



# Lecture.org

**Professoressa La Regina, Lei è autrice del libro *L'identificazione della voce nel processo penale* edito da CEDAM: qual è l'importanza della *speaker identification* nel nostro ordinamento?**

L'impulso ad un ricorso sempre crescente verso le investigazioni vocali si è progressivamente amplificato anche a causa del fatto che all'incremento di dispositivi per la comunicazione a distanza e delle forme svariate di comunicazione via *internet*, molto spesso, si accompagna l'utilizzo di tecniche volte a rendere anonima l'identità dell'utente. Si tratta di una problematica ben nota sia nel settore delle intercettazioni telefoniche, in cui si riscontra di frequente l'utilizzo di *sim card* acquisite con documenti contraffatti, o tramite prestanome, che nel campo delle intercettazioni ambientali, a contrasto delle quali si adottano contromisure che rendono impossibile procedere ad attività di osservazione e identificazione di soggetti che si trovano in luogo di monitoraggio. In questi casi, e più in generale laddove non sia possibile pervenire all'identificazione di un sospetto ricorrendo ad elementi ulteriori ed estrinseci del parlato, la traccia vocale registrata attraverso le intercettazioni può essere un punto fondamentale di partenza per una indagine.

A dimostrazione dell'importanza riconosciuta a questo settore delle investigazioni, del resto, si possono richiamare le sperimentazioni che investono il settore dell'indagine vocale. Tra queste si segnala il progetto di biometria vocale dell'Arma dei Carabinieri, esposto dal Generale Angelo Santo in occasione della conferenza sulla «Biometria vocale nelle investigazioni» che si è tenuta a Roma il 13 luglio 2017. Si tratta di un programma articolato su tre fasi. La prima fase ha comportato la selezione di alcuni nuclei investigativi da dotare di sistemi di riconoscimento c.d. 1 a molti. Una volta effettuato questo tipo di individuazione, che si ottiene attraverso un risultato di *score* in tutto simile a quello che restituisce l'AFIS (ad esempio: l'impronta vocale somiglia a quella di cinque persone), il progetto prevede il passaggio del risultato al Ris competente per territorio. Questo opera attraverso un sistema che procede ad effettuare il confronto c.d. 1 ad 1 e, dunque, consente di indicare quale tra le voci selezionate sia quella ricercata. La seconda fase, che si svilupperà nel corso del prossimo biennio, è finalizzata alla condivisione dei dati e alla verifica delle modalità di interscambio; si pensa alla creazione di un *repository* centrale presso il Ris di Roma nel quale convogliare e depositare tutte le impronte vocali che sono tratte dalle singole attività; sarà, poi, il Ris di Roma a mettere a disposizione i dati a favore degli altri reparti. L'obiettivo ultimo del progetto, la terza fase, prevede il coinvolgimento delle altre forze di polizia per l'adozione delle impronte vocali come ulteriore mezzo di identificazione personale e, dunque, per la creazione di una banca dati da adottare secondo il modello giuridico già previsto per la banca dati del DNA. Si tratta di un programma

complementare ad una ulteriore ricerca che, parimenti, vede coinvolta l'Arma dei Carabinieri. Si tratta del programma SIIP (*Speaker Identification Integrated Project*), un progetto finanziato dalla Commissione Europea che si prefigge di sviluppare un sistema per l'identificazione dei parlatori che, avvalendosi di un *data-base* in cui confluiscono chiamate intercettate, file audio-video e contenuti diffusi sui *social media* o su qualsiasi altro tipo di mezzo o canale di comunicazione, permetterà di superare il problema dell'utilizzo di identità nascoste o false da parte dei terroristi e dei criminali che utilizzano *Internet* per evitare di essere intercettati, identificati e monitorati.

### **Come si svolgono le investigazioni per la ricerca della traccia vocale?**

L'acquisizione della voce sospetta è solo il presupposto per l'attivazione di una indagine vocale perché essa, per approdare all'accertamento, presuppone la possibilità di comparare tale voce con quella dell'indagato. Pertanto, snodo essenziale di un percorso investigativo in cui si intenda valorizzare il reperto fonico per l'individuazione dell'autore del reato – snodo che generalmente si presenta anche nei casi in cui la traccia vocale non venga registrata ma resti custodita nella mente di un ricognitore- è quello che conduce all'acquisizione di un campione di voce, sicuramente appartenente a quella della persona sottoposta alle indagini, da confrontare con quella "bersaglio", per poter attribuire la paternità delle conversazioni captate o comunque ascoltate sulla scena del crimine. Da un punto di vista generale, le possibili tappe che conducono all'acquisizione di un campione vocale, sono due: la prima presuppone la collaborazione del parlatore, che – ove richiesto – può scegliere se prestare la propria voce per la predisposizione del saggio fonico. Può accadere, tuttavia, che l'interessato rifiuti di collaborare. L'ostruzionismo eventualmente frapposto all'esecuzione di una ricognizione, auditiva o all'esecuzione di un accertamento fonico è, tuttavia, aggirabile attraverso l'impiego di "sucedanei": così come nel caso in cui il riconoscimento investa il volto si può utilizzare una foto, nel caso in cui il riconoscimento riguardi la voce, si può utilizzare un saggio fonico già in possesso degli inquirenti, ad esempio perché acquisito nel corso di una intercettazione o di un interrogatorio documentato attraverso mezzi di riproduzione fonografica. Del pari, nulla preclude di intraprendere uno specifico itinerario investigativo volto all'acquisizione di una traccia vocale da utilizzare per l'accertamento fonico o per l'esecuzione di un riconoscimento vocale. Il primo pensiero va naturalmente alle intercettazioni, oggi peraltro consentite anche attraverso il captatore informatico, il quale darà sicuramente un notevole impulso allo sviluppo delle investigazioni vocali. Chiaro, tuttavia, che in questi casi – così come nella situazione in cui si ripieghi per l'individuazione fotografica – se non si pone un problema di legittimità dell'azione investigativa si apre comunque uno squarcio sul versante dell'attendibilità del risultato di tale azione. Così come è ben nota la fallacia dei riconoscimenti fotografici, statisticamente più elevata di quelli operati "dal vivo", ai fini della qualità del risultato di una perizia fonica o di una ricognizione vocale, è assolutamente determinante che l'accertamento o il riconoscimento avvenga su un segmento di parlato che riproduca le stesse parole e le medesime condizioni, non solo ambientali ma anche emotive.

## Quali sono i rischi connessi al riconoscimento vocale?

Quale che sia l'itinerario prescelto dagli organi investigativi, il percorso si mostra sempre carico di incertezze. Rispetto alle due strade che conducono all'emersione dell'identità vocale – quella che si affida all'innata capacità dell'uomo di riconoscere una percezione sensoriale e quella che, invece, si fonda sullo studio di connotazioni che siano peculiari di un individuo e come tali dotate di capacità individualizzante – quale che sia lo strumento prescelto, l'idea di fondo è sempre quella che la voce anonima possa essere comparata con la voce nota, dell'indagato e dell'imputato. Laddove questa operazione sia affidata interamente alle capacità mnestiche dell'individuo, la comparazione, e il conseguente riconoscimento, sono il prodotto di una serie di processi cognitivi, non sempre consapevoli, e del contesto fisico e psichico di riferimento. Qualora, invece, ci si affidi ad un esperto, la comparazione diviene una operazione tecnica che postula, o dovrebbe postulare, una elevata specializzazione ed il ricorso ad un metodo scientifico.

Non esiste un percorso privilegiato per l'accertamento, una tecnica migliore, una *best evidence*. Da un lato, infatti, l'approccio sensoriale restituisce un risultato gravido di insidie suscettibili di manifestarsi già nella fase percettiva della voce, la quale, per sua natura, non è caratterizzata dalla stessa precisione e dalla stessa determinatezza che connota l'omologo visivo. Del resto, sono basi magmatiche quelle su cui si fonda il fenomeno ricognitivo, le quali, inevitabilmente, determinano un elevato tasso di soggettivismo del riconoscimento, oltre che un consistente grado di fallibilità. Neppure la certezza che esprime il ricognitore nell'atto dell'identificazione può costituire un serio indice di affidabilità del risultato; per questo aspetto, infatti, nonostante diversi studi siano stati indirizzati ad approfondire il rapporto tra sicurezza nel riconoscimento e correttezza del medesimo, non è emersa alcuna evidenza che testimoni l'esistenza di una correlazione tra i due fattori. Anzi, è assolutamente diffuso il convincimento che, anche nel caso di familiarità con la voce da riconoscere, l'identificazione auditiva debba essere vagliata con particolare prudenza perché essa risente, ancor più del riconoscimento personale, di numerose variabili contestuali.

Dall'altro, l'indagine tecnica si incentra su un bioindicatore dotato di una capacità caratterizzante imperfetta. La voce a differenza di una impronta digitale o genetica, non è immutabile e non possiede caratteristiche sufficientemente univoche da consentire sempre e comunque di far emergere una identità dalla moltitudine indifferenziata. Il problema principale è connesso al fatto che la variabilità non si presenta solo a livello interindividuale – cioè da persona a persona – ma anche a livello intraindividuale, ovvero rispetto alla produzione vocale dello stesso individuo, la quale è soggetta a cambiamenti di breve termine, cioè legati a situazioni contingenti – lo stato d'animo, lo stato di salute, il fumo di una sigaretta, l'assunzione di una bevanda – che di lungo termine, questi ultimi dovuti al trascorrere del tempo.

Non si tratta solo di limiti derivanti da una fonte di prova cui è coesistente l'intrinseca incertezza. Non vi è un metodo univoco di indagine e non sussiste un preciso limite che

imponga di non utilizzare il materiale sonoro che superi alcuni limiti qualitativi. A differenza degli accertamenti aventi ad oggetto le impronte digitali, in materia di riconoscimento vocale non esiste neppure una soglia limite superata la quale il livello di compatibilità possa considerarsi un valore tendenzialmente assoluto. In altri termini, anche in questo settore, la compatibilità è sempre un valore relativo.

Quali metodi vengono adottati per il riconoscimento del parlatore?

A partire dagli anni '90, e ancora oggi, il metodo più diffusamente impiegato per il riconoscimento del parlatore è il c.d. metodo parametrico cioè una tecnica semi-automatica che studia la voce da un punto di vista fisico e, dunque, come sistema di onde acustiche. Si tratta di una indagine che non si concentra sull'intero segnale del parlato ma solo su determinate proprietà della voce ricavabili, in particolare, dalle vocali perché queste si ritiene che producano un suono misurabile con maggiore precisione. L'estrazione dei parametri di riferimento coinvolge essenzialmente le zone di maggiore concentrazione dell'energia - c.d. frequenza fondamentale (FF0) e frequenze formanti (FF1, FF2, FF3, FF4) - le quali presentano intrinseche caratteristiche di robustezza, poca variabilità inter-parlante.

L'analisi tecnica è articolata su tre diverse fasi operative. Innanzitutto, l'operatore deve scegliere il materiale fonico da utilizzare e, dunque, singole parole o intere frasi che posseggano determinate caratteristiche qualitative, in termini di rapporto segnale/rumore, larghezza di banda e durata; in secondo luogo, appositi programmi - che in Italia sono IDEM e SMART - isolano, nell'ambito del materiale selezionato, i parametri significativi per la caratterizzazione della voce; nella terza fase, infine, si effettua una interpretazione statistica dei dati ed un confronto tra le misure ottenute per stabilire la compatibilità tra la voce anonima e quella del soggetto noto.

Nonostante il progressivo ricorso a metodiche di riconoscimento sempre più sofisticate, anche dal punto di vista della tecnologia utilizzata, resta il fatto che, allo stato attuale della conoscenza, non vi è un sistema capace di cogliere, contemporaneamente, tutte le caratteristiche di una voce. Talvolta, poi, è lo stesso tipo di segnale sonoro che guida nella scelta dell'approccio da seguire perché presenta connotazioni tali da precludere un tipo di accertamento, consentendo invece di sperimentarne un altro. Tenendo conto di questi fattori, e della connotazione multidisciplinare del settore, la tendenza attuale è quella di ricorrere ad un approccio ibrido, denominato combined, che tiene conto sia degli aspetti strumentali che di quelli linguistici, per ottenere una analisi della voce più precisa e completa.

Come avviene l'identificazione del parlante?

Può avvenire o attraverso un accertamento di tipo tecnico (una perizia) che si avvale, come anticipato, di metodi automatici o semi-automatici, oppure attraverso una ricognizione vocale. Se nel primo caso una considerevole ipotesi sui risultati dell'analisi deriva dalla inesistenza di un metodo scientificamente riconosciuto come valido a fini identificativi, nel caso della ricognizione sono le stesse caratteristiche della memoria - e della memoria uditiva in particolare - a suggerire una valutazione decisamente cauta dei risultati del riconoscimento. Il discorso è in parte sovrapponibile a quello che si effettua in materia di ricognizione personale perché identici sono i passaggi cognitivi che caratterizzano lo scandire dei relativi processi mnestici. Così, se rispetto alla percezione visiva è da escludere che si compia una attività di mero immagazzinamento di informazioni fedeli all'impulso ricevuto, lo stesso è a dirsi rispetto alla percezione acustica la quale - lungi dall'essere unicamente un fenomeno di tipo fisico-meccanico, legato ad onde con frequenze che comprimono ed espandono l'aria provocando la vibrazione della membrana timpanica - implica sempre una

propria e personale interpretazione di un segnale, al quale vengono impresse caratteristiche soggettive dipendenti dalle modalità di percezione del suono stesso. La percezione uditiva è, dunque, un fenomeno complesso, in cui le caratteristiche fisiche del suono ricevuto, quelle fisiologiche dell'orecchio, e l'attività neurale del cervello, interagiscono non per restituire una immagine fedele del suono ricevuto, ma piuttosto una sua complessa – e non necessariamente fedele – elaborazione. Elaborazione che, del resto, caratterizza anche le successive fasi della conservazione e del recupero di un ricordo che rimane oggetto di una manipolazione cognitiva, amplificata in presenza di una molteplicità di fattori anche indipendenti dalla percezione rimasta impressa sull'earwitness.

Se la questione legata alla fisiologica attitudine creativa accomuna il processo di ricognizione quale che sia il materiale mnestico a venire in rilievo, i problemi legati alla valutazione di attendibilità si accentuano nei casi in cui oggetto del riconoscimento sia una voce, perché la memoria uditiva è più difficile da elaborare storicamente e, per questo, finisce per ancorarsi a basi assolutamente magmatiche. Significativi sul punto sono i risultati di un esperimento che venne condotto facendo assistere un gruppo di trenta persone alla proiezione di un breve film, seguita da una serie di domande tra le quali ne era stata inclusa una con cui si chiedeva cosa avesse detto il protagonista principale in una determinata situazione. Nonostante in quel contesto il protagonista non avesse affatto parlato, ventitré dei trenta partecipanti ricordavano non solo di averlo sentito parlare ma anche ciò che aveva detto.

Anche sotto il profilo della tenuta cronologica la percezione auditiva è profondamente più fragile di quella visiva. Il ricordo di un volto nel lungo periodo mantiene dei margini di attendibilità maggiori, potendo restare integro almeno nell'arco dei primi mesi; il ricordo di una percezione vocale, al contrario, è significativamente più esposto all'usura del tempo, tanto da essere soggetto a decadimento anche nel breve periodo. Così, ad esempio, misurando il decremento di accuratezza cui è esposto nel tempo il ricordo di una traccia vocale, è emerso che appena il 50% delle identificazioni erano esatte dopo una settimana; il 43% dopo due settimane e solo il 9% dopo tre settimane.

**Katia La Regina è Professore Associato di Diritto processuale penale presso l'Università degli Studi Giustino Fortunato di Benevento**