

AMLETO

Trasparenza vs. black box

un sistema automatico verificabile
per la comparazione forense della voce

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

COS'È AMLETO

Software di comparazione vocale basato su rete neurale ECAPA-TDNN su corpus telefonico italiano.

Il sistema implementa la metodologia Likelihood Ratio secondo gli standard ENFSI 2021 a 13 livelli.

Validato internamente secondo i criteri del consenso internazionale e già impiegato per perizie in parallelo ad altri strumenti

È un proof of concept funzionante, a pipeline verificabile per la comparazione forense della voce

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

COS'È AMLETO

AMLETO utilizza **SpeechBrain** per caricare il modello basato sull'architettura **ECAPA-TDNN**: una rete neurale che analizza le caratteristiche acustiche della voce nel tempo, dà maggiore peso alle informazioni utili a distinguere i parlanti e le riassume in un vettore numerico, chiamato **embedding**.

ECAPA Emphasized Channel Attention, Propagation and Aggregation
dà peso alle informazioni più utili e combina caratteristiche estratte a diversi livelli.

TDNN Time Delay Neural Network
analizza le caratteristiche vocali considerando anche il contesto temporale, cioè ciò che avviene nei momenti vicini.

COS'È AMLETO

Il modello ECAPA-TDNN pre-addestrato è stato sottoposto a fine-tuning su un corpus di telefonate in lingua italiana, con l'obiettivo di ottimizzare le performance nel dominio specifico delle intercettazioni telefoniche italiane. La fase di fine-tuning ha utilizzato un corpus di più di 60.000 file audio di telefonate in lingua italiana, producendo un modello specializzato per il parlato telefonico forense italiano.

COS'È AMLETO

Come popolazione di riferimento per la normalizzazione dello score di comparazione è stato utilizzato un corpus di circa 18000 registrazioni telefoniche in lingua italiana, corpus ampiamente utilizzato nella letteratura scientifica italiana e internazionale come benchmark per i sistemi di riconoscimento del parlante su parlato telefonico italiano.

COS'È AMLETO

I Corpus* contenenti migliaia di registrazioni sono stati acquisiti a suo tempo per scopi di ricerca, sviluppati con questo spirito e ricevuti con lo stesso spirito dal Dr. Falcone,

della Fondazione Ugo Bordoni.

*EVALITA 2009, CALI + Common Voice

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

PERCHÈ AMLETO

ENFSI dice espressamente:

combinare metodi automatici, semiautomatici e linguistico-fonetico-acustici

Tuttavia, periodicamente ci ritroviamo a dover leggere relazioni e dibattere con CTP, CTU e Periti, che presentano perizie con solo risultanze Automatiche, senza oltretutto avere la possibilità di testare quello che producono

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

PERCHÈ AMLETO

NON esiste un software automatico che possa risolvere da solo una qualunque perizia di comparazione vocale

Per fare una Perizia di Comparazione Vocale occorrono Sempre delle Metodologie Complementari

LA PERIZIA LA FA IL PERITO

Quando la Giustizia dà i numeri!

"E' la stessa voce, oppure sono due voci diverse?"

Sono anni che racconto ciò che mi è successo con i software più blasonati sul mercato.

Quando la Giustizia dà i numeri!

"E' la stessa voce, oppure sono due voci diverse?"

Ho scritto

Mi metto a disposizione per qualsiasi prova PRATICA,

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

Quando la Giustizia dà i numeri!

"E' la stessa voce, oppure sono due voci diverse?"

RISULTATO:

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

NASCE AMLETO

AMLETO

· Bayesian LR · ENFSI · ECAPA Fine-tuned

A — Voce Campione (Known)

/Users/marcperi/Downloads/Genova/Ignoto_Y.m4a

Seleziona file audio

B — Voce Incognita (Questioned)

/Users/marcperi/Downloads/Genova/Andrea/Audio 2.m4a

Seleziona file audio

TIPO DI AUDIO

Popolazione di Riferimento

✓ Popolazione di riferimento caricata

▶ AVVIA COMPARAZIONE

✓ Comparazione completata — Supporto forte per H2

RISULTATO LR

LR = 4.29e-04

$\log_{10}(\text{LR}) = -3.368$

Cosine similarity = -0.0227

$\log_{10}(\text{LR}) = -3.37 \rightarrow \text{LR} = 4.29\text{e-}04$

Le tracce vocali sono 2332 volte più probabili se provengono da persone DIVERSE.

SCALA ENFSI

■ ENFSI -3 LR 0,001 - 0,01

→ Indicazione verso persone DIVERSE

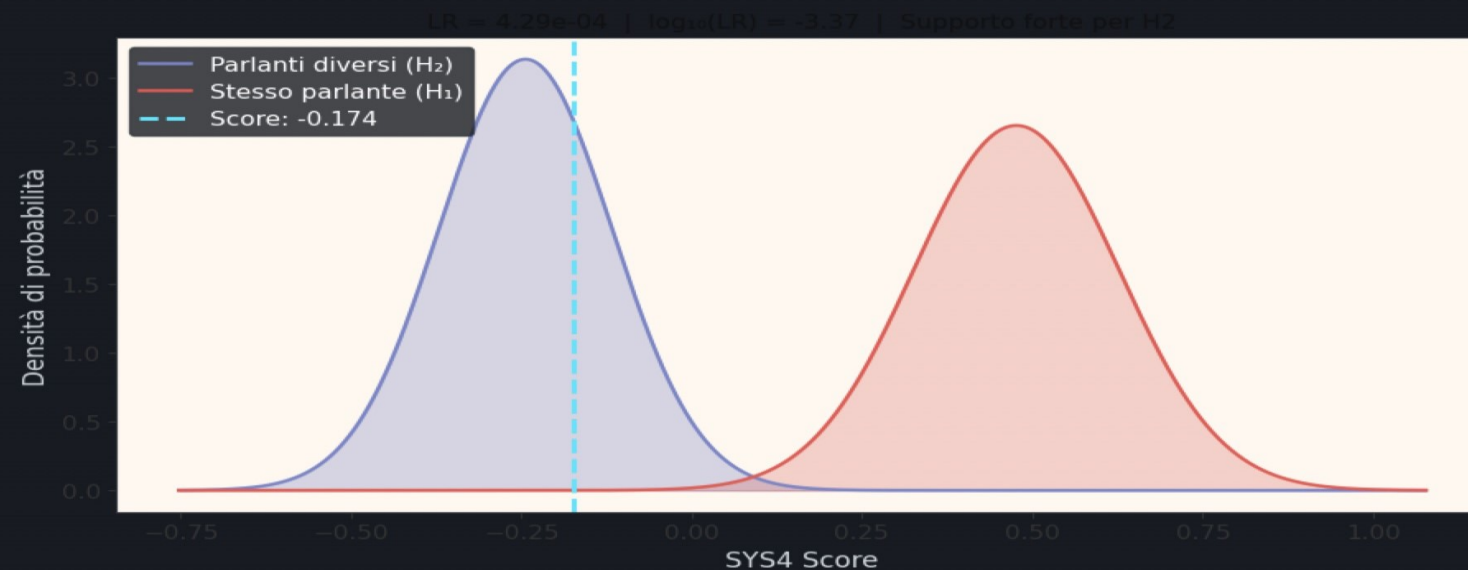
■ ENFSI -4 LR 0,0001 - 0,001

Supporto forte per H2

→ Probabilmente persone DIVERSE

■ ENFSI -5 LR 0,00001 - 0,0001

→ Molto probabilmente persone DIVERSE



Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

NASCE AMLETO

Il risultato espresso come Likelihood Ratio (LR) e interpretato secondo la scala ENFSI a 13 livelli da 'supporto estremamente forte per H2' a 'supporto estremamente forte per H1'.

Il sistema è stato sviluppato con piena trasparenza del pipeline, corpus noti, calibrazione documentata e riproducibile, che costituisce un requisito fondamentale per la verificabilità del risultato in sede processuale

PROVE PRATICHE CON AMLETO COME FUNZIONA

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

A - Voce Campione (Known)



Seleziona file audio

B - Voce Incognita (Questioned)



Seleziona file audio

TIPO DI AUDIO

Popolazione di Riferimento



✓ Popolazione di riferimento caricata

▶ AVVIA COMPARAZIONE

✓ Modello pronto. Carica due file audio e avvia la comparazione.

RISULTATO LR

 $\log_{10}(\text{LR}) = -$

Cosine sim = -

Avvia una comparazione per vedere la spiegazione.

SCALA ENFSI

ENFSI +6 LR > 100.000
Supporto estremamente forte per H1
→ Quasi certamente la STESSA persona

ENFSI +5 LR 10.000 - 100.000
Supporto molto forte per H1
→ Molto probabilmente la STESSA persona

ENFSI +4 LR 1.000 - 10.000
Supporto forte per H1

Densità di probabilità

In attesa di comparazione...

SYS4 Score

Come si arriva al risultato?

TRASPARENZA

Quali dati riceve il sistema?

Quali trasformazioni applica?

Quali parametri usa per produrre il LR?

Come può un altro perito controllare il calcolo?

La documentazione deve rendere verificabili questi passaggi.

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

La trasparenza del metodo

DOCUMENTAZIONE

Documentare audio, modello e preprocessing.

Rendere esplicite formule e costanti del "dietro le quinte di Amleto».

Descrivere i dati di calibrazione.

Consentire una verifica indipendente adeguata al quesito.

La sola disponibilità del risultato numerico non basta.

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

L'idea: **un motore che chiunque possa controllare**

AMLETO nasce per rendere ogni passaggio — dal file audio al Likelihood Ratio — documentato, calcolabile a mano, e verificabile da un esperto terzo.

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

Come funziona: la pipeline a 3 stadi

ECAPA-TDNN · SYS4 · CALIBRAZIONE

1

Embedding

Rete ECAPA-TDNN (SpeechBrain), fine-tuned su decine di migliaia di file telefonici italiani produce un vettore a 192 dimensioni per ogni registrazione.

Marco Perino · Sonia Cenceschi

2

Normalizzazione SYS4

«Quanto queste due voci somigliano anche alle altre voci del gruppo di riferimento?»

3

Calibrazione LR

Modello Gaussiano same-speaker / different-speaker

Preparazione dell'audio

01

Selezione

Segmenti pertinenti e parlato utile.

02

Decodifica

Il contenitore viene trasformato in campioni audio.

03

Ricampionamento

Il segnale viene portato a 16 kHz se necessario.

04

Mono

Se multicanale, la traccia viene mediata.

Prima del confronto

- Audio a 16 kHz, mono. Serve ad adattare il file all'ingresso atteso dal modello.
- Controllo di durata, qualità e parlato utile.

Selezione dei file A e B

- A: voce campione nota.
- B: voce da confrontare.
- AMLETO controlla che file, modello e popolazione di riferimento siano disponibili.
- Formati selezionabili: WAV, MP3, M4A, FLAC e OGG.

Ricampionamento a 16 kHz

- 44,1 kHz o 48 kHz diventano 16 kHz.
- Anche un segnale a 8 kHz viene portato a 16 kHz.
- Se è già a 16 kHz, il passaggio viene saltato.

Conversione dei canali in mono

- Se il file ha più canali, AMLETO ne fa la media.
- Per uno stereo: $\text{mono} = (\text{sinistro} + \text{destro}) / 2$.
- Il segnale prosegue in memoria a 16 kHz, mono.
- Se due interlocutori occupano canali distinti, il campione va preparato prima.

Estrazione delle caratteristiche

- Il segnale passa a SpeechBrain.
- Prima vengono ricavate le caratteristiche acustiche.
- Poi viene applicata la normalizzazione prevista dalla configurazione del modello.
- Questa normalizzazione è distinta dal volume audio.

Confronto e risultato

similarità coseno → score SYS4

si confrontano le due voci con un gruppo di voci di riferimento e si toglie quanto già somigliano a quel gruppo in generale.

E come correggere un voto con la media della classe. Ciò che resta è lo **score SYS4**.

SYS4 - Corregge il punteggio di somiglianza tra due voci, togliendo quanto ciascuna somiglia già a tutte le altre.

score SYS4 → calibrazione gaussiana

Lo score da solo non dice niente: 0,403 è tanto o poco? Per capirlo servono due “campane” costruite su confronti già noti.

calibrazione → $\log_{10}(\text{LR})$ trasformato

.

Il LR esprime il peso dell'evidenza rispetto a due ipotesi.

Due livelli di verifica indipendente

DISCLOSURE PROGRESSIVA, NON TUTTO-O-NIENTE

LIVELLO 1

Calcolo manuale

Calcolatrice + 4
parametri Gaussiani
dichiarati.

Chiunque ricalcola l'LR
dato lo score.

LIVELLO 2

Script di verifica

Script indipendente con
libreria standard.
Verifica a partire da
score già prodotti.

Due livelli di replica

CHE COSA SI VERIFICA

- 1 Calcolatrice: dallo score SYS4 al risultato numerico.
- 2 Script indipendente: stesso calcolo su più score, con controlli degli intermedi e degli arrotondamenti.

Livello 1 · Che cosa riceve il perito

PRATICAMENTE · DATI DI PARTENZA

Un perito di controparte riceve:

- lo score e i quattro parametri nella formula gaussiana.
- Istruzioni per calcolare il logaritmo in base 10 del rapporto delle due densità.

Potrà verificare il risultato a partire dallo score e dai parametri

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

Livello 1 · Che cosa riceve il perito

PRATICAMENTE · DATI DI PARTENZA

Può farlo con una **calcolatrice, livello 1,**

oppure

con uno **script indipendente, livello 2.**

In questa fase Non **servono gli audio.**

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

Calcolo manuale delle due densità

PRATICAMENTE · CALCOLATRICE

$$\ell(\mu, \sigma) = -\ln(\sigma) - \frac{1}{2} \ln(2\pi) - \frac{1}{2} [(s-\mu)/\sigma]^2$$

$$\begin{aligned}\ell_{SS} &= -\ln(0,1502) - 0,918939 - \frac{1}{2}(-0,0734/0,1502)^2 \\ &= 0,857444\end{aligned}$$

$$\begin{aligned}\ell_{DS} &= -\ln(0,1271) - 0,918939 - \frac{1}{2}(0,6468/0,1271)^2 \\ &= -11,804641\end{aligned}$$

$$x = [0,857444 - (-11,804641)] / \ln(10) = 5,499074$$

ℓ è il logaritmo naturale della densità. x è il $\log_{10}$ del LR gaussiano.

Trasformazione e risultato finale

PRATICAMENTE · CONFRONTO DEGLI OUTPUT

$$x = 5,4990738816$$

$$y = 4,0021623965$$

$$R = 10^{4.0021623965} = 10.049,91518$$

Il perito confronta x , y e R con gli output dichiarati.

Esito dell'esempio: circa 10.050.

Che cosa consegnare al verificatore

- Audio e segmenti analizzati.
- Checkpoint e cohort, quando necessari.
- Parametri, formule e versioni software.
- Hash, configurazione e risultati intermedi.

Il perito conserva un ruolo centrale

- Valuta idoneità dei campioni e ipotesi.
- Motiva la popolazione di riferimento.
- Esamina risultati discordanti e limiti.
- Integra il confronto con le altre analisi.
- Il risultato automatico contribuisce alla valutazione tecnica.

Validazione interna: che cosa sappiamo

STATO DELLA DOCUMENTAZIONE

$EER = 1,86\%$ $Cllr = 0,1294$ $d' = 5,195$

Marco Perino · Sonia Cenceschi

10° Convegno Nazionale ONIF — Genova, 16 ottobre 2026

Le metriche: quale domanda rispondono?

COME LEGGERE I RISULTATI

EER 1,86%

Alla soglia di equilibrio, falsi allarmi e mancate accettazioni si equivalgono;
È il punto in cui il sistema sbaglia allo stesso modo nei due sensi — tante volte dice "stessa persona" quando non lo è, quante volte dice "persone diverse" quando lo è. Più basso = meglio.

Cllr 0,1294

Valuta i LR considerando discriminazione e calibrazione; non misura solo se il sistema distingue bene le voci, ma se i *numeri* che dà (il LR) sono onesti — né troppo sicuri di sé né troppo timidi. Più basso = meglio

$d' = 5,195$

quanto sono "lontane" le due campane statistiche (stesso parlante vs parlanti diversi) — più sono separate, più è facile distinguerle senza sovrapposizioni. Più alto = meglio.

COLLOCHIAMO AMLETO

STRADA 1

LR calcolato

Software + popolazione di riferimento + modello statistico. Esce un numero.

Vale solo se valgono le sue premesse.

**AMLETO si
posiziona qui**

STRADA 2

LR verbale

Scala verbale calibrata e giudizio esperto strutturato.

Nessun modello statistico richiesto.

Legittimo e molto diffuso (Willis et al. 2015; ISO 21043).

IL GUAIO

In aula diventano una cosa sola

- Il LR calcolato viene spesso confuso con quello verbale.
- Senza numero, per molti la perizia non vale niente. Peccato che spesso quel numero non regga: calcolato su premesse che nel caso concreto non ci sono.

Il LR non è la probabilità che il sospettato sia il parlante! è importante continuare a ripeterlo.

I LIMITI DEL LR NELLA PRATICA Problema 1: pere e banane

Durante la conferenza IAFPA 2026 la comunità internazionale si è confrontata

International Association for Forensic Phonetics and Acoustics

Un LR calcolato presuppone una popolazione di riferimento rappresentativa e statisticamente robusta.

Se non c'è, confrontiamo una pera con un cesto di banane

«le voci sono diverse». Diverse da cosa?

Vale anche per AMLETO: il corpus è telefonico italiano, non è «tutta l'Italia»

Per questo servono spirito critico e molteplicità di metodi

SE C'È

«Allora è scientifico»

Il numero sembra dare oggettività e validità in automatico.

Non sempre viene richiesto come è stato ottenuto.

SE MANCA

«C'è qualcosa che non va»

Giudici, avvocati e anche alcuni periti si aspettano di trovarlo.

La sua assenza può diventare un difetto, anche quando il caso non lo regge.

IL SOFTWARE

Non alza la mano

Se le premesse saltano, il sistema restituisce comunque un valore. Non può avvisarti.

Con le prove ambigue, un valore debole diventa «prova scientifica».

Morrison 2011 · Braun 2021

Nessuna misura è buona o cattiva in sé: vale finché valgono le sue premesse.
Il nostro lavoro è farle valere in più casi reali possibile.

AMLETO NON È IN VENDITA

AMLETO oggi è un proof of concept funzionante in evoluzione

CRESCE IN RETE

Una rete internazionale

Colleghi e laboratori di più paesi: nuovi corpora, nuove condizioni, nuovi occhi sul metodo.

DIVENTA SOLIDO

Raffinato, poi validato

Corpora più ampi, calibrazione più robusta e validazione esterna, fatta da chi non l'ha costruito.

RESTA REPLICABILE

Oltre la black box

La rete neurale resta una scatola nera, come ogni rete. Tutto il resto, dati, parametri e calcolo, è documentato e rifattibile.

Ma una rete, per costruire, ha bisogno di risorse.

APERTO, MA NON GRATUITO

Come possiamo lavorare insieme

FORZE DELL'ORDINE

Prove alla cieca

Con RIS e Polizia Scientifica: casi a esito noto scelti da loro, risultati confrontati con la verità.

Gratuite, regolate da convenzione.

TRIBUNALI, CTU E PERITI

Supporto alle perizie

Analisi con AMLETO, calcolo documentato, verificabile, firmato, affiancamento all'analisi fonetica e ad altri tool.

R&D

Studi finanziati

Progetti strutturati con università, enti e laboratori: protocollo, budget e risultati pubblicati.

Budget per ricerca e sviluppo.

WORK IN PROGRESS: I PROSSIMI PASSI

COSA FAREMO, CONCRETAMENTE

NEI PROSSIMI MESI

Documentazione allineata alle linee guida ENFSI

Prime prove alla cieca con chi aderisce

Presentazione di richieste di finanziamento
congiunte con rete internazionale

Validazione esterna e indipendente

Corpora più ampi: varietà regionali, canali, stili.

Non è una presentazione... è un invito!

Marco Perino · Sonia Cenceschi | Domande

COMING SOON

Estensione allo spagnolo...

GRAZIE!

Un software per calcolare,
più metodi per confrontare,
un perito, sempre, per risponderne.