

PROCURA DELLA REPUBBLICA presso il TRIBUNALE DI PERUGIA

Dr. Giuliano Mignini Sost.

PROCURA DELLA REPUBBLICA
DI FIRENZE
- 2 NOV. 2005

4438

**CONSULENZA TECNICO-FONICA NEL
PROC. PEN. 9144/2001 R.G. N.R. MOD. 21
CONTRO ██████████ FRANCESCO + ALTRI.**


RELAZIONE

PROCURA DELLA REPUBBLICA
di Perugia

DEPOSITATO IN SEGRETERIA

IL 26 OTT 2005

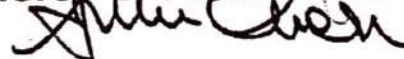
IL SEGRETARIO



I CONSULENTI

Dr.ssa Cinzia Sabatini

Cinzia Sabatini
M.A.s. UPS CC Simone Cesare



INDICE

CAPITOLO 1

1.1	CONFERIMENTO DELL' INCARICO.....	PAG. 3
1.2	QUESITO.....	PAG. 3
1.3	INIZIO DELLE OPERAZIONI TECNICHE.....	PAG. 3
1.4	AUTORIZZAZIONI CONCESSE E RNVIO.....	PAG. 5

CAPITOLO 2 - PREMESSE METODOLOGICHE

2.1	PREMESSA ALLA COMPARAZIONE DI VOCI.....	PAG. 6
2.2	PRODUZIONE DELLA VOCE - MODELLO TEORICO.....	PAG. 14
2.3	METODI UTILIZZATI PER LA COMPARAZIONE.....	PAG. 21
2.3.1	ESAME GENERALE E LINGUISTICO.....	PAG. 21
2.3.2	ESAME OGGETTIVO PARAMETRICO: IDE.M.®.....	PAG. 27

CAPITOLO 3 - ESAMI COMPIUTI

3.1	ANALISI DEL REPERTO.....	PAG. 38
3.2	PRETRATTAMENTO DEL SEGNALE.....	PAG. 44
3.3	ESAME GENERALE LINGUISTICO-FONETICO.....	PAG. 50
3.4	RISPOSTA AI QUESITI.....	PAG. 54
	BIBLIOGRAFIA.....	PAG. 56

ALLEGATI



CAPITOLO 1

1.1 CONFERIMENTO DELL'INCARICO

In data 18 Agosto 2005, in merito al Proc. pen. nr. 9144/2001 R.G. N.R. Mod. 21 nei confronti di Francesco ██████████ + altri, il dr. Mignini – Sostituto Procuratore presso la Procura della Repubblica del Tribunale di Perugia, conferiva agli scriventi Dr.ssa Cinzia Sabatini – operatore demo-antropologico e M.A.s.UPS Cc Simone CESARE - effettivo al Reparto Investigazioni Scientifiche di Roma, incarico di consulenza tecnico-fonica.

4140

1.2 QUESITO

*“ Presa visione della documentazione del Procedimento e ascoltato il CD-R relativo alle conversazioni telefoniche per cui è processo, accertino i C.T.U. quanto descritto nel decreto ”
di nomina datato 24.06.2005.*

Il decreto di nomina datato 19 luglio 2005, riporta i seguenti quesiti:

“ Previo esame delle trascrizioni e della fonìa delle telefonate per cui è processo, accerti il C.T.U. tutte le caratteristiche delle voci degli anonimi interlocutori di Falso Dorotea (in relazione all'avviso ex art. 415 c.p.p.), caratteristiche atte ad identificarli, vale a dire:

1. Sesso ed età degli interlocutori; timbro delle voci;

caratteristiche 'linguistiche' delle voci ed esatta individuazione dell'aerea geografica di provenienza alla luce dell'inflessione dialettale, specie in relazione alle note distintive fonetiche, all'andamento della tonalità e a particolarità lessicali e/o morfologiche;

2. *Comparazione dei risultati raggiunti con le caratteristiche linguistiche generali delle aree di Foligno (PG) e Cannara (PG), specie in relazione agli aspetti di cui all'ultima parte del punto 1;*
3. *Ulteriori particolarità espressive, atte a riferire le voci a particolari ambienti socio-culturali;*
4. *Eventuali anomalie di pronuncia e loro origine;*

Fornisca ogni ulteriore considerazione ritenuta necessaria ai fini dell'accertamento dei fatti. ”

1.3 INIZIO DELLE OPERAZIONI TECNICHE

La data dell'inizio delle operazioni tecniche veniva fissata contestualmente al conferimento dell'incarico con il ritiro del materiale oggetto di accertamento.

1.4 AUTORIZZAZIONI CONCESSE E RINVIO

Per lo svolgimento dell'incarico gli scriventi ottenevano l'autorizzazione a:

- ❖ ritirare il supporto ottico oggetto di verifica;
- ❖ impiegare per gli spostamenti da e per la residenza, in tutte le occasioni inerenti l'incarico demandato, il mezzo proprio;
- ❖ rispondere con relazione scritta circa le operazioni compiute nel termine di 55 giorni. Per la conclusione dell'incarico è stato necessario richiedere una proroga di gg. 10 al termine inizialmente fissato.



4/42

Ⓟ CF

CAPITOLO 2

PREMESSE METODOLOGICHE

2.1 PREMESSA ALLA COMPARAZIONE DI VOCI

La richiesta di comparazione della voce per l'individuazione del parlatore deve considerare che nell'atto della verbalizzazione si propongono essenzialmente due tipi di informazioni: il segnale acustico dovuto alle caratteristiche fisiologiche dell'apparato fonatorio individuale ed il messaggio enunciato, formalizzato secondo uno schema linguistico codificato.

Voce e messaggio sono dunque espressioni diverse di un'unica sorgente: il soggetto parlante.

In termini generali, inoltre, l'oggetto dell'accertamento sarà costituito da conversazioni che necessariamente devono essere state registrate (mediante apparati tecnologicamente più o meno avanzati) su un supporto fisico. Il risultato finale, il reperto, è dunque il frutto di un'operazione di mediazione strumentale e potrebbe, per gli effetti prodotti dagli strumenti stessi, non rappresentare esattamente il segnale originario in tutte le sue caratteristiche. Questi effetti¹, benché possano essere teoricamente controllabili e riproducibili, influenzano comunque la qualità del segnale e le variabili fisiche in esso contenute. Le variabili di tipo linguistico, invece, non sono in generale sensibili alla curva di risposta dei sistemi riproduttivi/trasmissivi ma possono dipendere da altre circostanze quali lo stato emotivo dei parlatori, la natura della conversazione etc...

¹ Dovuti alla non linearità della curva di risposta in frequenza dei componenti del sistema.

Da questi semplici ragionamenti preliminari si deduce che l'esame confrontuale tra voci potrà (e dovrà, per quanto possibile dal tipo di materiale) essere concentrato parimenti su due filoni d'indagine: la branca scientifica strumentale dell'analisi acustica e quella fonetico-linguistica, il tutto perché le informazioni ricavabili dalle due discipline sono tra loro indipendenti e complementari. L'obiettivo è quello di ottenere un quadro, il più descrittivo possibile, di quelle caratteristiche individuali che consentano la creazione di una sorta di un profilo tipo 'carta d'identità del parlatore', capace di essere quanto più oggettiva e completa per le successive comparazioni.

La storia della fonica insegna che una trattazione esclusivamente ingegneristica, che consideri la voce come un'impronta digitale o comunque come un marchio indelebile della persona (come è stato proposto verso la fine degli anni '60, con la realizzazione del sonografo, denominato troppo fiduciosamente *voiceprint*) dopo una breve fase di entusiasmo finì con lo smorzare le iniziali ambizioni. La variabilità, ineliminabile, del segnale vocale proveniente da qualsiasi parlatore rende il confronto tra le impropriamente nominate 'impronte vocali' decisamente diverso e molto più complesso rispetto al confronto tra impronte digitali[2]. Come si vedrà nel seguito l'ambito del confronto vocale si sviluppa su un numero elevato di variabili 'dinamiche', ed è pertanto impossibile effettuare confronti che non richiedano una trattazione statistica; le risposte stesse, quindi, non saranno categoriche ma dovranno contenere aspetti di gestione degli errori.

La difficoltà principale per un approccio scientifico è dato proprio dalla variabilità intrinseca della voce che, sommata alla variabilità delle condizioni di registrazione, pone un problema sopra un altro problema.

Vedremo, tuttavia, per tornare al discorso della valenza duale segnale + messaggio, che esistono caratteristiche che possono essere considerate stabili all'interno di opportuni intervalli di tolleranza e/o invariante rispetto ad alcune condizioni al contorno, alcune per esempio non sono influenzate dal sistema di registrazione, altre non lo sono dal tipo di linguaggio... Poter contare su differenti branche d'indagine risulta quindi non solo un potenziamento del metodo, ma anche un sistema di verifica della stabilità dei risultati finali. Dall'insieme delle considerazioni fin qui riportate nasce spontanea l'esigenza di associare all'esame fisico-acustico anche un esame generale linguistico².

Anche per tutte le caratteristiche linguistiche esiste, comunque, una variabilità intrinseca più o meno accentuata. Alcune peculiarità sono tra esse indipendenti e meritano una trattazione separata. In particolare menzioniamo aspetti:

FONETICI: acquisiti con l'educazione e consolidati nel tempo, caratterizzano le persone per le modalità dei processi di articolazione e collegamento dei suoni. Rappresentano

² Gli esperti dell'Europa del Nord, invece, tradizionalmente fonetisti, sono giunti alla medesima conclusione partendo dalla considerazione che gli esami linguistico-fonetici risultano insufficienti, da soli, a fornire un quadro sufficientemente completo per poter consentire una decisione statisticamente fondata.

[Handwritten initials]

un fattore seminconscio³, sostanzialmente stabile nel parlato spontaneo, e scarsamente variabile nel medio termine. La sovrapposibilità delle mappature delle produzioni fonetiche tra individui è stringente per qualsiasi giudizio di compatibilità;

SEMANTICO-LESSICALI: incidono sull'identità e il significato del messaggio pronunciato in relazione ai suoni emessi o alla scelta di vocaboli operata (parola/significato, secondo un codice semantico), forniscono informazioni di carattere prevalentemente socio-linguistico, con riferimento alla zona geografica di appartenenza ed al livello culturale raggiunto dal soggetto;

PROSODICI: riguardano l'andamento (temporale e intonativo) dell'esposizione frastica, possono essere connessi con il carattere della persona e/o con l'ambiente di sviluppo/lavoro.

Gli aspetti intonazionali sono i più condizionati dallo stato emozionale, tuttavia altri aspetti temporali possono risultare estremamente stabili rispetto alle emozioni e quindi maggiormente utilizzabili.

FONIATRICI: riguardano le modalità di emissione acustica dei suoni dovute alle condizioni morfo-fisiologiche dell'apparato respiratorio.

³ Quindi ragionevolmente robusto anche nei casi di alterazione volontaria della voce.

Rappresentano la categoria meno controllabile dal cervello e più oggettivamente legata alla biometria dell'individuo. Sono i responsabili delle frequenze fonatorie emesse, e rappresentano l'oggetto degli esami strumentali.

Cerchiamo ora di spiegare le cause e le conseguenze del fattore 'intrinseca variabilità': la voce è soggetta a continue alterazioni non solo per l'aumentare dell'età o per l'insorgere di patologie specifiche dell'apparato respiratorio (più precisamente parleremo di variazioni 'di lungo periodo' del sistema Rino-Laringo-Faringeo), ma anche per cause 'di breve periodo' come un'anomala respirazione momentanea, l'assunzione di alcool o per improvvisi sbalzi umorali del parlatore.

Queste variazioni influiscono anche sulle componenti più prettamente fisiche della voce come è stato dimostrato, per esempio, per la frequenza di vibrazione delle corde vocali che dipende anche, per esempio, dall'intensità della voce e dalla tensione delle corde stesse [3].

Da tutto ciò si evince che, a livello *intraparlante*, esiste una doppia variabilità che non può essere trascurata in fase confrontuale. Si sottolinea 'doppia' perché vi è la componente a breve termine, di natura occasionale (dovuta prevalentemente allo stato emotivo, quindi transitoria o idiosincratca) e quella sistematica dovuta all'invecchiamento fisiologico degli organi e delle membrane, talvolta affetti da patologie croniche.

[Handwritten initials]

Oltre alle variazioni intrinseche, dovute alla sorgente del segnale audio, vi sono poi le alterazioni indotte dal sistema strumentale di registrazione, che dipendono essenzialmente da:

- microfono
- canale di trasmissione
- apparecchio di registrazione
- supporto di registrazione

Ogni elemento del sistema introduce modifiche al segnale originale. Tali variazioni possono risultare trascurabili o meno in funzione della qualità dei materiali, delle condizioni generali di configurazione del sistema di registrazione e, generalmente, la qualità finale del prodotto sarà determinata dalle caratteristiche del peggior componente.

Tra i fenomeni più critici, che possono alterare una registrazione fino a renderla inutilizzabile a fini comparativi, menzioniamo la saturazione del segnale, il taglio in frequenza dovuto al canale di trasmissione e i rumori di fondo ambientali (o voci sovrapposte)...

Per queste ragioni è necessario porre dei limiti 'ab initio' alla possibilità di comparare voci. La richiesta di attribuzione della voce ad un dato individuo dovrà basarsi sulla possibilità di disporre di segnali che, a livello di qualità, soddisfino una minimale serie di requisiti.

Possono considerarsi ragionevolmente utili ad un confronto completo coppie di voci:

[Handwritten initials]

- di soggetti adulti, (cioè linguisticamente e fisiologicamente formati),
- concernenti idiomi linguistici omogenei,
- presenti su registrazioni che distino, temporalmente, un intervallo di tempo sufficientemente breve da poter considerare trascurabili gli effetti dovuti al normale invecchiamento (stimabile intorno ai sei anni) o ad eventuali modifiche strutturali del cavo orale (operazioni chirurgiche, insorgenza di malformazioni...)
- registrate da conversazioni ove siano simili le condizioni emotive, l'ideale sarebbe poter trattare conversazioni nelle quali la voce si possa considerare 'naturale'.
- incise con sistemi che rispettino maggiormente la fedeltà del segnale vocale originario (senza distorsioni, senza eco, senza disturbi che coprano il segnale...) o, comunque, siano il più possibile omogenei tra loro, per minimizzare le eventuali differenze dovute al canale di trasmissione o ai singoli componenti del sistema di registrazione.

È comunque da evidenziare che questi requisiti, pur soggiacendo alla corretta fattibilità delle varie analisi, strumentali e linguistiche, non sono *tutti* contemporaneamente necessari per poter comunque effettuare *qualcuna* delle analisi utili ad un confronto di voce; solamente si otterranno risposte più limitate rispetto a quanto si sarebbe potuto ottenere avendo la possibilità di approfondire tutti gli aspetti possibili.

È anche importante sottolineare come l'opzione di poter quantificare la probabilità nel processo di identificazione sia determinante soprattutto in caso di *positività* tra due voci. L'esame linguistico, che al momento non offre un risultato di tipo numerico⁴, comporta una *condizione necessaria non sufficiente* per sostenere un'identificazione. Le analisi linguistiche, in effetti, consentono la possibilità di discriminare se due voci appartengano o meno ad un '*gruppo socio-gergale*' circoscrivendo cioè la voce anonima ad una connotazione geografica definita e limitata ad un gruppo di persone ma non così stringente da poter appartenere *univocamente* ad un solo individuo. Eventuali sistematiche discrepanze sugli elementi linguistici possono condurre, invece, a collocare le voci dei soggetti a gruppi *disgiunti*, facendo così cadere la condizione necessaria all'identificazione e consentendo il giudizio di *incompatibilità* (= non identità) tra le voci. Per meglio far comprendere questo concetto è possibile ricorrere ad un semplice esempio: se si dovesse valutare l'identità delle persone sulla base di parametri quali altezza e peso, affermare che l'anonimo ed il sospettato presentano gli stessi valori significa indicare la *compatibilità* (che non è identità!) tra gli individui, tuttavia se si dimostrasse una differenza di 10 cm in altezza e 10 Kg in peso, a parità di condizioni temporali, saremmo sicuramente in grado di affermare la non identità tra i soggetti.

⁴ Perché ad oggi non è disponibile un database dei parametri linguistici nazionali

AN
413
A f

2.2 PRODUZIONE DELLA VOCE - MODELLO TEORICO

La voce da un punto di vista fisico è un'onda acustica che necessita di un mezzo di propagazione per diffondersi, tipicamente l'aria.

Come ogni fenomeno ondulatorio è caratterizzato da alcune grandezze, alcune scalari, quali il periodo o la frequenza di oscillazione, altre vettoriali come il numero d'onda o la velocità di propagazione.

Alcuni parametri dipendono dal mezzo di propagazione (velocità, vettore d'onda...), altri sono determinati solo dalla sorgente emissiva; tra questi ultimi, più importanti per un'eventuale caratterizzazione, annoveriamo la frequenza.

Per lo studio di un'onda complessa come il suono vocale l'analisi strumentale verrà quindi condotta sullo *spettro del segnale* che è, giustappunto, la rappresentazione del segnale nella scomposizione delle proprie componenti in frequenza.

La voce è un suono complesso perché è il risultato acustico di più fenomeni:

- il passaggio di aria attraverso la glottide che, a seconda della posizione delle corde vocali, genera rumore ma può produrre anche un *treno d'onde*;
- il seguente passaggio attraverso un condotto acustico che, fungendo da cassa di risonanza, modula il segnale entrante;
- la variabilità nel tempo della morfologia del condotto acustico, che permette l'emissione dei diversi suoni.

Tale condotto, chiamato *tratto vocale*, ha come estremità le labbra e dall'altra parte la glottide. La glottide è il sito della prima fonte di eccitazione (quasi) periodica, le corde vocali. L'uomo, nell'esposizione verbale, è capace di generare suoni in tre forme sonore basilari⁵:

- VOCALICHE: sono suoni prodotti eccitando il tratto vocale con un flusso d'aria costituito dal treno di impulsi, quasi periodico, generato dalla vibrazione delle corde vocali;
- FRICATIVE: forme prodotte formando un restringimento nel tratto vocale provocando, al passaggio forzato di aria, una turbolenza e quindi un particolare rumore (es. le consonanti *f, l, r, s...*);
- OCCLUSIVE: prodotte chiudendo completamente l'estremità superiore del tratto vocale. All'apertura rapida, per effetto della pressione dell'aria forzatamente racchiusa, si genera il fono occlusivo (es. le consonanti *b, p...*).

Una differenza sostanziale tra le tre forme consiste nella rapidità con la quale il suono viene emesso.

⁵ Esistono anche altre classi intermedie e sottoclassi per distinzioni più raffinate (nasali, laterali ...) ma questa trattazione appesantirebbe il discorso

Tutte e tre le forme corrispondono ad una eccitazione a larga banda del tratto vocale e sono rappresentabili matematicamente da un filtro che varia lentamente nel tempo e modifica con la propria risposta lo spettro di eccitazione. Il tratto vocale è intrinsecamente caratterizzato dalle sue frequenze naturali, che dipendono esclusivamente dalla forma e dalle dimensioni, che corrispondono a risonanze nelle sue caratteristiche di trasmissione. Dette frequenze prendono il nome di *formanti* e rappresentano le variabili cardine dei metodi strumentali per il riconoscimento del parlatore, per la riferibilità all'aspetto fisiologico del soggetto.

Delle tre forme basilari di fonazione la produzione delle vocali è quella che fornisce le condizioni di maggior stazionarietà al sistema matematico associato, nel senso che la conformazione del tratto vocale, nell'emissione di questi suoni, varia in tempi relativamente lunghi (anche oltre 50 ms) e può ragionevolmente considerarsi stabile nell'intervallo di misurazione.

Poiché i risultati delle analisi così condotte sono riproducibili, ed essendo la riproducibilità degli eventi il pilastro di ogni affermazione scientifica, ne segue che questa metodologia riveste un importante grado di attendibilità.

Supponendo che la sorgente di eccitazione e la forma del tratto vocalico siano approssimativamente indipendenti, la generazione della voce può essere rappresentata analiticamente da un sistema la cui uscita è data dalla convoluzione della risposta all'impulso del tratto vocale con la forma d'onda di eccitazione. Il modello proposto prende il nome di 'filtro lineare' (Fig. 1).

[Handwritten initials]

Un tale modello è idoneo non solo per l'analisi ma anche per la sintesi della voce, in tale caso $s(t)$ sarà un generatore di frequenze, $n(t)$ un generatore di rumore bianco e $h(t)$ rappresenterà un filtro variabile nel tempo.

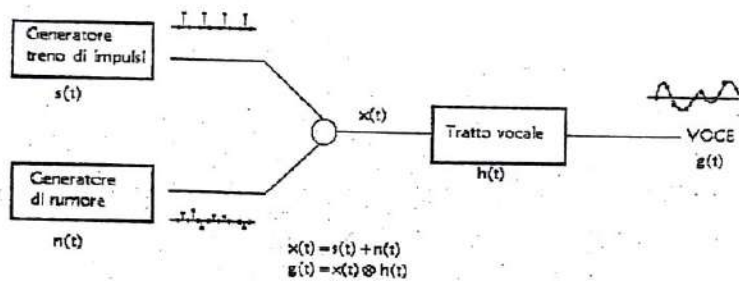


FIG. 1- MODELLO DI GENERAZIONE DELLA VOCE

Nell'analisi delle vocali, per quanto detto, il filtro numerico rappresenta il tratto vocale come fisso nella produzione del particolare fonema e può pertanto essere caratterizzato dalla sua risposta in frequenza (o dalla sequenza di coefficienti relativi ad un filtro con risposta ad un impulso di durata infinita).

Dalla teoria dei sistemi di trasmissione si ha che la funzione di trasferimento, definita come la trasformata z della risposta all'impulso, rappresenta matematicamente l'ambiente risonante dovuto al cavo orale.

Questo filtro viene eccitato da un treno di impulsi $s(n)$ in cui la distanza tra le creste corrisponde al periodo fondamentale, detto 'pitch'. La frequenza associata è nominata *Frequenza Fondamentale* (FF0).

e cf

Le frequenze di risonanza del tratto vocale (da cui le formanti FF_j) sono i poli della funzione di trasferimento e si ottengono dall'estrazione dei coefficienti c_k della funzione di trasferimento [5] fattorializzata e complessa associata al sistema stazionario⁶.

Data la difficoltà computazionale si comprende l'importanza di poter utilizzare un algoritmo di supporto che possa agevolare la ricerca di questi valori a partire dallo spettro generico (ottenuto come Trasformata di Fourier discreta) del segnale originario, uno di questi è basato sulla trasformata di Fourier del logaritmo dello spettro di potenza. Bogert e altri [6] chiamarono *Cepstrum* (anagramma di *spectrum*) tale funzione, e questo nome si è mantenuto fino ad oggi dall'epoca della pubblicazione del lavoro originale del 1963.

L'approccio mediante funzione *Cepstrum*, ancora oggi base di sviluppi applicativi, si adatta particolarmente bene all'analisi di classi di sequenze numeriche aventi trasformata z razionale fratta (caso in argomento), poiché fornisce con semplicità, e grazie ai moderni elaboratori anche con celerità, la possibilità di ottenere i valori delle FF_j , ricavabili dallo studio dei punti di massimo relativo della funzione rappresentata (poli). Per completezza bisogna ricordare che tra le evoluzioni del *Cepstrum* figurano anche altri algoritmi (con relativi coefficienti, noti in letteratura) tra i quali citiamo, in particolare, il *Mel-Cepstrum*, l'*LPC (Linear Predictive Coding-Cepstrum Derived)* [15]. Per un sistema di misura di tipo semi-automatico, dove la misura è sempre controllabile dall'operatore

⁶ in formule :

$$V(z) = \frac{A}{\prod_{k=1}^p (1 - c_k z^{-1})(1 - c_k^* z^{-1})}$$

significa, a livello grafico, rendere più facilmente interpretabile lo spettro ottenuto dal diagramma FFT (con coordinate *frequenza e intensità relativa*), evidenziando la curva di modulazione dovuta agli effetti di risonanza del tratto vocale. Grazie a questa *facility* i sistemi analitici di riconoscimento del parlatore di tipo semi-automatico sono particolarmente adatti per le esigenze forensi perché, a differenza di quelli totalmente automatici, consentono non solo la misura, ma anche la verifica (ed eventuali interventi correttivi) di ogni singola misura effettuata, a garanzia di un totale controllo.

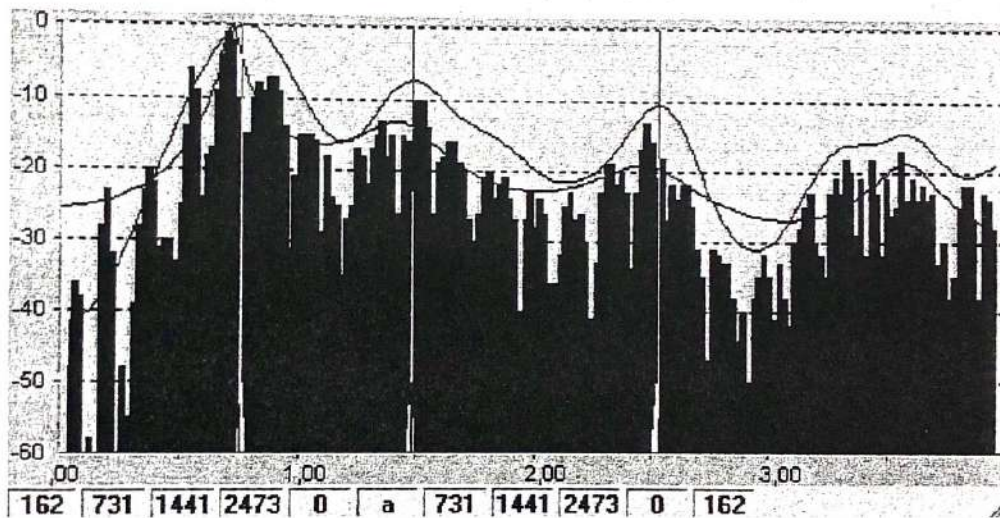


FIGURA 2: Spettro di una vocale 'a' maschile. La curva rossa rappresenta la funzione LPC, quella blu la funzione cepstrum; le barre indicano le prime tre formanti..

Per l'interpretazione statistica dei dati è doveroso domandarsi quante, delle variabili misurabili, possano effettivamente considerarsi significative (o indipendenti).

Per quanto riguarda le formanti FFj, risonanze della cavità orale, è necessario puntualizzare che se il condotto fosse ipoteticamente un 'tubo' senza asperità con una estremità chiusa (la glottide) ed una aperta (la bocca), bisognerebbe aspettarsi la produzione di n risonanze

legate dalla relazione matematica $F_n=(2n-1)F_1$ (cioè verrebbero emesse solo armoniche dispari della prima frequenza di risonanza [18]), e significherebbe una rigida interdipendenza dei parametri (in pratica si avrebbe una sola frequenza indipendente e tante armoniche correlate). Questa situazione, fortunatamente, non si verifica nel parlato quotidiano, poiché il condotto vocale contiene strozzature dovute alla posizione della lingua, la protuberanza dell'epiglottide e tutti i restringimenti più o meno volontari riscontrabili nel canale stesso, che provocano risonanze indipendenti (seconda, terza formante ecc...) e limitano la produzione delle armoniche correlate. In alcune rare circostanze, tuttavia, si possono ritrovare condizioni assimilabili al caso del 'tubo piatto', con forte dipendenza delle frequenze emesse dalla prima formante, sono i casi dei fonemi particolarmente centralizzati, quale per esempio la forma allofonica del fonema 'e' inglese presente nell'articolo *the* o delle grida (talvolta anche di parole pronunciate con volume sostenuto), casi nei quali il tratto vocale viene realmente 'appiattito' nell'emissione sonora (effetto *Lombard*).

Un limite per l'applicazione del modello del filtro lineare è il presupposto che la risposta della cavità orale si comporti, appunto, in modo lineare; questa condizione si verifica con buona approssimazione nel parlato naturale o, comunque, non eccessivamente sforzato.

4132

2.3 METODI UTILIZZATI PER LA COMPARAZIONE

Per quanto esposto in precedenza risulta evidente che la voce possiede caratteristiche fisiche parametrizzabili, misurabili e quindi comparabili.

Gli esami 'fisici' sul segnale, condotti sullo spettro delle vocali, offrono garanzie di oggettività e stabilità rispetto a variabili di ordine emotivo o idiosincratico, e producono risultati generalmente svincolati dall'intensità della voce e dal testo pronunciato, gli esami fonetico-linguistici, viceversa, sono generalmente stabili rispetto ad altre variabili quali il canale di trasmissione, il rumore di fondo e gli apparati impiegati, ma possono essere condizionati dal testo.

Descriviamo di seguito brevemente gli esami, generale-linguistico e fisico-oggettivo parametrico, che rappresentano i due pilastri del metodo di riconoscimento.

2.3.1 ESAME GENERALE E LINGUISTICO

Normalmente è suddiviso in quattro analisi.

La prima è basata sull'ascolto ripetuto del segnale audio e ha lo scopo di stimolare la memorizzazione 'inconscia' delle voci. Dall'osservazione è noto che ogni voce diviene per tutti più familiare e, quindi, facilmente riconoscibile anche in presenza di condizioni sfavorevoli (qualità scarsa del segnale audio, presenza di alterazioni...) tanto maggiore è il tempo che questa viene elaborata dal nostro cervello.

Si è soliti effettuare la distinzione *tra memoria a breve e lungo termine* sulla base sia della quantità di tempo dedicata

[Handwritten initials]

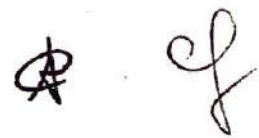
all'ascolto sia sulla capacità di ricordarne le caratteristiche a distanza.

Come esempio di memoria a breve termine possiamo riferirci ad ascolti ripetuti e concentrati nell'intervallo di una-due settimane, mentre per memoria a lungo termine l'esempio classico è configurabile dalla capacità che ognuno ha di riconoscere voci talmente 'metabolizzate' nel tempo (quali quelle dei congiunti stretti) che, di fatto, sono divenute patrimonio permanente della memoria. In simili condizioni sapremmo riconoscere queste voci anche dopo anni e, soprattutto, in qualsiasi condizione.

Le analisi di puro ascolto, chiamate anche 'uditive', sono basate sulla memoria a breve termine. Per questo genere di prova sono sufficienti pochi secondi di segnale, mentre per gli ulteriori accertamenti linguistici è necessario disporre di quanto più materiale possibile.

L'esame linguistico prosegue con tre analisi separate che tendono ad evidenziare:

aspetti fonetici, consistenti nella mappatura delle forme allofoniche utilizzate per vocalizzare i fonemi all'interno delle parole. La lingua italiana, a differenza di altre (es. il francese), non viene scritta ponendo in risalto tutte le componenti soprasegmentali (p.es. gli accenti) e, pertanto, alcuni fonemi subiscono variazioni sistematiche nella pronuncia a seconda delle varie influenze dialettali-regionali. Per fare un esempio il



numero tre può essere letto come *tré* o *trè* senza che ciò cambi il significato fonologico⁷. Essendo queste caratteristiche seminconscie del parlatore, dovute principalmente all'ambiente di crescita (famiglia, città, scuola primaria...) e consolidate con gli anni, generalmente forniscono un quadro indipendente da fattori esterni o emozionali. La mappatura può quindi fornire indicazioni sull'area geografica di provenienza o, comunque, di scolarizzazione. L'analisi fonetica, inoltre, è tesa ad individuare non solo quale allofono dell'alfabeto IPA⁸[19] sia stato utilizzato, ma anche ad evidenziare la presenza di eventuali difetti fonatori, quali, per esempio, la pronuncia gutturale del fonema 'r' o sibilante della 's' ecc..., circostanze, cioè, che possono essere più discriminanti perché meno dipendenti da fattori ambientali. Il risultato di un'eventuale concordanza nel confronto delle mappature è condizione necessaria non sufficiente per l'identificazione. In caso di particolarità significative tale quadro può assumere un'importanza fondamentale.

Aspetti semantico-lessicali, che insistono su elementi più variabili, quali la lingua-dialetto, la pianificazione semantica e sintattica del discorso e di tutte le forme espressive utilizzate per esprimere significati. È evidente l'importanza che assume tale analisi per i confronti interni (che in genere si riferiscono a conversazioni del medesimo genere/contenuto) mentre,

⁷ In parole più tecniche il fonema 'e', nella lingua italiana, consente i due allofoni 'é' ed 'è'

⁸ International Phonetic Alphabet

Handwritten initials/signature

tuttavia, non sempre si possono ottenere riscontri attendibili a fini investigativi, poiché spesso la terminologia utilizzata nelle conversazioni informali differisce, di norma, da quella impiegata in un interrogatorio o nell'acquisizione di un saggio fonico. Un esame linguistico 'perfetto' richiederebbe, infatti, di poter disporre di tutto l'universo espressivo di un locutore [24].

aspetti prosodici, infine, che investono le caratteristiche dell'andamento del discorso (*speech flow*), l'eventuale 'contorno intonativo', lo studio della velocità o della distribuzione delle pause ecc... Alcune di queste variabili, tuttavia, sono particolarmente sensibili alle condizioni emotive del parlante, pertanto a fini comparativi sarebbe preferibile lavorare con tratti che preservino quanto più possibile questo aspetto.

Lo scopo finale dell'esame generale-linguistico dovrebbe essere quello di ottenere un documentato quadro delle caratteristiche espressive che sia il più articolato possibile, con l'evidenziazione in particolare degli aspetti linguistici e fonetici generali, delle peculiarità del parlatore, degli eventuali difetti di articolazione dei fonemi o della frase ed eventuali elementi prosodici o lessicali di utilità, quali intercalari caratteristici, termini o frasi preferenziali ricorrenti, errori grammaticali o sintattici. La presenza di eventuali elementi ricorrenti di notevole rarità (da alcuni denominati

contrassegni) può aumentare enormemente il valore identificativo del confronto.

In sintesi, agli effetti di una attribuzione del parlato, gli aspetti di carattere linguistico conducono in genere ad una associazione del parlatore ad un'area socio-geografica di provenienza, la corrispondenza delle mappature fonetiche tra sospettato ed anonimo sarà condizione necessaria per un giudizio di compatibilità ma non sufficiente per l'identificazione, mentre l'ulteriore determinazione di caratteristiche individualizzanti, siano esse prosodiche o particolarità articolatorie personali o rare forme di dislessia, oltre a consentire un giudizio di sicura incompatibilità (qualora non riscontrate nella voce in comparazione), possono essere di significativa importanza per l'attribuzione, poiché restringono la dimensione della popolazione dei possibili individui con caratteristiche fonatorie compatibili.

In tema di studi prosodici sono state proposte differenti possibilità, dalla misurazione della lunghezza delle vocali atone e toniche [8] allo studio della durata delle consonanti occlusive [9], tuttavia appare più percorribile e particolarmente discriminante, in ultima analisi, la misurazione di un'altra variabile, la velocità media di fonazione. Operativamente significa misurare, nei tratti di parlato continuo o *real speech* (cioè riferibili a brevi frasi, generalmente della durata di 1-2 secondi, che non contengano pause, né respiratorie né di esitazione o di micropianificazione del linguaggio), il numero

Q *cf*

di battute di sillabazione o *sillabe fonetiche* al secondo, seguendo la definizione di *Articulation Rate* come formulata nella letteratura internazionale forense [17].

Test sull'impiego di questa variabile, applicata alla lingua italiana, che è ad isocronia sillabica e non accentuale (a differenza p.e. dell'inglese), hanno recentemente mostrato l'utilità del parametro nel riconoscimento a scopo forense [34], anche in virtù di una buona stabilità rispetto allo stato emotivo del parlatore (che, invece, ha riflessi sul numero e sulla durata delle pause [25]).

2.3.2 ESAME OGGETTIVO PARAMETRICO: IDE.M.[®]

L'esame strumentale condotto per l'espletamento di questo lavoro viene svolto avvalendosi del pacchetto software applicativo 'IDE.M.[®]' (IDEntification Method)[23], sviluppato appositamente per l'identificazione del parlatore dall'istituto di ricerca "Fondazione Ugo Bordoni" e utilizzato nel campo forense in più Nazioni. Il metodo è semi-automatico ed è denominato 'oggettivo' perché consente non solo di estrarre i valori fisici oggettivi che caratterizzano la produzione della voce (le formanti) ma permette anche di rappresentare e controllare (sia graficamente che analiticamente) le misure ad ogni passo ed, infine, consente una scelta decisionale articolata su test statistici consolidati che non comportano alcun intervento 'soggettivo' nella computazione.

L'intervento manuale nelle operazioni è, invece, limitato alla correzione nella misurazione delle formanti, lasciando all'operatore la possibilità di scartare spettri che, per esempio, non rispettano i requisiti di qualità o non siano sufficientemente interpretabili.

Il pacchetto, che prevede una preventiva digitalizzazione del segnale utile e successiva memorizzazione in file tipo *wav*, è modulare ed è composto da più programmi che possono essere avviati anche separatamente.

In particolare il programma *ARES* fornisce lo spettro di potenza mediante trasformata veloce di Fourier (*FFT*),

consentendo la selezione delle vocali idonee all'estrazione delle formanti e la successiva misurazione con l'ausilio degli algoritmi Cepstrum e/o LPC già discussi. I valori ottenuti vengono storicizzati in un file separato e costituiranno la base delle componenti per la vettorializzazione delle caratteristiche di ogni singolo parlatore nello spazio parametrico multidimensionale.

4/6

Le variabili oggetto di misurazione normalmente sono:

- la "Frequenza Fondamentale" FF0. E' la frequenza naturale di vibrazione (1° armonica) dovuta alle corde vocali e rappresenta la più bassa frequenza presente nello spettro vocale. Essa é legata alle caratteristiche anatomofisiologiche del parlatore ed al suo atteggiamento fonatorio;
- le "Frequenze Formanti" FFj. Sono risonanze del tratto vocale connesse con i parametri biometrici del parlatore; per le già citate condizioni di maggior stabilità del sistema matematico associato consideriamo le prime tre formanti riferibili alle sole vocali.

Le frequenze F0, F1, F2 ed F3 saranno estratte dalle quattro vocali *a*, *e*, *i* ed *o*, (la *u* isolata è rara nelle conversazioni abituali).

Il programma consente di visualizzare su una finestra del monitor (della durata di 2,5 sec.) la forma d'onda del tratto di segnale prescelto, dopodiché, mediante un puntatore, si localizza la vocale da misurare, con possibilità di ascolto per

Handwritten initials/signature

4

l'operatore tramite cuffia o altoparlante. Questa frazione di segnale, espansa nella scala temporale, è visualizzata su un'altra porzione dello schermo mentre un'altra finestra mostra, contemporaneamente, anche l'andamento spettrale, ottenuto mediante una FFT discreta pesata con finestra di Hamming, nonché la rappresentazione visiva degli involucri cepstrale (in blu) ed LPC (in rosso). Su tali immagini, in cui risultano ben visibili i picchi di energia, viene effettuata la misurazione dei valori delle formanti.

L'indeterminazione della misura soddisfa l'equazione $\Delta f \cdot \Delta t = 1$, il che significa che più è piccola la finestra di selezione, maggiore sarà l'incertezza della misura in frequenza. Solitamente si utilizzano finestre a 512 campioni su segnali campionati a 11.025 Hz, il che corrisponde ad una indeterminazione dell'ordine dei 20Hz.

La F_0 , invece, può essere valutata in tre modi:

- direttamente dal valore espresso dall'algoritmo 'Cepstrum',
- dalla differenza tra le armoniche della fondamentale presenti nello spettro,
- dalla forma d'onda, trattandosi della frequenza più bassa presente.

Tutti i valori relativi alle formanti sono trasmessi automaticamente in un file per la storicizzazione.

La misurazione è quantificata e riproducibile oggettivamente.

Non è previsto alcun intervento dell'operatore nelle fasi decisionali che saranno sviluppate dal successivo programma matematico-statistico denominato 'SPREAD'.

Il processo operativo globale che caratterizza l'indagine fonica si compone di tre fasi, la prima delle quali riguarda la scelta dei campioni di voce da analizzare. Questa operazione è la più delicata perché occorre selezionare parole o frasi che, oltre ad essere effettivamente rappresentative dell'espressività del parlatore, devono anche possedere caratteristiche qualitative tali da consentire le necessarie misurazioni strumentali⁹; i valori relativi alle voci anonime sono prelevati dal materiale fornito dall'A.G., mentre i termini di paragone sono generalmente estratti da saggio fonico.

La seconda fase consiste nell'estrarre da ciascun campione i parametri prescelti per la caratterizzazione della voce.

L'operatore interviene nella scelta delle parole o dei suoni da utilizzare nell'indagine, al fine di selezionare i soli fonemi che rispettino le condizioni di misurabilità (rapporto segnale/rumore, leggibilità dello spettro...).

Nella terza fase, infine, questi vettori vengono elaborati dal programma 'SPREAD', che fornisce una valutazione statistica del confronto, utile per rispondere ai due quesiti fondamentali per poter giungere ad una possibile attribuzione:

⁹ Sono preferibili vocali tipiche dell'alfabeto italiano classico, toniche, di lunghezza non inferiore alla finestra FFT utilizzata e con rapporto s/n maggiore di 12 dB.

- l'appartenenza o meno di un campione di voce anonima ad uno o più campioni di persone sospettate, con un prefissato livello di significatività (nel qual caso sarà conseguentemente possibile affermare la positività dell'esame, intendendo con ciò che le voci analizzate presentano le medesime caratteristiche acustiche e, pertanto, possono ritenersi indistinguibili entro il regime di tolleranza prefissato);
- la probabilità, in caso positivo, che un'altra persona, diversa dal sospettato, possa avere le stesse caratteristiche in fase di produzione acustica (*probabilità di falsa identificazione*).

Molti esperti, in Europa, si chiedono se sia meglio fornire ad un Giudice un numero o piuttosto esprimere un giudizio secondo una scala verbale ben definita [29]. Alcuni di questi hanno all'uopo studiato una specie di *tabella di conversione*, che possa tradurre un esito numerico in una frase più comprensibile alle orecchie della Giuria, del tipo '*Gli esiti degli esami forniscono forti/deboli/neutrali elementi a supporto dell'ipotesi di (non)identificazione*' [30, 31]. In ogni caso, anche la risposta 'verbale' deve essere fondata su una stima matematica e non sul semplice intuito o su una presunta e comunque soggettiva 'esperienza'.

Per calcolare la probabilità di falsa identificazione (in letteratura conosciuta anche come *False Acceptance*

Probability) è necessario disporre di un adeguato database comprensivo dei dati di una popolazione di riferimento; a questo scopo sono attualmente rivolti gli sforzi della comunità scientifica operante nel settore [26, 27]. Il sistema IDE.M. ha implementato un database riferito ad una popolazione maschile italiana, contenente circa 20.000 misure.

4/11

L'assetto del test di identificazione utilizzato dal programma SPREAD per rispondere alla prima domanda, cioè analizzare la possibilità di associazione di un campione di voce anonima ad un campione di riferimento (o eventualmente saggio fonico), appartenente ad una persona nota (*test si/no*), ha le seguenti caratteristiche generali:

- Dati K saggi si ricavano le popolazioni di $n(k)$, ($k=1, \dots, K$), vettori parametrici p -dimensionali, descritte ognuna da un vettore centroidico $C(k)$, e da una matrice di covarianza calcolata $W(k)$, stimata con $n(k)-p-1$ gradi di libertà;
- assegnati M campioni anonimi, per ciascuno si estraggono $n(f,m)$ misure dei parametri acustici dai fonemi f , assunti come descrittori per il modello statistico dato;
- è richiesto un test di accettazione dell'ipotesi di appartenenza di ciascuno degli M campioni a ciascuno dei K saggi;
- se l'ipotesi è accolta per il campione m -esimo al confronto del saggio k -esimo, il campione m -esimo viene classificato

Q *Q*

come di voce indistinguibile dalla voce del saggio dell'indiziato k -esimo¹⁰;

• se l'ipotesi viene respinta, il campione m -esimo viene definito di voce diversa da quella del k -esimo saggio.

I test statistici capaci di rispondere alla domanda posta sono ricavati in funzione delle proprietà statistiche delle popolazioni di vettori, tenendo conto delle ipotesi discusse in precedenza.

Si assume che campione e saggio abbiano la stessa matrice di covarianza W e che vi siano abbastanza dati per stimarla dalle misure disponibili; la distanza calcolata tra le voci è quella canonica di Mahalanobis [10] espressa dalla forma bilineare:

$$D = Ad' W^{-1}(m,k) d$$

dove

d è la differenza tra i due vettori: $c(m) - c(k)$

d' è il vettore d trasposto

A è un fattore di normalizzazione

¹⁰ All'interno dell'intervallo di tolleranza, dipendente dal livello di confidenza α che si vuole stabilire. L'espressione italiana che meglio traduce il valore 'si' che dovesse esprimere il test è: 'Le due voci risultano strumentalmente indistinguibili all'interno del dominio determinato dalla soglia prefissata';

$W^{-1}(m,k)$ è la matrice inversa della matrice di covarianza, che può essere costruita a partire dagli $n(m)$ dati del campione m -esimo e dagli $n(k)$ dati di quello k -esimo.

Disponendo di una stima della matrice intraparlatores W , calcolata su una banca dati fornita insieme al programma, tale da poter essere assunta di fatto coincidente con la matrice del modello effettivo¹¹, la funzione statistica di distribuzione, per il test di accettazione, si semplifica divenendo una distribuzione χ^2 con p gradi di libertà [11], nel caso di 4 vocali e 4 frequenze si avrebbe $p=4 \times 4=16$. Per una corretta applicazione della funzione χ^2 è altresì necessario specificare entro quale livello di confidenza α il test si considera positivo; generalmente si considera $\alpha=0,01$, corrispondente al valore di circa $2,5\sigma$ di distanza massima dalla media entro la quale poter considerare lo scostamento come compatibile con la variabilità intrapersonale.

In altri casi, cioè quando non è disponibile una matrice-modello W per il campione dell'anonimo e del saggio¹², il programma SPREAD è in grado di calcolare autonomamente la matrice W in base alla popolazione fornita dal test-utente; si seguirà, a tal punto, la statistica T^2 di Hotelling, applicabile anche al caso minimale di due campioni [14].

¹¹ Nel database vi sono i dati di circa 400 parlatori per un totale di oltre 15.000 misure

¹² È il caso, p.es., delle voci femminili o relative a stranieri.

Generalmente si considerano come fonemi utili le prime quattro vocali della lingua italiana, perché più numerose nel parlato e più stabili nella misura; misurando quindi la F0 e le prime tre formanti si ottengono così 16 variabili e la W è, in questo caso, una matrice quadrata¹³ 16x16.

Fissando $\alpha=0.01$ per 16 gradi di libertà il valore limite di χ^2 è 32, per 4 gradi di libertà, invece, il limite di χ^2 vale 13,28. Il test sarà considerato positivo se il valore di χ^2 (relativo ai dati della comparazione) è inferiore a quello della prefissata soglia. Il test completo, inoltre, oltre a confermare l'identità tra voci anonime e saggi, consente di stimare l'errore massimo di falsa identificazione nel senso di cui al punto 2 della premessa al paragrafo, cioè valuta quale probabilità ha di esistere una persona che, dato l'esito positivo, possa avere caratteristiche vocali indistinguibili da quelle da noi evidenziate per la coppia anonimo-saggio (è la *probabilità di falsa identificazione* o, in altre parole, la probabilità di avere un *falso positivo*).

Ciò premesso, è chiaro che gli elementi essenziali da calcolare sono le probabilità di *falsa reiezione* (= *falso negativo*) e di *falsa identificazione*, le sole che danno contributo all'aspetto decisionale.

Operativamente si inizia, pertanto, fissando la già citata soglia, che per *default* è assunta all'1% ($\alpha=0,01$) per fonema, il che

¹³ Il test può, tuttavia, essere scomposto separando le 4 vocali in 4 sottomatrici 4x4, in questo caso L'ipotesi di reiezione di identità è da accettarsi se anche uno solo dei test fornisce questo risultato.

attesta un livello di confidenza¹⁴ del test strumentale pari al 99%. Se il test dà esito negativo, significa ammettere, per costruzione stessa, che la probabilità di falsa reiezione è inferiore all'1%.

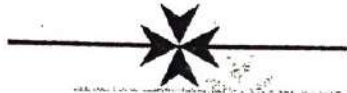
Una volta superato il test dicotomico accettazione/reiezione con esito positivo, diviene necessario valutare analiticamente la *P.f.i.*. Il programma 'SPREAD' utilizza, oltre ai valori medi contenuti nei vettori p -dimensionali, anche il livello di confidenza impostato $(1-\alpha)$ e il numero totale dei campioni estratti da ogni coppia¹⁵ saggio-voce anonima $(n1+n2)$. Su queste basi la *P.f.i.* è valutabile come il volume di spazio della funzione di densità di probabilità della distribuzione dei campioni di voce intercettata dall'ellissoide multidimensionale che, al livello di confidenza prefissato definisce la regione nello spazio multidimensionale entro la quale si determina l'esito positivo del test di identificazione. Questo integrale viene approssimato mediante un algoritmo Montecarlo.

Riassumendo la duplice funzione del programma SPREAD con un esempio: se stessimo trattando un tipo di riconoscimento biometrico basato su due caratteristiche anatomiche, p.es. l'altezza ed il colore dei capelli, il primo test *yes/no* calcola se una persona bionda alta 207,1 cm può essere

¹⁴ In altre parole poiché ogni individuo ha una naturale variabilità, è necessario fornire al sistema un limite, una soglia per la decisione di attribuzione ad una medesima persona. $\alpha=0,01$ significa tarare il sistema in modo da ammettere un riconoscimento della medesima persona 99 volte su 100.

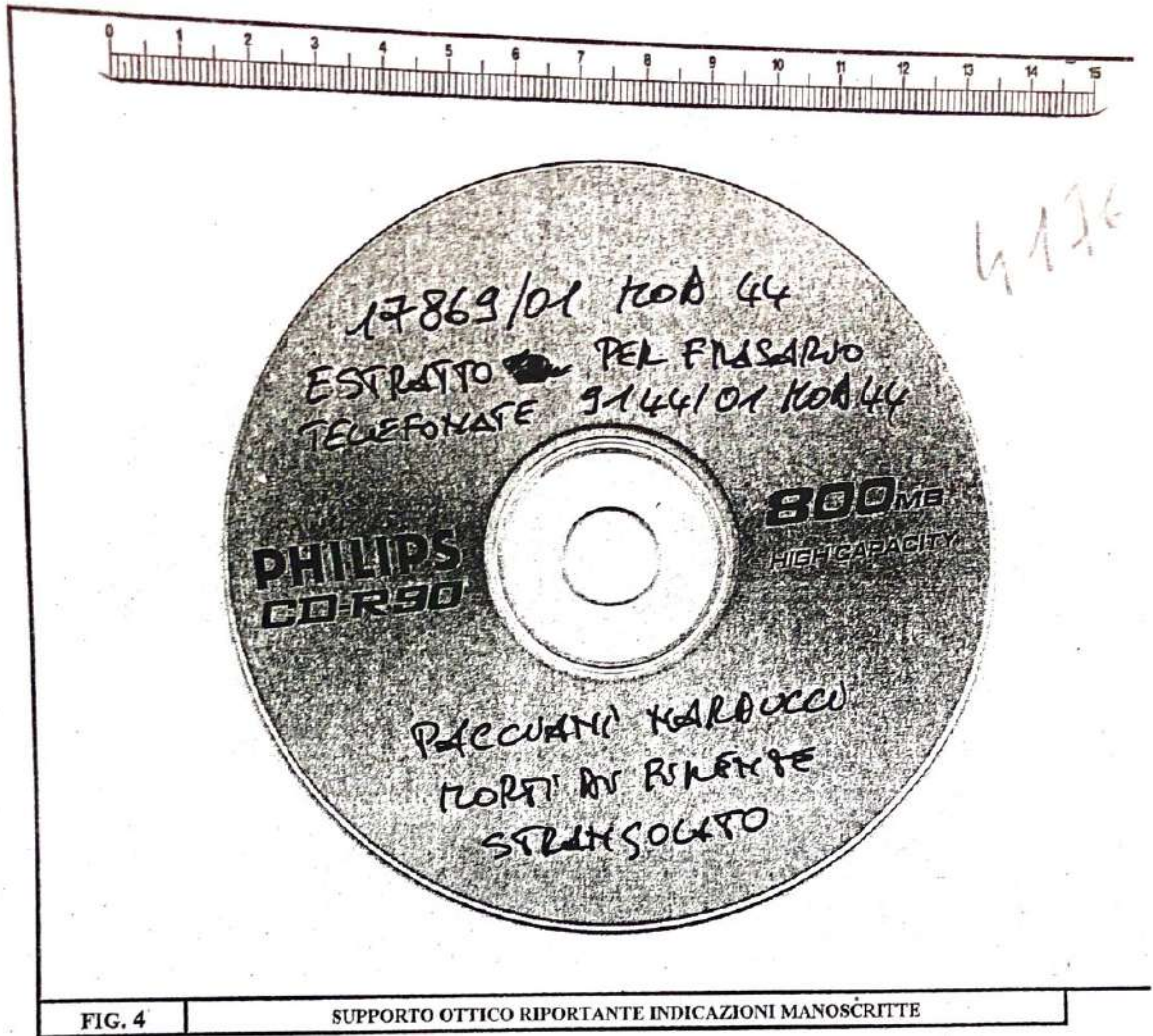
¹⁵ Infatti è noto che, per una variabile gaussiana, la deviazione standard del valor medio è legata al numero di osservazioni dalla formula $\sigma_{VM} = \sigma \sqrt{(n_1+n_2)/n_1 n_2}$

distinguibile o meno da un'altra, bionda, di 207 cm, mentre il secondo test esplicita che, essendo simili valori inconsueti per la popolazione italiana, il dato ha un elevato rilievo identificativo (cioè piccolo errore di falsa identificazione).



Handwritten signature or initials

Handwritten initials or marks



Il CD-R contiene 20 conversazioni telefoniche.

La prima attività è stata quella di riversare sulla memoria di massa di un Personal Computer dotato di adeguato processore e idonea scheda audio il segnale audio nel suo complesso. Successivamente si è proceduto:

- ♦ a separare ciascuna conversazione;
- ♦ a isolare la voce dei locutori anonimi, con creazione di files distinti.

Q f

U 114

Ogni singolo file è stato quindi riascoltato ripetutamente per stimolare la memoria a breve termine, fondamento per l'inizio dell'attività inerente gli esami linguistici.

L'utilità del materiale a comparazioni strumentali è stata valutata mediante il programma QUALITÀ (inserito in IDE.M.) che permette di esaminare i parametri generali del segnale (rapporto s/n, banda passante), fornendone al contempo una rappresentazione grafica.

Il metodo descritto nel Cap. 2 per il riconoscimento del parlatore richiede infatti che il segnale vocale abbia alcune caratteristiche qualitative minime per condurre ad un'analisi attendibile:

- rapporto S/N: almeno 12 dB
- banda passante: ≥ 3 KHz
- quantità di segnale: almeno 5 campioni per vocale

Si ricorda che:

- Il rapporto segnale/rumore (o s/n), è il rapporto tra l'energia del segnale utile (il parlato) e quella del segnale disturbante (rumore)¹⁶.
- Per banda passante si intende quella parte dello spettro di frequenza compreso tra le due frequenze di taglio superiore ed inferiore.
- Per frequenza di taglio si intende quel valore al quale corrisponde una potenza sul carico pari a metà di quella massima (abbassamento di 3 dB).

¹⁶ Si misura in decibel, definito come misura relativa tra due stati di intensità del livello sonoro: s/n (dB) = $10 \log_{10} (I_s/I_r)$, 3 dB corrispondono a I_s doppia di I_r , 10 dB a I_s 10 volte superiore a I_r e 20 dB 100 volte.

- I fonemi vocalici su cui è 'tarato' il programma sono le vocali appartenenti all'alfabeto fonetico in uso alla lingua italiana¹⁷.

Le operazioni di pretrattamento delle conversazioni hanno, in sintesi, seguito questo iter:

- I. Ricerca delle voci di interesse;
- II. Riversamento del segnale sulla memoria di massa di un PC, digitalizzando mediante campionamento a 11.025 KHz e risoluzione a 16 bit, creando un file per ogni voce/conversazione;
- III. Editing (Cut & Paste) con creazione di 20 files relativi ad altrettante voci presenti sulle conversazioni rilevate;
- IV. Controllo qualità segnale per prove comparative strumentali;
- V. Familiarizzazione con le voci mediante ascolto ripetuto (stimolazione della memoria a breve termine);

¹⁷ non necessariamente la conversazione deve essere in italiano, per le misure è sufficiente considerare le vocali che sono 'tipiche' della lingua italiana

R J

Le prime analisi del segnale relativo alle conversazioni oggetto di consulenza hanno portato al seguente risultato:

448

nomefile "tel 01_cut.wav"			
durata ¹⁸	registrazione	Rapporto s/n	Banda passante
~36"	Telefonica	<p>18 dB</p>	300-3300 Hz

nomefile "tel 02_cut.wav"			
durata	registrazione	Rapporto s/n	Banda passante
~54"	Telefonica	<p>21 dB</p>	300-3300 Hz

¹⁸ Riferita alla solo durata del parlato del locutore di interesse.

Handwritten initials or marks.

		nomefile "tel 03 cut.wav"	
durata	registrazione	Rapporto s/n	
~1'03"	Telefonica	<p>SNR : 19 TRH : -30</p> <p>19 dB</p>	
		Banda passante 300-3300 Hz	

		nomefile "tel 04 cut.wav"	
durata	registrazione	Rapporto s/n	
~31"	Telefonica	<p>SNR : 17 TRH : -28</p> <p>17 dB</p>	
		Banda passante 300-3300 Hz	

		nomefile "tel 05 cut.wav"	
durata	registrazione	Rapporto s/n	
~37"	Telefonica	<p>SNR : 16 TRH : -27</p> <p>16 dB</p>	
		Banda passante 300-3300 Hz	

Handwritten signature or initials

		nomefile "tel 06_cut.wav"	
durata	registrazione	Rapporto s/n	
~1'24"	Telefonica		
			Banda passante 300-3300 Hz

		nomefile "tel 07_cut.wav"	
durata	registrazione	Rapporto s/n	
~48"	Telefonica		
			Banda passante 300-3300 Hz

		nomefile "tel 08_cut.wav"	
durata	registrazione	Rapporto s/n	
~1'12"	Telefonica		
			Banda passante 300-3300 Hz

[Handwritten signatures]

54



durata ~1'10"	registrazione Telefonica	nomefile "tel 09 cut.wav"	
		Rapporto s/n	
			Banda passante 300-3300 Hz

durata ~1'30"	registrazione Telefonica	nomefile "tel 10 cut.wav"	
		Rapporto s/n	
			Banda passante 300-3300 Hz

durata ~1'24"	registrazione Telefonica	nomefile "tel 11 cut.wav"	
		Rapporto s/n	
			Banda passante 300-3300 Hz

Handwritten initials or signature




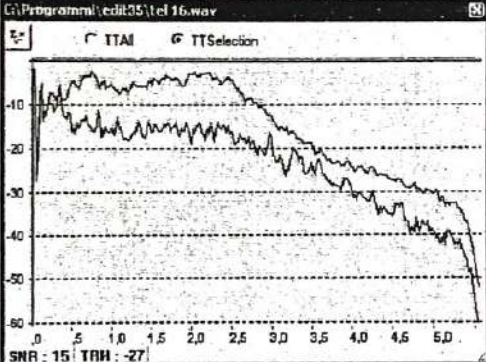
		nomefile "tel 12 cut.wav"	
durata	registrazione	Rapporto s/n	
~1'14"	Telefonica		
		Banda passante 300-3300 Hz	

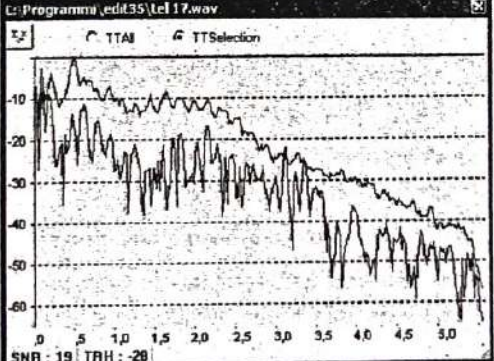
		nomefile "tel 13 cut.wav"	
durata	registrazione	Rapporto s/n	
~2'31"	Telefonica		
		Banda passante 300-3300 Hz	

		nomefile "tel 14 cut.wav"	
durata	registrazione	Rapporto s/n	
~2'38"	Telefonica		
		Banda passante 300-3300 Hz	

Q Q

		nomefile "tel 15 cut.wav"	
durata	registrazione	Rapporto s/n	
~1'45"	Telefonica	Banda passante	
		300-3300 Hz	
			
		18 dB	

		nomefile "tel 16 cut.wav"	
durata	registrazione	Rapporto s/n	
~2'22"	Telefonica	Banda passante	
		300-3300 Hz	
			
		15 dB	

		nomefile "tel 17 cut.wav"	
durata	registrazione	Rapporto s/n	
~4'29"	Telefonica	Banda passante	
		300-3300 Hz	
			
		19 dB	

Ⓢ Ⓣ

durata ~1'59"	registrazione Telefonica	nomefile "tel 18_cut.wav"	
		Rapporto s/n	Banda passante 300-3300 Hz

durata ~1'31"	registrazione Telefonica	nomefile "tel 19_cut.wav"	
		Rapporto s/n	Banda passante 300-3300 Hz

durata ~1'39"	registrazione Telefonica	nomefile "tel 20_cut.wav"	
		Rapporto s/n	Banda passante 300-3300 Hz

[Handwritten signature]

Nel ribadire che per gli esami richiesti sono da considerarsi idonei i files audio che soddisfano i requisiti minimi di qualità (≥ 12 dB) e di quantità, riguardo all'utilizzabilità delle conversazioni per gli esami comparativi si segnala che tutti i files audio sono idonei sia dal punto di vista qualitativo che quantitativo.

4186

a, 54

IAN

[Handwritten initials]

3.3 ESAME GENERALE LINGUISTICO-FONETICO

Seguendo quanto previsto dalla metodologia si sono effettuate, separatamente e con i limiti esposti già al §2.3.1, le analisi che, nel loro complesso, costituiscono l'esame comparativo generale di tipo linguistico-fonetico.

L'ascolto dei campioni vocali (tel 01.wav, tel 02.wav... tel 20.wav) individua la presenza di due locutori anonimi le cui voci sono da riferirsi a un soggetto maschile e a un soggetto femminile.

1. Soggetto maschile

La stessa voce maschile è rilevabile nelle telefonate nn. 2, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19 e 20.

A supporto dell'ipotesi dell'unicità del parlatore:

- la sensazione a livello percettivo;
- la circostanza che in tutte le conversazioni il soggetto ponga in essere la medesima artefazione della propria naturale fonazione, attuata mediante laringalizzazione con conseguente voce arrochita e forzata;
- le particolarità linguistiche rilevate;
- la modalità nello "sghignazzare" (p.e. nelle tel. 7, 8, 10, 11, 12...);

Handwritten initials

	consonantico atipico con leggero avanzamento della mandibola per le dentali "t" e "d" in cui il punto di articolazione della "d" è più arretrato (postalveolare) rispetto a quello standard (p.e. nella telefonata 10 ... <i>il demone ti corrode la tua anima...</i> - nella telefonata 13 ... <i>da quella ... della tua...</i>) e nella "t" avvertibile nella telefonata 13 ... <i>sfregiata stati...</i>
Intensità	Non valutabile stante l'artefazione attuata
Intonazione	Non valutabile stante l'artefazione attuata
Lessico	Informale - corretto - di estrazione culturale medio-alta (p.e. dalla tel. 10 ... <i>il demone ti corrode la tua anima...</i> dalla tel. 16 ... <i>balanzosa...</i>) - tecnico settoriale (p.e. dalla tel. 7 ... <i>le tenebre di satana...</i> dalla tel. 9 ... <i>il nostro adepto...</i> dalla tel. 12 ... <i>sarai sacrificata in nome di satana...</i> dalla tel. 13 ... <i>sacerdoti di satana...</i>)
Velocità di articolazione	Non misurabile stante l'artefazione attuata
Particolarità fonetiche:	Individuali relative alla pronuncia della liquida "r" in modalità uvulare vibrante (p.e. ... <i>dottore...</i> → ... <i>dottore...</i>) della plosiva "d" in modalità sonora retroflessa (p.e. ... <i>demone...</i> ... <i>corrode...</i>) e della fricativa "s" in modalità arretrata (p.e. ... <i>schifo...</i> ... <i>scimmia...</i>)
Andamento prosodico	Regolare con assenza di pause che denotano una certa padronanza nel discorso trattato e una organizzazione ben strutturata in fase di micropianificazione del discorso.

2. Soggetto femminile

La stessa voce femminile è rilevabile nelle telefonate nn. 1, 3, 4, 5, 6 e 17.

A supporto dell'ipotesi formulata si segnala:

- la sensazione a livello percettivo;
- le particolarità linguistiche rilevate;
- l'impiego di locuzioni ricorrenti (p.e. "nostro signore satana", "maledetta puttana", "tuo figlio", "sacrificato", "servo"...) con

R *f*

argomentazioni frequentemente ripetute. ("...la morte è vicina..." "...tuo figlio lo prendiamo noi...");

4193

Dal materiale si sono estrapolate alcune caratteristiche utili per il profilo del parlatore, riportate in tabella.

Registrazione:	Telefonica
Sesso	Femminile
Età	<u>Adulta</u> (è possibile ipotizzare una fascia compresa dai 32 ai 45 anni)
Timbro ²¹	Mezzo soprano
Lingua:	Italiana
<u>Inflessione:</u>	<u>Tipica dell'Italia settentrionale</u> (dagli elementi rilevati dal punto di vista della costruzione frastica si rileva la marcata intonazione ascendente e discendente con apice nella penultima sillaba (piana) del termine finale con voce querula - è possibile ricondurla verosimilmente a quella piemontese) (p.e. nella tel 6 ... <i>ma credi che avrà un futuro... il tuo cornuto di tuo marito... sulle colline...</i>).
Intensità	Media
Intonazione	Cantilenante
Lessico	Informale - corretto - di estrazione culturale medio-alta (p.e. dalla tel. 3 ... <i>sappilo...</i> dalla tel. 4 ... <i>il tuo ciarlatano è un sacrilego...</i>) - tecnico settoriale (p.e. dalla tel. 1 ... <i>con nostro signore satana...</i> dalla tel. 4 ... <i>il tuo ciarlatano è un sacrilego...</i> dalla tel. 5 ... <i>con il nostro rito per satana...</i>)
Velocità di articolazione	~5,4 s/s
Particolarità fonetiche:	Individuali relative alla pronuncia della liquida "r" in modalità uvulare vibrante (p.e. nella tel. 5 ... <i>questo avverrà...</i>)
Andamento prosodico	Regolare con assenza di pause che denotano una certa padronanza nel discorso trattato e una organizzazione ben strutturata in fase di micropianificazione del discorso.

²¹ Secondo la classificazione di R. T. Sataloff (*La voce umana*, da *Le Scienze*, 294/1993, pag.66).

Handwritten initials or signature.



3.4 RISPOSTA AI QUESITI

- IL SUPPORTO OTTICO ESAMINATO CONTIENE LA REGISTRAZIONE DI 20 (VENTI) CONVERSAZIONI TELEFONICHE, NUMERATE DAI PERITI PROGRESSIVAMENTE DA 1 A 20 SECONDO L'ORDINE DI PRESENTAZIONE;
- LE PROVE DI ASCOLTO E L'ANALISI LINGUISTICA INDIVIDUANO COME AUTORI DELLE TELEFONATE: UN MEDESIMO LOCUTORE MASCHILE PRESENTE NELLE CONVERSAZIONI NN. 2, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20 E UN MEDESIMO LOCUTORE FEMMINILE NELLE RESTANTI CONVERSAZIONI NN. 1, 3, 4, 5, 6 E 17;
- IL TEMA TRATTATO, I TERMINI IMPIEGATI E ALTRE PECULIARITÀ INSOLITE (COME P.E. IL MODO DI SGHIGNAZZARE) SONO IN COMUNE A ENTRAMBI I PARLATORI ANONIMI: LA CIRCOSTANZA IMPLICA CHE L'UNO SIA AL CORRENTE DELLE MODALITÀ ATTUATE DALL'ALTRO IN FASE DI CONDUZIONE DELLA CONVERSAZIONE (È RAGIONEVOLE PENSARE QUINDI CHE SUSSISTA UN TENTATIVO DI EMULAZIONE);
- LE PARTICOLARITÀ ESPRESSIVE RILEVATE MANIFESTANO PER ENTRAMBI I LOCUTORI IN VERIFICA UN LESSICO INFORMALE, CORRETTO, TECNICO-SETTORIALE CHE PORTEREBBE AD IPOTIZZARE L'APPARTENENZA AD UNA SETTA O CONGREGAZIONE (PER IL TEMA TRATTATO - ESOTERICO) E UNA ESTRAZIONE CULTURALE MEDIO-ALTA.
- IL PARLATORE ANONIMO PONE IN ESSERE UNA ARTEFAZIONE DELLA PROPRIA NATURALE FONAZIONE MEDIANTE LARINGALIZZAZIONE CON CONSEGUENTE VOCE ARROCHITA E FORZATA. LA VOCE DELLA PARLATRICE ANONIMA È VEROSIMILMENTE PRIVA DI ARTEFAZIONE E, PERTANTO, UTILE A EVENTUALI FUTURI CONFRONTI BASATI ANCHE SU ANALISI DI TIPO STRUMENTALE;
- LE CARATTERISTICHE LINGUISTICHE RILEVATE DI CIASCUNO DEI PARLATORI

Handwritten initials/signature



4191

SONO MEGLIO DESCRITTE NEL § 3.3;

○ DALL'ANALISI LINGUISTICA (FONETICO-ARTICOLATORIA) SI INDIVIDUA
VEROSIMILMENTE COME ZONA DI ORIGINE:

- PER LA VOCE MASCHILE COMPATIBILE CON QUELLA DI CANNARA (§ 3.3 PUNTO 1);
- PER LA VOCE FEMMINILE COMPATIBILE CON QUELLA PIEMONTESE (§ 3.3 PUNTO 2).

Unitamente alla presente relazione, si restituisce il CD-R oggetto di accertamento.

Roma, li 26 ottobre 2005

I Consulenti del P.M.

Ernesto Laferla
Luigi Cione