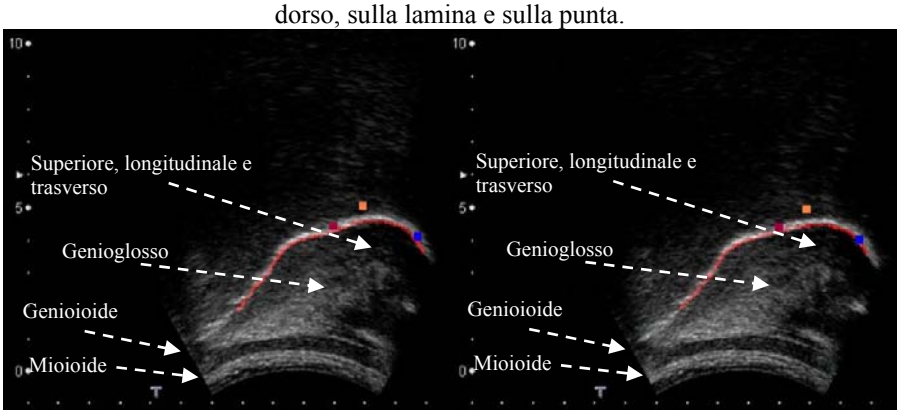

Figure 7-8: immagine sagittale del frame corrispondente alla fase di rilascio sulla sinistra ed alla fase di frizione del fonema /tʃ/ sulla destra.

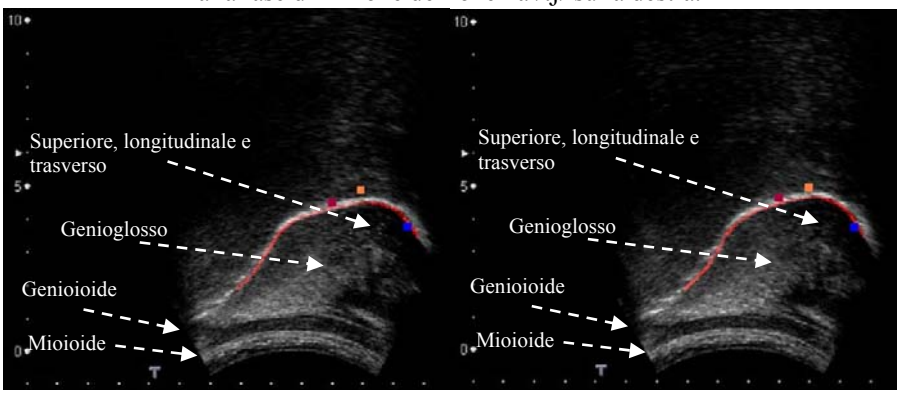

Figure 9-10: immagine sagittale del frame corrispondente alla fase interlocutoria di preparazione del processo di innalzamento nella realizzazione del fonema /i/ sulla sinistra e alla immediata fase di anteriorizzazione sulla destra.

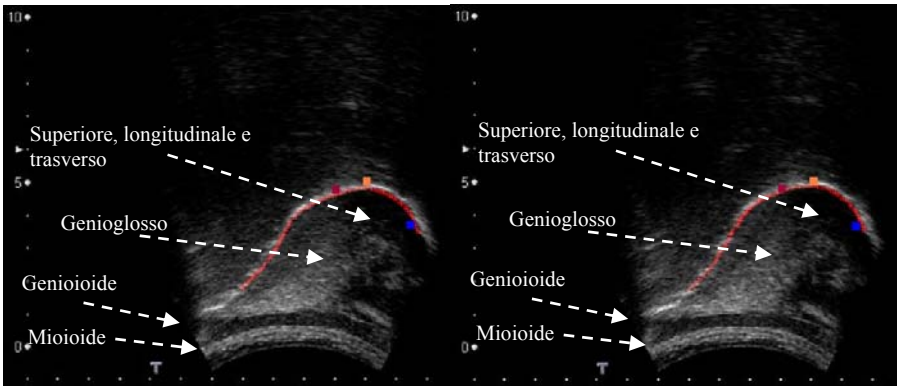

Figure 11-12: da sinistra a destra immagini sagittali dei frame corrispondenti al progressivo innalzamento ed anteriorizzazione della lingua durante la realizzazione del segmento /i/.

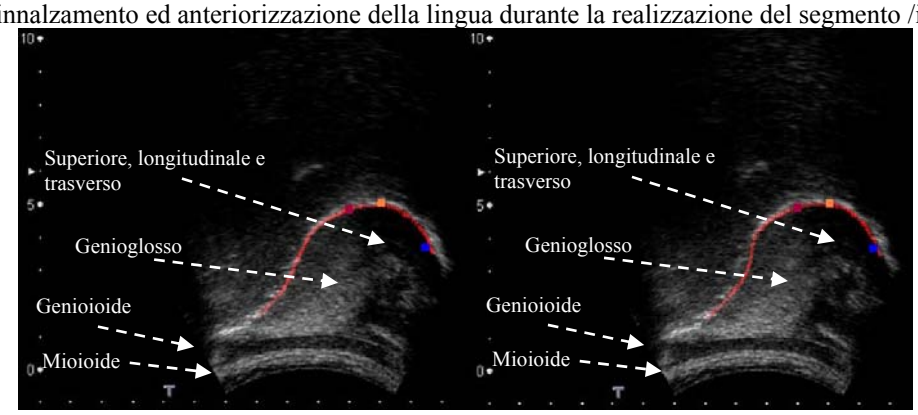

Figure 13-14: da sinistra a destra immagini sagittali dei frame corrispondenti alle fasi finali di innalzamento della lingua durante la realizzazione del segmento /i/

Anche una semplice analisi dei frame in sequenza sembra offrire sufficienti dati per indagini più sistematiche in direzione dell'ipotesi secondo cui i goal articolatori della lingua, nella realizzazione dei segmenti fonologici, siano raggiunti attraverso l'attivazione differenziata di muscoli o di parti di muscoli di volta in volta diversi. In particolare, relativamente al caso preso in esame, si può ipotizzare che la realizzazione del segmento consonantico affricato e del segmento vocalico alto anteriore coinvolga un set di muscoli in modo differenziato. Come si può notare nelle Figure 6-8:

- nelle fasi di occlusione, di rilascio e di frizione sembrerebbe siano completamente coinvolti sia il muscolo genioglosso che i muscoli superiori longitudinali e trasversali.

Dalle Figure 9-14 emergerebbe invece che:

- nella realizzazione di /i/ l'anteriorizzazione precede la fase di innalzamento.

- le fasi di innalzamento e anteriorizzazione della lingua sono prodotte con l'espansione e la compressione solo della parte posteriore del muscolo genioglosso, unitamente ai muscoli superiori longitudinali e trasversali.

- l'espansione e la compressione della sola parte posteriore dei muscoli genioioide e mioioide paiono contribuire incisivamente nella realizzazione del tratto ATR che caratterizza il fonema /i/.

Dunque, limitatamente alla lingua – il principale organo dell'apparato vocale coinvolto nella fase cruciale in cui i suoni assumono la loro 'forma' definitiva – si potrebbe ipotizzare che un segmento fonologico complesso sia il risultato dell'attivazione differenziata e nello stesso tempo coordinata delle fibre di diversi muscoli linguali: in sostanza, il goal articolatorio finale sarebbe la somma di goal articolatori intermedi.

Questo complesso fenomeno può essere riconsiderato da un'altra prospettiva grazie alla possibilità di ottenere una rappresentazione grafica del tracciato dei sensori EMA e dei contorni sagittali della lingua nel tempo, che può essere relativa a un solo segmento fonologico, a una parola intera, o addirittura a una frase. Elaborando i dati US ed EMA precedentemente ottenuti, attraverso i punti x e y di ogni contorno e i valori t del tempo, possiamo infatti ricostruire la superficie della lingua in un ambiente 3D. Nel nostro caso il software GUI è in grado di estrarre automaticamente la rappresentazione 3D della lingua sia dal contorno sagittale ottenuto tramite US sia dal tracciato dei sensori EMA collocati sulla stessa.

Tuttavia le diverse tecniche di acquisizione dei dati articolatori offrono possibilità di rappresentazioni grafiche diverse. Abbiamo visto in 5.1 e 5.2 che il sistema US acquisisce ad una frequenza di 25 Hz (e quindi è in grado di registrare 25 immagini al secondo), mentre l'EMA ad una frequenza di 200 Hz (e quindi è in grado di registrare 200 posizioni al secondo per ogni sensore). Volendo ottenere una rappresentazione 3D di un segmento come /tʃi/, la cui durata si aggira intorno ai 0,6 secondi, risulta evidente che i dati ottenuti con l'EMA ci permettono di rappresentare proficuamente le diverse fasi articolatorie anche di segmenti che hanno breve durata, mentre l'US ci consente solo di rappresentare il segmento per intero.

Correlando con il software GUI il segnale acustico ai dati articolatori forniti dai sensori EMA (cf. Grimaldi *et alii*, 2007) e utilizzando la tecnica dell'interpolazione delle spline, possiamo ottenere rappresentazioni efficaci delle fasi di occlusione, rilascio, frizione, anteriorizzazione e innalzamento relativi alla produzione del segmento /tʃi/. Nelle Figure 15-20, possiamo infatti osservare le rappresentazioni grafiche 3D della lingua derivate dal tracciato dei sensori EMA collocati sul dorso, sulla lamina e sulla punta della lingua.

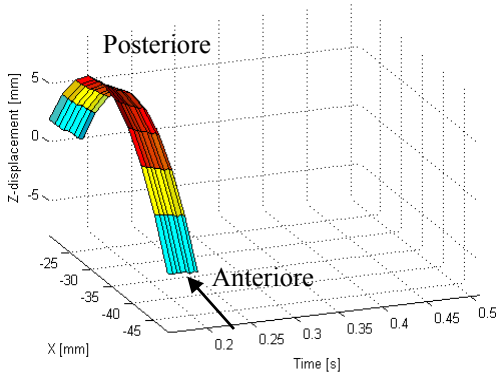

Figura 15: fase di occlusione. La freccia nera indica il probabile punto di contatto della punta della lingua con i denti.

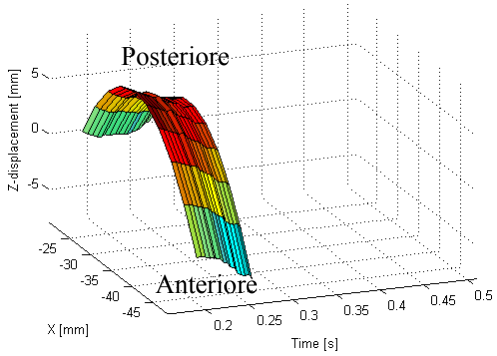

Figura 16: fasi di rilascio e frizione.

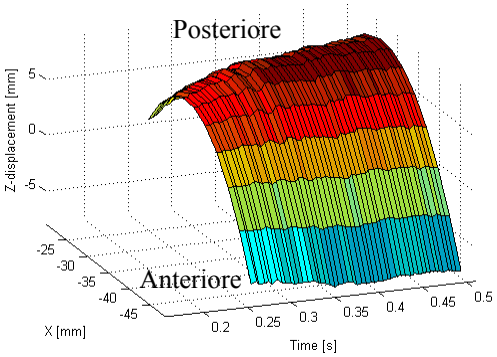

Figura 17: fasi di innalzamento e anteriorizzazione.

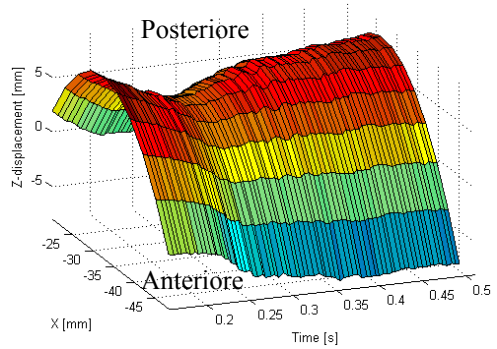

Figura 18: rappresentazione completa delle fasi di occlusione, rilascio, frizione, innalzamento e anteriorizzazione.

Allo stesso modo i contorni sagittali della lingua ricavati dalle immagine US (ed evidenziati nelle Figure 6-14), se proiettati nel tempo, posso essere elaborati per realizzare una rappresentazione 3D, come illustrato nella Figura 19:

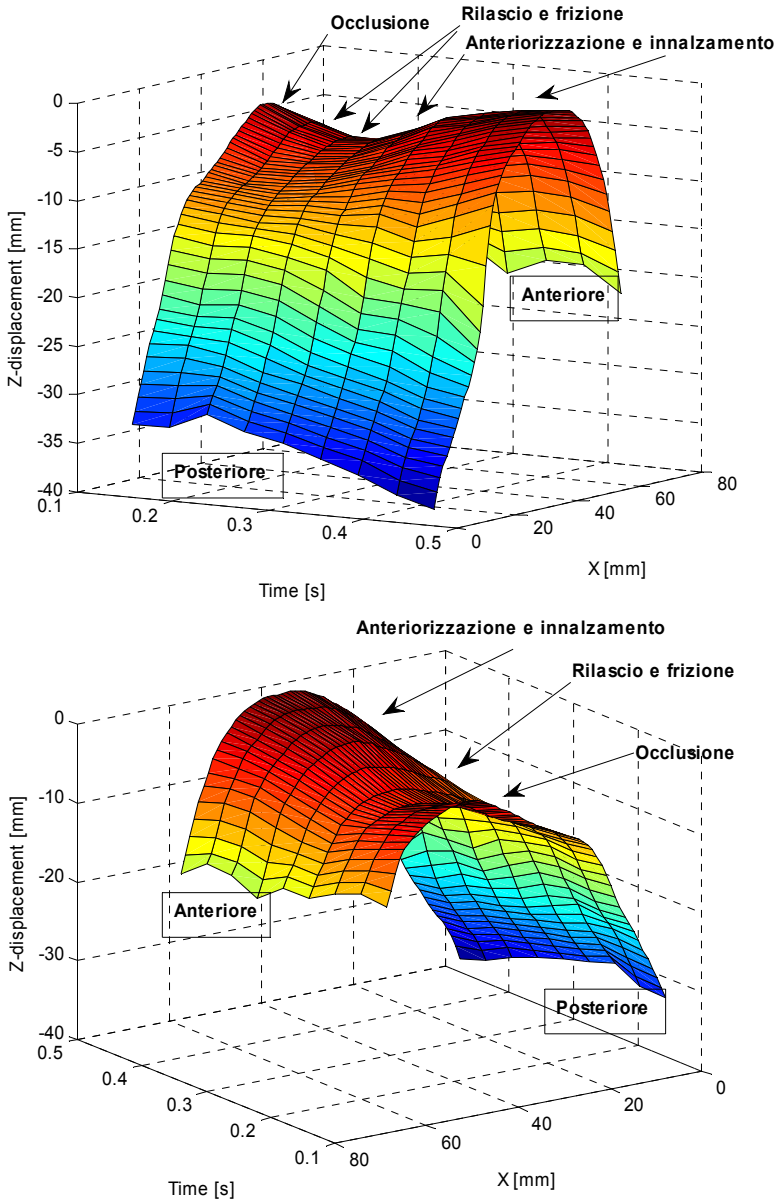

Figura 19: rappresentazione 3D da prospettive diverse dei contorni sagittali della lingua ottenuti tramite sistema US durante la realizzazione del segmento /tʃi/.

Se i dati EMA forniscono maggiori informazioni relativamente alla punta della lingua (che come abbiamo visto in 5.1 gli US non sempre riescono a monitorare), i dati US ci possono offrire informazioni più dettagliate in merito alla radice della lingua, in quanto i sensori EMA non possono essere collocati al di sotto del dorso.

Tentando di comparare, anche solo da un punto di vista impressionistico, le Figure 6-14 con le Figure 15-19 si può evincere come dalle rappresentazioni 3D relative ai dati EMA (e in particolare dalle Figure 15-18), è facile recuperare un particolare dato articolatorio inerente la punta della lingua: ovvero il graduale passaggio dalla posizione più elevata – in cui il contatto con gli incisivi superiori permette di creare la fase di occlusione – alla posizione di contatto con gli incisivi inferiori, attraverso il processo di interiorizzazione, innalzamento e di inarcamento della lingua, che permette il successivo allungamento e abbassamento della punta. Il fenomeno, nonostante i limiti del sistema US nel catturare i movimenti della punta della lingua, è sufficientemente visibile anche nel secondo grafico della Figura 19, che quindi conferma l'osservazione ricavata dai dati EMA.

7 CONCLUSIONI

In conclusione si può dire che mentre il sistema EMA – in quanto capace di decodificare in tempo reale la traiettoria di ogni sensore collocato su un punto della lingua – offre dati più robusti dal punto di vista temporale, il sistema US offre dati più efficaci da un punto di vista spaziale, in quanto l'onda sonora che si diffonde sotto il mento e si riflette cattura la maggior parte del corpo della lingua. Di queste potenzialità ci si può rendere meglio conto ricorrendo alle funzioni del software GUI che permette di assemblare in un unico ambiente i dati ottenuti durante la realizzazione del segmento /tʃi/ dal sistema US, insieme alla rappresentazione 3D di tutti i sensori collocati sulla lingua (cf. Figura 7), aggiungendo le informazioni ottenute dall'Elettroglottografo (EGG), insieme allo spettrogramma correlato e al filmato del soggetto durante l'esperimento, come si può vedere cliccando sulla Figura 21:

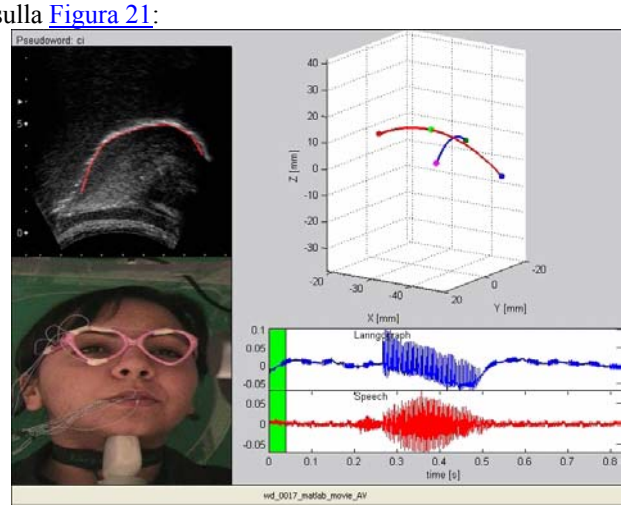

Figura 21: dati US, EMA, EGG insieme allo spettrogramma e al filmato del soggetto si sesso femminile che produce il segmento /tʃi/.

Al di là delle potenzialità delle singole tecnologie ora descritte, il perfezionamento del processo di acquisizione sincronizzata di dati EMA ed US, lo sviluppo e la verifica dell'algoritmo che permette di associare i sensori EMA all'immagine sagittale della lingua, e l'implementazione continua del software GUI, insieme alla stabilizzazione della testa dei soggetti monitorati (attualmente già possibile al CRIL), potrà sicuramente offrire possibilità di analisi articolatorie molto più complesse e mirate di quelle qui superficialmente

abbozzate a solo fine esemplificativo. Penso ovviamente a tutta la vasta gamma di processi coarticolatori che caratterizza le varietà romanze di area italiana, che si sono già rivelati di enorme interesse per le teorie fonologiche contemporanee, e che ora potrebbero essere guardati sotto una nuova luce e meglio interpretati anche in un prospettiva multidisciplinare.

BIBLIOGRAFIA

Cho, T. (2002), The Effects of Prosody on Articulation in English, Routledge, New York, NY.

Gick, B., (2002), The use of ultrasound for linguistic phonetic fieldwork, *Journal of the International Phonetics Association*, 32(2): 113-122.

Gick, B., Bernhardt B., Bacsfalvi P., Wilson I. (2008), Ultrasound imaging applications in second language acquisition, in J. G. Hansen Edwards & M. L. Zampini, *Phonology and Second Language Acquisition*, Benjamins, Amsterdam-Philadelphia, 309-322.

Gray, H., (1825–1861), *Anatomy of the human body*, 20th ed., thoroughly rev. and re-edited by Warren H. Lewis, Philadelphia: Lea & Febiger, 1918.

Grimaldi, M., Gili Fivela, B., Tavella, M., Sigona, F., Fitzpatrick, P., Metta, G., Craighero, L., Fadiga, L., Sandini, G., (2007), Synchronized acquisition of Italian speech articulatory data using ultrasound, 3D-articulograph and laryngograph: First results, in *Ultrafest VII*, New York, September 2007: http://jerome.linguistics.fas.nyu.edu/ultrafest.html.

Grimaldi M., Gili Fivela B., Sigona F., Tavella M., Fitzpatrick P., Metta G., Craighero L., Fadiga L., Sandini G., *New Technologies for Simultaneous Acquisition of Speech Articulatory Data: Ultrasound, 3D Articulograph and Electroglottograph*, in Cristina Delogu, Mauro Falcone (eds.), *LangTech 2008*, Fondazione Ugo Bordoni, Roma: 81-85.

Guenther, F.H., Espy-Wilson, C., Boyce, S.E., Matthies, M.L., Zandipour, M., Perkell, J.S., (1999), Articulatory tradeoffs reduce acoustic variability during American English /r/ production, *Journal of Acoustical Society of America*, 105, 2854-2865.

Guenther, F.H., Ghosh, S.S., Tourville, J.A., (2006), Neural modeling and imaging of the cortical interactions under-lying syllable production, *Brain and Language*, 96, 280-301.

Hardcastle, W. J., (1976), *Physiology of Speech Production*, London, Academic Press.

Hardcastle, W.J., (1985), Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences, *Speech Communication*, Vol. 4, 247-263.

Hardcastle, W. J., (1999), Electromyography, in W. J. Hardcastle & N. Hewlett, *Coarticulation. Theory, Data and Techniques*, Cambridge University Press, 270-283.

Hardcastle William J., Hewlett Nigel (editors), (1999), *Coarticulation, Theory, Data and Techniques*, Cambridge University Press, 1999.

Hardcastle, W. J., Vaxelaire, B., Gibbon, F., Hoole, P., Nguyen, N., EMA / EPG study of lingual coarticulation in /kl/ clusters, in *1st ESCA Tutorial and Research Workshop on Speech Production Modeling*, Autrans, May 1996, 123-131.

Hirose, H., (1997), Investigating the Physiology of Laringeal Structures, in W. J. Hardcastle, J. Laver, *The Handbook of Phonetics Sciences*, Blackwell, Oxford, 11-32.

Hoole, P. & Nguyen, N., Electromagnetic articulography in coarticulation research, in: W.H Hardcastle, Instrumental Studies of Coarticulation. CUP (in press).

Hoole, P., Nguyen-Trong, N. & Hardcastle, W. (1993), A comparative investigation of coarticulation in fricatives: electropalatographic, elctromagnetic and acoustic data, *Language and Speech*, 36:235-260.

Kent R. D., (1997), *The Speech Sciences, Singular Publishing*, Group, San Diego-London.

Jou, S.-C., Maier-Hein L., Schultz T, Waibel A., Articulatory Feature Classification Using Surface Electromyography, in *ICASSP 2006*, I, 605-608.

Li, M., Kambhamettu, C. & Stone, M., (2005a), Automatic Contour Tracking in Ultrasound Images, *International Journal of Clinical Linguistics and Phonetics*, 19 (6-7), 545-554.

Perkell, J. S., (2007), Sensory Goals and Control Mechanisms for Phonemic Articulations, in *ICPhS*, Saarbrücken, 6-10 August 2007, 169-174.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., Jordan, M. I., (1993), Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study, *Journal of Acoustical Society of America*, 2948-2961.

Recasens, D., (2002), An EMA study of VCV coarticulatory direction, *Journal of the Acoustical Society of America*, 111, 2002, pp. 2828-2841.

Rizzolati G., Craighero L., (2004), The Mirror-Neuron System, *Annual Review of Neuroscience*, 27, 169–192.

Stone, M. & Lundberg A., (1996), Three-dimensional tongue surface shapes of English consonants and vowels, *Journal of the Acoustical Society of America*, 99, 3728–3737.

Stone M., Epstein M. A., Iskarous K., (2004), Functional Segments in Tongue Movement, *Clinical Linguistics and Phonetics*, 18, 6-8, 507-521.

Stone, M. (1997), Laboratory Techniques for Investigating Speech Articulation, in W. J. Hardcastle, J. Laver, *The Handbook of Phonetics Sciences*, Blackwell, Oxford, 11-32.

Stone, M., (2005b), A Guide to Analyzing Tongue Motion from Ultrasound Images, Clinical Linguistics and Phonetics, 19, (6-7), 455-502.

Takemoto, H. (2001), Morphological analyses of the human tongue musculature for three-dimensional modeling, Journal of Speech and Language Hearing Research, 44, 95–107.

Tavella, M., (2007), *Simultaneous recording of phono-articulatory parameters during speech production*, unpublished Master Thesis, University of Genova, LiraLab (Laboratory for Integrated Advanced Robotics), 2006-2007.

Wilhelms-Tricarico, R. (1995), Physiological modeling of speech production: Methods for modeling soft-tissue articulators, *Journal of the Acoustical Society of America*, 97, 3085–3098.

Wrench, A. A,. (2007), Advances in EPG palate design, Advances in Speech-Language Pathology, 99, Issue 1, March 2007 , 3 – 12.

Zierdt, A., Carstens, B., Method and a device to determine the spatial arrangement of a directionally sensitive magnetic field sensor, *United States Patent Number*, 5, 453, 687, 1995.

Zierdt, A., Hoole, P., Tillmann, H.G., (1999), Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements", in *Proceedings. ICPhS'99*, San Francisco. August.