Poliziamoderna

A tutt’orecchio

Come le impronte digitali anche la voce di una persona è unica: saperla riconoscere e riuscire a individuare i “parlatori” di una conversazione assumono una grande importanza a livello investigativo. Di questo si occupa l’area indagini foniche della Sezione indagini elettroniche, con a capo il direttore tecnico capo Giovani Tessitore, incastonata all’interno della IV Divisione del Servizio polizia scientifica. «È un settore che potrebbe sembrare un po’ desueto nell’ottica del nostro processo di digitalizzazione – sottolinea Lorenzo Rinaldi, primo dirigente tecnico ingegnere della Polizia di Stato responsabile della IV Divisione – invece non lo è affatto, perché quello che è cambiato negli anni è stato il mezzo attraverso il quale la voce viene veicolata, ma non certamente l’importanza, nel quadro delle attività investigative, delle indagini sulla voce». Ma quando nasce l’esigenza del riconoscimento del parlatore? Il laboratorio per le indagini foniche è stato introdotto nella polizia scientifica verso la fine degli Anni ’70 come ci racconta il sostituto commissario Giuseppe Feliciani che ha passato tutta la sua carriera in questo ufficio: «Con il sequestro Moro, l’inizio del periodo degli Anni di piombo, delle telefonate estorsive, dei rapimenti, i dirigenti di allora ebbero l’intuizione di creare un’attività investigativa per la voce sulla falsariga delle impronte digitali». Da quel momento sono iniziati gli studi e le ricerche per trovare gli strumenti e le metodologie più efficaci. Ricerche che sono continuate e si sono sviluppate negli anni con l’avvento di sempre nuove tecnologie e nuovi sistemi.

Quella del riconoscimento del “parlante” è una delle attività dell’indagine fonica, ma non l’unica; un altro lavoro importante che viene svolto è quello della pulizia dei segnali. «Il primo step della filiera dell’accertamento prevede l’acquisizione e il miglioramento dell’audio, il cosiddetto filtraggio – spiega Rinaldi – Fondamentale è il mezzo con il quale l’audio viene registrato: ogni strumento o applicazione ha una sua banda di acquisizione, per problemi di compressione e trasmissibilità del file non memorizzano tutti lo stesso segnale; questo non influisce sulla nostra capacità di comprendere il contenuto dell’audio, ma cambia enormemente la capacità di confrontarlo». L’attività di filtraggio è importantissima nel corso delle indagini e viene richiesta di continuo sia dagli uffici investigativi che da quelli di prevenzione e non è finalizzata esclusivamente a individuare il soggetto che parla ma anche a capire cosa dice. Scoprire questo richiede un’attività non semplice perché non è facile distinguere le parole in un’intercettazione ambientale il cui audio è complesso. Infatti sia che venga registrata in un’automobile o in un luogo chiuso come la casa c’è sempre molto rumore di fondo. La microspia, poi, può essere stata posizionata dietro qualcosa che produce rumore; o magari c’è un segnale che si sovrappone come la radio o la televisione. «Per “pulire” l’audio e capire cosa hanno detto – continua il primo dirigente – utilizziamo delle tecniche specifiche. Ultimamente abbiamo acquisito un software eccezionale, sviluppato a livello internazionale, che consente il filtraggio attivo degli audio; ad esempio avevamo una conversazione, che ritenevano fondamentale per dare una svolta a un’indagine, che si svolgeva con la televisione accesa. Il primo passo dei tecnici è stato quello di capire cosa si sentisse oltre le voci. Poi sono riusciti a individuare quale fosse il programma televisivo e a sottrarre l’audio al segnale, togliendo quello che noi chiamiamo “rumore” lasciando solo le voci che ci interessavano».

Rinaldi ha parlato di gruppi internazionali che anche in questo settore sono preziosissimi per lo scambio delle esperienze e l’aggiornamento continuo. «Sono tornato da poco dall’Armenia – racconta il commissario capo tecnico ingegnere Giacomo Rogliero che si occupa delle indagini foniche – dove c’è stato l’ultimo meeting dell’*European network of forensic science institutes* (*Enfsi*) la rete europea degli istituti di scienze forensi, fondata nel 1995 per facilitare il dialogo tra i professionisti del settore in Europa. Ci siamo confrontati sull’analisi e il miglioramento dei file audio e dell’individuazione del parlatore. In queste occasioni vengono messe a fattore comune, tra tutti gli istituti forensi europei che fanno parte di questo network, una serie di nuove metodologie e tecniche o si espongono nuove problematiche. Questa volta sono stati presentati anche dei casi studio su come viene modificata la voce portando la mascherina, sia chirurgica che FFP2, perché chiaramente nell’epoca post Covid sono aumentati i casi di registrazioni di persone che la indossano. È emerso che sicuramente la voce viene in parte modificata, perché la mascherina l’attenua sia in termini di potenza che di frequenza; i risultati sono stati però confortanti e, anche se con una maggiore incertezza, il parlatore risulta comunque individuabile».

L’attività più importante delle indagini foniche è quella del riconoscimento del parlatore. Ma come si fa a stabilire se una voce appartenga proprio a quella persona? Non è un’impronta digitale, ma comunque porta con sé delle caratteristiche individuali, che dipendono anche dall’anatomia e che

possono variare con lo stato di salute, l'età e l'umore. «Si riconosce un parlatore da un altro andando a fare l'analisi fonica delle frequenze, chiamate formanti, delle vocali – spiega Rinaldi – Questo perché quando si pronuncia una vocale all'interno della cavità orale l'aria entra in oscillazione, vibra e porta con sé un'informazione utile che può essere estrapolata con un'analisi spettrale da cui si vedono le frequenze caratteristiche di una persona. Nelle consonanti questo invece non succede perché non si instaura nessuna oscillazione dell'aria. Nel 2003 – continua Rinaldi – dopo l'arruolamento fui assegnato alla Scientifica e il mio primo incarico fu proprio la fonica. Arrivai e sentii nei corridoi: "A… A… A… E… E… E…": gli operatori ripetevano nei microfoni tutte le vocali per creare una banca dati. Infatti il passo successivo per l'individuazione del parlante è stabilire quanto le sue "vocali" si discostino da quelle di altre persone e per questo abbiamo sviluppato nel tempo dei database con molte voci divise a seconda della provenienza geografica». «Affinché il risultato sia affidabile la popolazione di riferimento deve essere il più possibile omogenea alle voci messe a confronto – specifica Rogliero – Per fare un esempio, se si confrontano due voci maschili italiane, bisogna farlo rispetto a una popolazione di riferimento che è composta da un certo numero di voci maschili italiane. Questo permette all'algoritmo sia di calcolare una sorta di similarità tra le voci, quindi l'anonimo e il sospettato, ma anche di andare a definire quanto sia distante la voce del sospettato rispetto a una popolazione di riferimento, la cosiddetta "tipicità" di quella voce per dare il risultato in termini di rapporti di verosimiglianza». Quindi, quando viene chiesto ai tecnici di riconoscere il parlante da un audio la prima cosa che chiedono è di ascoltarlo. Questo per due motivi: per vedere la qualità della registrazione ma soprattutto per sapere quanto è lungo e qual è la quantità di parlato. Per poter almeno tentare di conoscere un parlatore, è necessario avere un audio di discreta qualità, sufficientemente lungo e con un certo numero di vocali.

E se si ha un audio in cui un sospettato parla in una lingua diversa e quello noto da confrontare in italiano? «Abbiamo fatto dei test dai quali è emerso che gli algoritmi automatici riescono a riconoscere la persona anche se non parla nella sua lingua nativa – spiega Rogliero – Chiaramente c'è un margine di errore maggiore, ma il software è *language independent*. Abbiamo fatto anche altri studi per capire quanto influisca il parlato spontaneo oppure il testo letto quando si tratta di una lingua straniera e abbiamo visto che nel testo letto i risultati sono migliori rispetto al parlato spontaneo. Il software automatico riesce a riconoscere il timbro, quindi la persona, anche se non parla nella sua lingua nativa e i risultati sono accettabili ai fini delle indagini. Cosa diversa, invece, per l'analisi semiautomatica dove c'è una problematica maggiore nell'estrapolazione delle vocali». Software automatico e analisi semiautomatica, in cui si usano delle strumentazioni per analizzare il segnale, ma la parte di confronto viene lasciata all'operatore, sono le due metodologie utilizzate per riconoscere il parlatore. Spesso vengono usate entrambe per avere la conferma dei risultati ma naturalmente ognuna ha un suo ambito di utilizzo.

Un altro quesito che viene posto di frequente ai tecnici è se in un audio venga detta una determinata parola fondamentale per le indagini o di riconoscere altri rumori. «In un'intercettazione ambientale che ci hanno sottoposto – ricorda Rinaldi – si sentiva un rumore che l'investigatore pensava potesse essere l'estrazione e il caricamento di un'arma e ha chiesto a noi di confermarlo e magari, in caso positivo, anche di stabilire quale arma fosse. Abbiamo iniziato a fare delle prove: l'indagine restringeva il campo a delle armi che abbiamo nel museo della sezione balistica e quindi siamo saliti su un'automobile e abbiamo registrato i rumori delle varie armi, ma anche dell'apertura e chiusura delle cinture di sicurezza, del vano potaoggetti, del posacenere, del finestrino, dello spostamento dello specchietto retrovisore, insomma tutti gli eventuali altri suoni che potevano esserci all'interno dell'abitacolo. E successivamente abbiamo eseguito le comparazioni che ci hanno permesso di confermare l'uso delle armi e di definirne il tipo».

Uno dei motivi per cui le analisi foniche vengono svolte dalla polizia scientifica e non dagli investigatori è perché spesso chi ascolta è condizionato dalle informazioni che conosce, in questi casi si parla di "miraggi acustici". Quindi questa è un'attività tecnica che va svolta in maniera asettica: l'operatore non deve farsi influenzare dalle informazioni dell'investigatore perché il risultato potrebbe essere non attendibile. «Anni fa abbiamo fatto dei test in collaborazione con la "Fondazione Ugo Bordoni" proprio per fronteggiare questo problema dei miraggi acustici – spiega Feliciani – i risultati sono stati inquietanti. Ci hanno sottoposto cinque file di cui uno era soltanto una modulazione del rumore. Non c'erano frasi. Bene, tutti e dieci, nonostante fossimo operatori con esperienza, abbiamo sentito delle parole anche in quella circostanza. Il cervello, in un contesto particolare, cerca di associare qualsiasi cosa a quel che già conosce». Alla base del miraggio acustico ci sono due variabili spiega il commissario capo Rogliero: «Una è il condizionamento, così detto *bias*, e l'altra è la bassa qualità dell'audio. Le due variabili insieme sono molto pericolose perché il condizionamento predispone l'ascoltatore a sentire quello che gli viene detto e la bassa qualità crea ancora più confusione. Per questo, facendo riferimento alle *best practice* adottate a livello europeo, quello che si cerca di fare è stabilire un limite, in gergo tecnico si dice "rapporto segnale-rumore", ossia quanto segnale utile abbiamo rispetto alla quantità di rumore; tanto più alta è la qualità del segnale, tanto più intelligibile è quello che ascoltiamo. Quindi si adotta una soglia minima, al di sotto della quale si preferisce non

procedere, proprio per evitare il problema del miraggio acustico». Questa è una delle principali difficoltà nelle trascrizioni, attività che normalmente espletano gli investigatori e a volte anche la Scientifica per verificarne l'attendibilità: «Un nostro collega ormai in pensione, Stefano Delfino, qualche anno fa è riuscito a dimostrare che la trascrizione di un consulente di parte era completamente errata – ricorda Feliciani – Secondo il consulente in una conversazione si sentiva una persona che affermava che il colpevole fosse un certo soggetto. Se fosse stata detta quella frase, attraverso le misure fatte con le strumentazioni sarebbero dovute apparire delle vocali, che come abbiamo detto producono un'oscillazione delle corde vocali. In quella circostanza Delfino dimostrò che non c'era niente, non c'era una vocale, quindi la registrazione non conteneva una conversazione ma semplicemente rumore, e il consulente aveva sentito quello che voleva sentire».

Le indagini foniche sono iniziate, come abbiamo detto, per le telefonate anonime dei sequestri: qualcosa che sembra lontano nel tempo, ma in realtà è solo cambiato il contesto. Oggi abbiamo un sistema di comunicazioni assolutamente favorevole a questo tipo di indagini, perché tutti lasciano quantità immense di vocali su app di *instant messaging*. «La sfida più grande è quando ci chiedono di confrontare un audio su WhatsApp con un'intercettazione ambientale – spiega Rinaldi – è molto complicato, perché le applicazioni tagliano, comprimono il segnale e quindi la diversità di canali e di caratteristiche della voce registrata porta con sé una maggiore incertezza. Ecco perché si cerca sempre di confrontare voci che provengono dalla stessa sorgente. Qualora non fosse possibile si procede con l'analisi, ma è chiaro che c'è una maggiore incertezza di misura».

Una recente preoccupazione per gli operatori viene dal mondo dei *fake*. «Per il momento, per fortuna, casi di analisi di voci sintetizzate non ce ne sono capitati – sottolinea Rinaldi – perché c'è un controllo alla fonte del dato da analizzare; di solito i file arrivano da intercettazioni o comunque da canali noti, ma sicuramente nel prossimo futuro i casi aumenteranno». I *deep fake* vengono generati con l'uso di algoritmi di intelligenza artificiale che riescono a riprodurre, quindi a falsificare una voce. Dal punto di vista dell'ascolto, l'orecchio umano non riesce a percepire la differenza tra quella falsificata e la vera. «Questa situazione adesso è da monitorare – spiega Rogliero – quello che è emerso nell'ultimo meeting europeo, è che si stanno rivolgendo sempre di più gli studi e gli approfondimenti sull'analisi dei *deep fake*. Fra poco ci verrà chiesto di capire chi parla ma soprattutto di capire se la voce registrata appartiene veramente a quel personaggio conosciuto. Immaginiamo un politico a cui viene attribuita una frase che può incidere sul risultato elettorale; ormai è talmente facile riprodurre la voce di una persona soprattutto nelle fonti che provengono dal Web. L'orecchio umano non riesce a distinguere la differenza ma ci auguriamo che l'analisi fonica, con le tecnologie che si stanno studiando, ci permetterà di capire se quella è una voce sintetizzata o no. Questa è la sfida del futuro».

Rapporto di verosomiglianza - la scala Enfsi In contesto giudiziario per ottenere valutazioni obiettive delle prove in cui si chiede un'identificazione personale, ottenuta confrontando una caratteristica biometrica (voce, volto, ecc.), i metodi generalmente utilizzati implicano l'estrazione di parametri che possono essere codificati e confrontati. Queste valutazioni richiedono interpretazioni statistiche appropriate: *Enfsi* da venti anni promuove l'utilizzo dell'approccio Bayesiano per determinare la così detta forza dell'evidenza di un accertamento, con la quale vengono valutate le probabilità delle prove sia nel caso dell'ipotesi accusatoria (voce appartenente allo stesso soggetto) sia nel caso dell'ipotesi difensiva (voce appartenente a soggetti differenti). Tale metodologia ha il vantaggio di essere trasversale a tutte le branche delle scienze forensi. In questo ampio lavoro dell' *Enfsi* il tema centrale è l'utilizzo del rapporto di verosimiglianza (LR) che è il rapporto fra la probabilità che un dato riscontro sia riconducibile al "nostro" sospettato (similarità) e la probabilità che lo stesso riscontro possa appartenere a una qualsiasi altra persona diversa dall'indagato del caso in esame (tipicità). In tale contesto risulta essenziale avere una popolazione di riferimento (di voci nel caso di specie) con la quale confrontare le caratteristiche del sospettato. Se le due quantità, similarità e tipicità, si equivalgono l'esame è inconcludente non favorendo alcuna delle due ipotesi in esame - accusatoria e difensiva. Salendo nella scala (verso il verde) la probabilità che l'evidenza in esame supporti l'ipotesi accusatoria (stesso soggetto) è sempre più alta; scendendo (verso il marrone) è sempre maggiore supporto verso l'ipotesi che siano due persone differenti. Questa scala di valutazione viene ormai riconosciuta in tutta l'Unione europea.

04/12/2023