Illusione e Scienza nella Fonetica Forense: Una Sintesi

M. Grimaldi, S. d'Apolito, B. Gili Fivela, F. Sigona

Abstract. *Questo articolo presenta una sintesi sulla fonetica forense, focalizzando l'attenzione sul problema della comparazione di registrazioni vocali, nella quale un campione intercettato della voce del reo (anonimo) viene comparato con il campione registrato della voce del sospettato o dei sospettati (saggio). Dopo aver discusso i principi fondamentali della fonetica acustica, il presente lavoro pone l'accento sul perché metodi non scientifici, come quello dell'impronta vocale, non siano più accettati dalla comunità scientifica, che si sta invece orientando verso metodi tecnico-scientifici, implementati da software specializzati: nell'ambito di tali metodi, viene infine illustrato quello basato sull'approccio bayesiano e sul calcolo del rapporto di verosimiglianza (Likelihood ratio) per il confronto delle distribuzioni statistiche delle frequenze formantiche e della frequenza fondamentale.*

Keywords: Forensic voice comparison, Speaker identification, Forensic phonetics, Likelihood ratio framework, Bayesian approach

1. Introduzione

Una questione fonetica di rilevanza forense può essere così formulata: qualcuno può essere riconosciuto in base alle caratteristiche della propria voce oltre ogni ragionevole dubbio? In altre parole, si può essere sicuri che la voce intercettata sia proprio quella del sospettato? Una risposta ragionevole è: dipende dal metodo di comparazione applicato.

Questo contributo ha l'obiettivo di fornire un breve quadro critico di come la fruttuosa integrazione di teorie e metodi nel campo della fonetica acustica con le teorie e i metodi propri dell'ingegneria e dell'informatica possa essere utile nel confronto di registrazioni vocali (in letteratura si utilizzano anche le espressioni Forensic Speaker Recognition, FSR, e Technical Forensic Speaker Identification, TFSI, quest'ultima da riterenersi preferibile). In questa sede ci soffermeremo solo sull'aspetto principale della FSR: ovvero la comparazione della voce (rimandiamo a [3] Jessen 2008 per gli altri aspetti, come il voice profiling e l'analisi dell'identificazione del parlante da parte di vittime e testimoni).

In genere, nella comparazione della voce il parlato registrato della voce anonima viene messo a confronto con il parlato registrato della voce nota. Tutte le parti coinvolte (polizia giudiziaria, giudici e avvocati) vogliono sapere se la voce dell'anonimo appartenga alla voce nota. A seconda del sistema legale in cui ci si trova a operare, intercettazioni telefoniche e/o ambientali oppure registrazioni di interrogatori possono essere utilizzate come evidenza nel caso in cui il sospettato sia poco o per nulla collaborativo. In caso contrario, si può anche ricorrere all'acquisizione di ulteriore materiale audio dal sospettato tramite 'saggio fonico' (opportunamente costruito sulla base del materiale intercettato e con esso coerente). Le registrazioni possono quindi essere messe a confronto rispetto a un'ampia varietà di tratti peculiari della voce e sulla base di metodi differenti.

La comparazione della voce può essere richiesta sia dalla Polizia giudiziaria sia da privati al di là di un dibattimento in Tribunale; ma in genere si rende necessario depositare una perizia scientificamente motivata che sarà utilizzata come evidenza in un processo e che deve essere discussa e difesa oralmente in dibattimento da parte dell'esperto responsabile della perizia ([3: 673]. Dal momento che un processo, dopo tutto, è un evento in cui si 'decide' sulla base di evidenze, un modo ragionevole di porre la questione è: qual è la probabilità che, data l'evidenza delle voci comparate, il parlato registrato della voce dell'anonimo e quello della voce nota appartengano alla stessa persona? (cfr. [7])

Prima di rispondere a questa domanda, è necessario capire in modo sintetico come si può descrivere il segnale vocale sulla base di alcuni principi di fisica acustica.

2. In principio era la voce

Per quanto riguarda le vocali, il parlato può essere definito come un segnale periodico prodotto da tre effetti: (i) il movimento periodico delle corde vocali che genera la frequenza fondamentale (F0) correlata con il tono della voce di ciascun individuo; (ii) il rumore prodotto dalla fonazione; (iii) le modificazioni del flusso d'aria da parte degli articolatori all'interno del cavo orale. Questi tre effetti generano uno spettro di frequenza, la cosiddetta *Struttura Formantica.*

La struttura formantica è caratterizzata da una serie di picchi discreti nello spettro di frequenza che sono il risultato dell'interazione tra la frequenza di vibrazione delle corde vocali e le risonanze che si generano all'interno del tratto vocale del parlante. La frequenza di questi picchi, che corrisponde alle frequenze formantiche, come anche la frequenza relativa tra i picchi, varia in base ai differenti suoni realizzati poiché sono coinvolti differenti articolatori (lingua, denti, palato, labbra, ecc.). La struttura formantica del parlato interagisce con la struttura armonica del parlato (rappresentata da multipli interi della frequenza fondamentale). Le armoniche che sono vicine alla frequenza di risonanza del tratto vocale sono chiamate *Formanti*.

Lo spettrogramma rappresenta le componenti del suono in un grafico a tre dimensioni, in cui il tempo è posto sull'asse delle ascisse, la frequenza sull'asse delle ordinate e l'intensità attraverso il maggiore o il minore annerimento delle

frequenze (oppure attraverso una scala di colori). La frequenza di questi picchi, generalmente espressa in Hz, come anche la frequenza relativa tra i picchi, varia in base ai differenti suoni prodotti. La frequenza più bassa è nota come prima formante (F1) e le formanti successive sono la F2, F3, ecc. Generalmente, le vocali sono classificate considerando i primi due picchi dell'inviluppo spettrale [5] vedi Fig.1. La prima formante è inversamente proporzionale al movimento della lingua nella dimensione verticale (alto/basso), mentre la seconda formante riflette il luogo di articolazione nella dimensione orizzontale (anteriorità/posteriorità) del cavo orale. La F2, insieme con la frequenza della terza formante, può dare utili indicazioni sull'arrotondamento delle labbra [9].

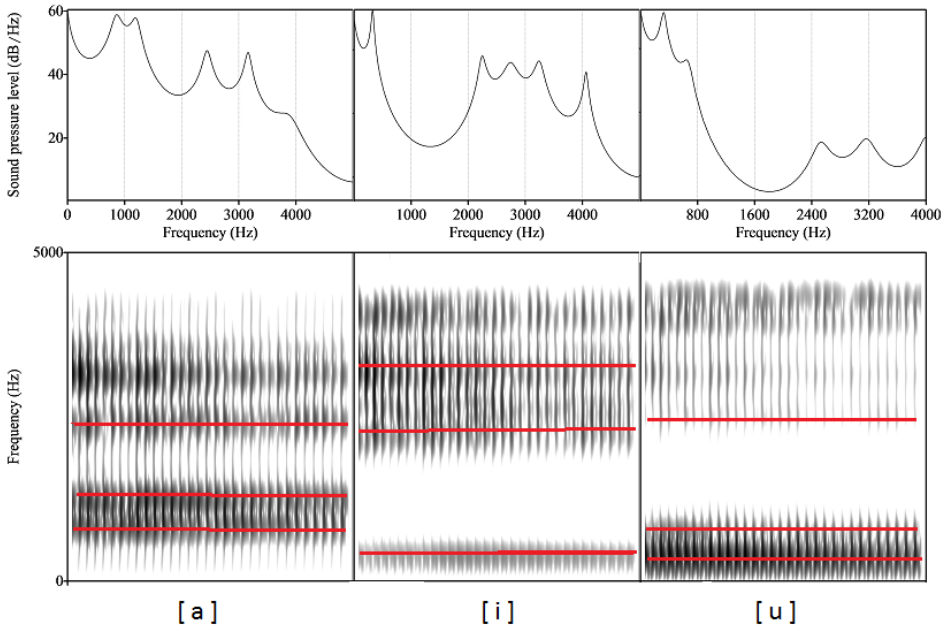

Figura 1 - Inviluppi spettrali (in alto) e spettrogrammi (in basso) delle vocali cardinali [a], [i] e [u] realizzate da un parlante italiano di sesso maschile. Le prime tre formanti sono messe in evidenza dalle linee tratteggiate.

3. L'illusione dell'impronta vocale

Agli inizi degli anni '60 quando ancora i fonetisti non avevano ancora dato un importante contributo nella disciplina, un primo esperimento fu condotto presso i Laboratori Bell [4] in cui si testò se la comparazione visiva degli spettrogrammi poteva essere utile per l'identificazione del parlatore. L'esperimento dimostrò che tale comparazione poteva avere successo. Nel corso del tempo, tuttavia, la maggior parte degli scienziati assunse un atteggiamento scettico sull'affidabilità di questo metodo poiché non era stato sufficientemente validati [8] e in alcuni casi fu respinto in modo completo [2] Bolt et al. [1] hanno criticato il metodo sollevando numerose questioni a riguardo:

1) quando due spettrogrammi sono simili, tale similarità indica che si tratta dello stesso parlante o semplicemente che si tratta della stessa parola pronunciata? 2) le similarità irrilevanti possono fuorviare una giuria composta da persone non esperte?

3) quanto sono costanti i parametri della voce?

4) quanto tali parametri sono caratterizzanti per il soggetto?

5) questi parametri possono essere simulati o camuffati?

Nonostante questo metodo sia stato strenuamente difeso da [10], il 24 luglio del 2007 una risoluzione *dell'Associazione Internazionale di Fonetica Forense e Acustica* (IAFPA) ha definitivamente deliberato che questo metodo è privo di fondamenti scientifici, dichiarando esplicitamente che esso non deve essere utilizzato in ambito forense.

Sfortunatamente, questo metodo risulta ampiamente adottato nei Tribunali italiani, grazie anche al fatto che, ancora oggi, il Codice di Procedura Penale non riconosce la figura del perito in Fonetica Forense ed Acustica (FPA). Contrariamente a quanto accade negli altri paesi stranieri, questo implica che periti e consulenti, nella migliore delle ipotesi, siano ingegneri o tecnici informatici i quali non hanno competenze di linguistica o di fonetica acustica e non sono a conoscenza dei metodi scientifici da utilizzare per il riconoscimento del parlante.

4. Riconoscimento del parlante secondo un metodo scientifico

Come evidenziato in [1], nel moderno approccio alla TFSI l'identificazione del parlante si ispira alla identificazione del DNA, cioè assumendo una prospettiva probabilistica. Occorre, innanzitutto, specificare che la comparazione della voce non avviene sulla base di tutte le proprietà del parlato, ma su solo su determinate peculiarità della voce umana, cioè le prime tre formanti che abbiamo illustrato prima insieme alla frequenza fondamentale delle vocali.

È necessario, inoltre, chiarire un concetto importante: l'esperto forense non deve e non può fornire la probabilità che il parlato registrato dell'anonimo sia stato prodotto dal sospettato. In altre parole, per molteplici ragioni, lo scienziato forense non deve presentare la probabilità di colpevolezza o di non colpevolezza. È compito del giudice giungere a queste probabilità e decidere sulla base di tutte le evidenze forensi (e non) che emergono durante il processo. Allo scienziato forense deve essere solo richiesta la forza dell'evidenza. Al fine di espletare questo compito, lo scienziato forense deve considerare due importanti aspetti: 1) la similarità, cioè stabilire quanto siano simili o differenti i campioni di parlato dell'anonimo e del sospettato rispetto ai parametri di interesse; e 2) la tipicità, cioè stabilire quanto siano tipiche o rare le caratteristiche fonetiche tra i due campioni di parlato rispetto a una popolazione di riferimento. A parità di condizioni, l'evidenza circa l'identità dei due parlanti è più forte tanto più la tipicità è bassa rispetto al caso contrario. Questo approccio si ispira alla teoria Bayesiana e in particolar modo al rapporto di verosimiglianza

- Likelihood Ratio (LR) - [7: 49-54] in cui il rapporto dell'evidenza viene valutato come segue:

$$LR = \frac{P(EH_p)}{P(EH_D)}$$

Il numeratore corrisponde alla probabilità di ottenere una data evidenza E, se i due campioni hanno la stessa origine, mentre il denominatore esprime la probabilità di ottenere una data evidenza se i due campioni hanno una origine differente. Se il LR ha un valore maggiore di 1, maggiore è l'evidenza che i due campioni provengano dallo stesso parlante, mentre se il suo valore è minore di 1 è maggiore l'evidenza che i due campioni provengano da voci differenti. Il numeratore cattura la similarità: se la similarità è alta, la probabilità che la sorgente dei due campioni sia la stessa è anch'essa relativamente alta; se la similarità è bassa, la probabilità che la sorgente dei due campioni sia differente è anch'essa relativamente molto bassa. Il denominatore cattura, invece, l'aspetto della tipicità: se la tipicità è alta, la probabilità che qualcun altro possa aver dato origine alla voce anonima è relativamente alta, mentre se la tipicità è bassa la probabilità che sia stato qualcun altro, piuttosto che il sospettato, è relativamente bassa [3: 682-683].

In base al metodo utilizzato per calcolare il rapporto di verisimiglianza, il numeratore di LR può anche essere espresso come:

$$P(EH_p) = 1 - P.f.rej.$$

Dove P.f.rej indica la probabilità di rifiutare l'identificazione quando la voce dell'anonimo appartiene alla voce nota, mentre il denominatore può essere riscritto come:

$$P(EH_D) = P.f.id$$

Dove P.f.id indica la probabilità di accettare l'identificazione tra la voce dell'anonimo e quella del sospettato quando il campione dell'anonimo non è stato pronunciato dalla voce nota.

Al fine di quantificare la tipicità, è necessario creare o avere accesso ad una popolazione di riferimento basata sulle proprietà del parlato che devono essere utilizzate ai fini della comparazione. Questo è un aspetto fondamentale del LR e che necessità di ulteriori ricerche poiché le banche dati della popolazione di riferimento costruite sulle caratteristiche del parlato sono rare. Ancora più rare sono le banche dati che prendono in considerazione la variazione dialettale. Dal momento che le caratteristiche del parlato possono, in parte, essere determinate da differenze dialettali, queste caratteristiche dovrebbero essere escluse dal punto di vista percettivo e acustico nella comparazione di voci secondo il metodo Bayesiano. Inoltre, una popolazione di riferimento di questo tipo è molto interessante per il *voice profiling*, quando è disponibile solo una

registrazione di una voce anonima senza alcun possibile sospettato. Questo accade spesso durante le prime fasi di investigazione. In una situazione del genere, si può chiedere all'esperto forense di delineare un profilo in base alle caratteristiche della voce e questo può aiutare la polizia a restringere il campo dei possibili sospettati o di trovare il sospettato in base anche ad alcune caratteristiche dialettali.

Il nostro gruppo di ricerca presso il CRIL ha intrapreso un progetto di ricerca su questi aspetti. Si sta sviluppando un software semi-automatico basato sull'approccio Bayesiano del LR insieme alla creazione di una popolazione di riferimento caratterizzata da parametri acustici estratti dalle voci di parlanti di differenti varietà dialettali del Salento (una banca dati in continuo aumento). Nella Fig.2 è riportato uno screenshot del software per il riconoscimento del parlante, in fase di sviluppo al CRIL. L'interfaccia grafica permette di importare le tabelle dei valori formantici, precedentemente elaborate, relative alla voce nota (saggio) e della voce dell'anonimo (anonimo), una banca dati dei campioni registrati, di poter selezionare una qualsiasi combinazione dei parametri che si vuole considerare, i parametri in entrata per il test da eseguire, come anche i grafici relativi alla distribuzione di probabilità delle formanti, calcolata in base a test statistici parametrici multivariati. Il programma può anche esportare un resoconto delle operazioni effettuate da includere nella perizia dell'esperto.

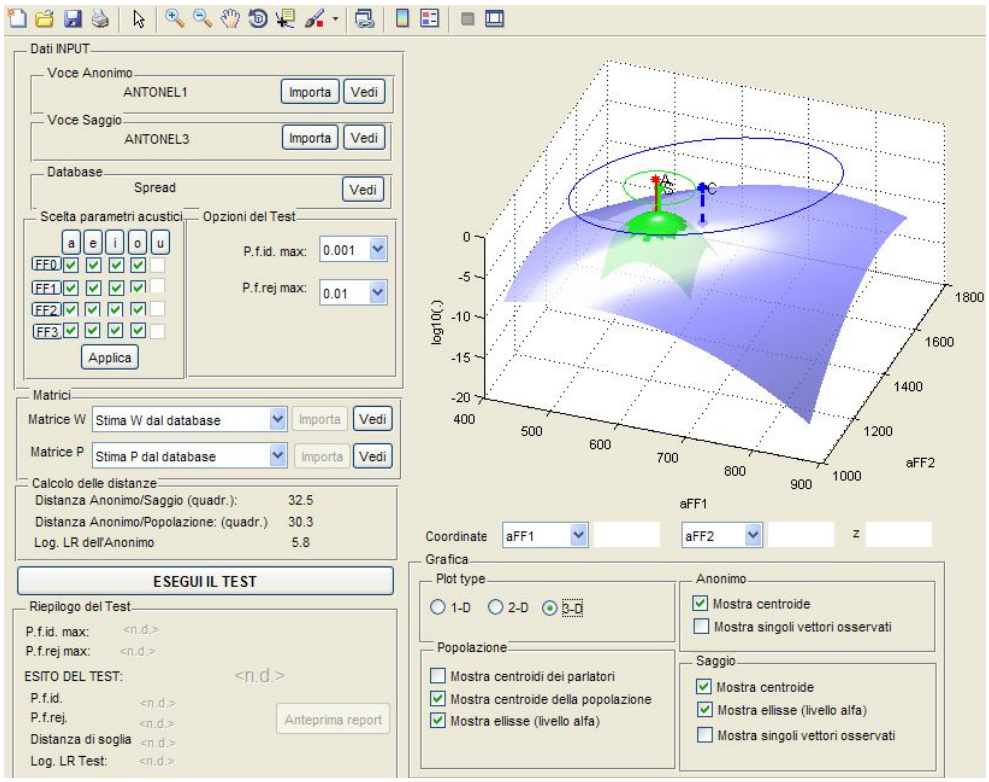

Figura 2: Uno screenshot del software in fase di sviluppo al CRIL

Conclusioni

Il moderno approccio alla TFSI è sempre più consapevole che la comparazione della voce finalizzata al all'identificazione del parlante deve essere eseguita scientificamente adottando il metodo Bayesiano non è solo un modo per comparare similarità e differenze tra determinati parametri acustici di campioni di voci differenti, ma anche il modo di conoscere quanto comuni siano le voci in base ad una popolazione di riferimento. Il ricorso alle conoscenze della fonetica e della linguistica da parte dei modelli ingegneristici ha sicuramente giocato un ruolo importante in questo processo. Ad ogni modo, sarebbe sbagliato dedurre che il LR nell'identificazione del parlante parlante in ambito forense sia ovunque accettato e istituito. Il grado in cui tale approccio è utilizzato, o anche solo compreso (data la sua complessità), differisce da nazione a nazione (cfr. [6: 67-68] per ulteriori dettagli). Allo stesso tempo, non vi è alcun dubbio che, dato l'interesse crescente nella corretta valutazione dell'evidenza relativa all'identificazione forense, ignorare tale approccio è a proprio rischio e pericolo.

Bibliografia

[1] Bolt, R. H. Cooper, F. S., David, E. E., Denes, P. B., Pickett, J. M., Stevens, K. S., Speaker identification by speech spectrograms: some further observations, Journal of the Acoustical Society of America, 54, 2, 1973, 531–53.

[2] Hollien, H., Status report of "voiceprint" identification in the United States, Occasionally, 2, 1977, 29–40.

[3] Jessen, M., Forensic Phonetics, Language and Linguistics Compass 2, 4, 2008, 671–711.

[4] Kersta, L. G., Voiceprint identification, Nature, 196, 1962, pp. 1253–1257.

[5] Peterson, G. E. and Barney, H. L., Control methods used in a study of the vowels, Journal of the Acoustical Society of America, 24, 2, 1952, 175–184.

[6] Rose P., Forensic Speaker Identification, Taylor and Francis, London & New York, 2002.

[7] Rose, P., Forensic speaker recognition at the beginning of the twenty-first century – An overview and a demonstration, Australian Journal of Forensic Sciences, 37, 2, 2005, 4–30.

[8] Stevens, K. N, Carl. E W., Carbonell, J. R. and Woods B., Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material, Journal of the Acoustical Society of America, 44, 1968, 1596–1607.

[9] Stevens, K. N., Acoustic phonetics. Cambridge, MA: The MIT Press, 1998.

[10] Tosi, O.I. (1979) Voice Identification: Theory and Legal Applications. Baltimore: University Park Press.

Biografia

Sonia d'Apolito si è laureata in Lingue e Letterature Moderne Euroamericane presso l'Università del Salento nel 2007 e nel 2012 ha conseguito il titolo di Dottore di Ricerca presso l'Università del Salento. Durante gli anni di dottorato si è interessata allo studio delle caratteristiche acustiche ed articolatorie (movimenti della lingua) nella lingua francese. In particolare, sono stati osservati gli aspetti coarticolatori e fonologici (assimilazione di sonorità e del luogo di articolazione) all'interno di sequenze eterosillabiche di sibilanti realizzate da studenti italofoni di francese L2 e da parlanti nativi. L'obiettivo è stato quello di osservare come gli apprendenti italofoni realizzassero sequenze fonotatticamente marcate nella lingua materna e come la loro produzione si differenziasse da quella dei parlanti nativi. I risultati di questo lavoro sono stati presentati a convegni nazionali e internazionali. Attualmente si interessa di fonetica forense con particolare attenzione alla comparazione di voci.

Email: sonia.dapolito@gmail.com

Barbara Gili Fivela è professore associato di Linguistica Generale e Fonetica e fonologia presso l'Università del Salento, è vicedirettore del Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) e presidente del Corso di Laurea in Scienza e Tecnica della Mediazione Linguistica/Traduzione e Interpretariato della stessa Università. Dal 2010 è anche membro del comitato direttivo dell'Associazione Italiana Scienze della Voce (AISV). Dal 1995, dopo aver svolto attività di ricerca allo CSELT, laboratorio per le telecomunicazioni (oggi NUANCE), ha perfezionato la sua formazione presso la Scuola Normale Superiore di Pisa, specializzandosi nello studio fonetico-fonologico della prosodia; durante il triennio, ha studiato presso il Dipartimento di Linguistica dell'Ohio State University, Columbus (U.S.A.) e svolto attività di ricerca presso l'Universität des Saarlandes, Saarbrücken (Germania), l'IPDS di Kiel (Germania) e l'LPL di Aix-en-Provence (Francia). È autrice di una monografia e di più di ottanta contributi su argomenti di fonetica e fonologia di laboratorio, pubblicati in volumi, atti di convegni e riviste specialistiche nazionali ed internazionali.

Email: barbara.gili@unisalento.it

Mirko Grimaldi è professore associato di Linguistica Generale presso la Facoltà di Lingue e Letterature Straniere dell'Università del Salento, dove insegna anche Psicologia del Linguaggio. Ha ideato e dirige il *Centro di Ricerca Interdisciplinare sul Linguaggio* (CRIL), realizzato grazie a un co-finanziamento della Comunità Europea (PON 2000-2006, Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione). Il CRIL è il luogo ideale per individuare spazi di ricerca di confine, non ancora ben delineati, fra discipline linguistiche, psicologiche, mediche, ingegneristiche, informatiche e fisiche che, pur partendo da presupposti, metodologie e tradizioni diverse, possono dare un contributo per comprendere non solo la fisiologia del linguaggio ma anche e soprattutto l'organizzazione anatomico-funzionale del linguaggio nel cervello. I suoi interessi di ricerca riguardano: (i) la fonetica, la fonologia e la comparazione della voce; (ii) le basi neurofisiologiche dei processi di percezione e produzione del

linguaggio; (iii) i processi acustici, uditivi e neurofisioligici nell'acquisizione della della seconda lingua; (iv) i processi sociolinguistici e pragmatici nella comunicazione mediata dal computer.
Email: mirko.grimaldi@unisalento.it

Francesco Sigona, nato a Bari nel 1973, si è laureato in ingegneria elettronica presso il politecnico di Bari nel 1998 e successivamente abilitato all'esercizio della libera professione. Ha svolto attività di ricerca nel campo della Quality of Service (QoS) per reti Wireless LAN (IEEE 802.11), presso ST Microelectronics (STM). Dal febbraio 2007 é responsabile tecnico del Centro di Ricerca Interdisciplinare sul Linguaggio (C.R.I.L.) dell'Università del Salento, nel quale è impegnato nel supporto alla ricerca nel campo dell'elaborazione numerica di segnali biometrici relativi alla produzione del parlato (speech/kinematics/ imaging), e nello studio di algoritmi di "forensic voice comparison / speaker identification" con relativo sviluppo di applicazioni software.

Email: francesco.sigona@unisalento.it