



UNIVERSITA' DEGLI STUDI DI GENOVA
FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN MEDICINA E CHIRURGIA

Tesi di Laurea

STUDIO DEI PARAMETRI STRUMENTALI CARATTERIZZANTI LA MIMESI VOCALE

Relatore

Prof. Alberto Ballestrero

Correlatrice

Dott.ssa Beatrice Travalca Cupillo

Candidato

Gabriele Poggi

Matricola 3373943

Anno accademico 2020-2021

STUDIO DEI PARAMETRI STRUMENTALI CARATTERIZZANTI LA MIMESI VOCALE

INDICE

INTRODUZIONE.....	2
1. MIMESI VOCALE.....	2
1.1 MIMESIS E IMITAZIONE: definizione.....	2
1.2 DISTINZIONE TRA MIMICA E IMITAZIONE.....	4
1.3 ESISTE UNA PEDAGOGIA DELL'IMITAZIONE VOCALE?	5
2. IMPRONTA VOCALE E SUA IMITAZIONE.....	6
2.1 IMITAZIONE E RICONOSCIMENTO VOCALE.....	9
CONCLUSIONI.....	16
BIBLIOGRAFIA.....	18

INTRODUZIONE

L'imitatore, specialmente se professionista, appartiene ad una tipologia di artista, che da sempre affascina, diverte ed incuriosisce. Egli esprime una vera forma d'arte facendoci percepire, attraverso la modificazione della sua vocalità e non solo, la sensazione che, a parlare piuttosto che cantare, sia un'altra persona che, se a noi nota, identifichiamo perfettamente.

Si tratta di un'arte, corredata da varie sfumature, esistente da moltissimo tempo, che spesso ci ha visto fruitori ma della quale poco si parla e poco si conosce.

Scopo di questa tesi è revisionare la letteratura al fine di evidenziare se siano stati individuati parametri vocali strumentali che caratterizzino per l'appunto quella che viene definita la mimesi vocale.

1.MIMESI VOCALE

Il tema della mimesi e dell'imitazione, dal punto di vista della vocologia artistica, viene affrontato per la prima volta in Italia in occasione del X Corso Internazionale Foniatria e Logopedia, Voce Artistica nel 2015, di seguito si riporta quanto rappresentato in tale contesto da F. Fussi (7)

1.1. MIMESIS E IMITAZIONE: definizione

Rispetto alla gestione neutrale dello speech, sappiamo che la nostra voce può essere modificata sia agendo sul vocal tract per modificare il timbro e la proiezione, sia agendo sulla sorgente glottica, ad esempio modificando i registri. Potremmo definire ogni tecnica

vocale di canto un travestimento della voce naturale.

A volte il travestimento che si realizza nell'imitazione è tale da rendere difficile la distinzione.

I termini “imitazione” e “mimica” derivano dal greco *mimesis*, termine ampiamente utilizzato da Platone e Aristotele e da allora mai più scomparso dalla storia dell'estetica e del pensiero.

La parola “imitazione” ha un significato tecnico di tipo estetico, cioè arte intesa come imitazione, ma è anche una parola utilizzata correntemente nell'ambito di tutti i giorni.

Il termine “mimica” ha un utilizzo corrente molto più ristretto e viene in genere adoperato come termine tecnico nella critica teatrale e cinematografica per esprimere gli atti espressivi a carattere motorio (mimica corporea, mimica facciale dell'attore).

Il docente di mimica ribadisce spesso che l'atto mimico è molto diverso dall'atto imitativo: mimare non è imitare. L'imitazione, di fatto, non concerne unicamente gli artisti, ma è tanto comune quanto variegato in tutti i processi di apprendimento: ad esempio il bambino impara il linguaggio tramite il processo di imitazione.

La stessa terapia logopedica del disfonico cerca di far eseguire al paziente anche esercizi vocali di imitazione.

Ogni parlante modifica gradualmente il proprio “habitus” fonatorio e discorsivo per adattarsi a quello degli altri (per esempio il tipo di lessico, la sintassi regionale, la prosodia vocale). Questo avviene secondo il principio di allineamento interattivo della conversazione, di cui la convergenza inconscia (o automatica) è parte.

L'imitatore di professione ha invece l'obiettivo di convincere che sia un'altra persona a parlare o cantare, evocandone le caratteristiche per essere scambiato con l'oggetto della sua imitazione, e mima con i gesti il corpo, la voce, il trucco, quelli che sono i caratteri essenziali per essere assimilato al personaggio interpretato, pur tuttavia senza perdere del tutto la sua identità. In fondo è proprio la non totale assimilazione a creare il risultato artistico nell'arte dell'imitare.

La tecnica del grammelot viene usata dagli attori con il fine di farsi comprendere anche senza sapere articolare frasi di senso compiuto in una lingua straniera, oppure per mettere in parodia parlate o personaggi stranieri con uno pseudo-linguaggio che acquisisce un surplus di espressività musicale in grado di comunicare emozioni e suggestioni. Pare che questo artificio recitativo fosse utilizzato da giullari e compagnie di comici dell'arte che recitavano usando intrecci di lingue e dialetti misti a parole inventate, affidando alla gestualità e alla mimica quel tessuto connettivo che rendeva la comunicazione possibile a prescindere dalla

lingua parlata dall'uditorio. Un esempio di grammelot cinematografico è rappresentato dal monologo di Charlie Chaplin nel film Il Grande Dittatore.

1.2. DISTINZIONE TRA MIMICA E IMITAZIONE

Distinguiamo l'imitazione, nella quale A imita l'oggetto B producendo C, dalla mimica nella quale A mima l'oggetto B producendo C

Mentre nel caso dell'imitazione l'oggetto C è prodotto coscientemente al fine di rassomigliare all'oggetto B in modo pressoché indipendente dal soggetto A (quindi il soggetto A perde la sua soggettività), nel caso della mimica (o mimazione) l'oggetto prodotto dipende fortemente dal soggetto A.

Quindi quando usiamo il termine "imitazione" ci riferiamo sempre a un atto in cui l'oggetto C che è imitazione di un altro oggetto B tende ad assomigliare in modo pressoché indistinguibile all'oggetto originale (i capi di moda piratati, le banconote false sono imitazioni). Il soggetto che opera l'imitazione in un certo senso non pone nulla di proprio. La bravura dei falsari/imitatori consiste proprio nell'annullare il proprio estro personale e la propria capacità creativa per produrre un oggetto indistinguibile dall'originale.

L'atto mimico al contrario è sempre un atto di soggettivazione, un atto cioè in cui il soggetto, le sue inclinazioni, le sue predilezioni o addirittura le sue abitudini entrano come ruolo fondamentale.

I comici e gli artisti che rivestono i panni e ricreano un personaggio noto assumono perciò un ruolo mimico e non imitativo, in quanto non annullano la loro personalità e riconoscibilità nell'assumere le sembianze di un altro, ma anzi proprio per questo mostrano un'arte simbiotica dove la loro personalità al servizio di un'illusione scatena simpatia e riconoscimento di bravura.

L'imitazione è una funzione che, a partire da oggetti qualsivoglia, porta ad oggetti determinati quasi indistinguibili o possibilmente indistinguibili, ma la mimazione invece no. L'oggetto che viene mimato non determina mai il risultato della mimazione che dipende sempre e comunque da colui che compie l'atto.

Un bambino non imita un aeroplano perché non ha alcuna accuratezza nel riprodurre le caratteristiche fisiche dell'aeroplano, si limita a distendere le braccia.

L'imitazione, talvolta, vuole essere ingannevole, ma la mimica non lo è mai. La mimica ricorda il personaggio, ne esalta alcune caratteristiche per renderlo riconoscibile, ma

mantiene la riconoscibilità dell'interprete ed è questo il lato artistico che nasce dall'imitatore professionista.

Siccome, a causa della variabilità intraindividuale, noi non possiamo mai riuscire a dire neanche la stessa cosa due volte nello stesso modo, è necessario per un imitatore fare una selezione dei principali fattori caratteristici ed esaltarli per essere identificato con il target, oppure addirittura confuso col target. L'imitatore fa tutto questo trascurando alcune caratteristiche per rinforzarne altre in base a quelle che ritiene più rilevanti ma mai nella direzione dell'esagerazione del tratto a differenza degli imitatori non professionisti che cercano invece al massimo di approssimare le caratteristiche del target. Se il professionista imitatore trascura alcuni tratti e ne rinforza altri è perché è conscio della necessità di mettere in luce questi aspetti, non nel tentativo di riprodurre la voce di qualcuno, ma di essere identificato con questo.

Il livello richiesto al compito mimetico e imitativo si basa su due tipi di strategie: una strategia di alto livello (le discorsive tematiche, la scelta del vocabolario, le scelte sintattiche, i manierismi...) e una strategia di basso livello (l'altezza tonale, l'intonazione, l'articolazione, la fluenza...) che richiede talento nell'acuità percettiva e nella flessibilità vocale molto meno comune nella popolazione. La frequenza fondamentale F0 appare essere un parametro facilmente modificabile. Per quanto riguarda la fluenza dello speech invece si è rilevato che il ritmo articolatorio rimane molto legato all'habitus fonatorio dell'imitatore.

La convergenza, ossia l'adattamento dell'imitatore alle caratteristiche globali del target (ad esempio cambiare l'altezza tonale della voce o la velocità dello speech), presenta comunque dei limiti per esempio quelli laringei del soggetto imitatore che potrebbero non permettere determinate estensioni.

Quando la convergenza raggiunge i suoi limiti allora ci si muove verso la sincronia che consiste nell'imitazione delle variazioni istantanee (curva intonativa, collocazione e durata delle pause riguardanti il ritmo).

Pertanto un'imitazione di successo è in realtà una combinazione di strategie di convergenza e di sincronia come allineamento istantaneo.

1.3 ESISTE UNA PEDAGOGIA DELL'IMITAZIONE VOCALE?

Dal vaglio della letteratura sembrerebbe non esistere una vera e propria pedagogia strutturata in merito, il che fa ipotizzare che, come non di rado accade in campo artistico, l'artista

coniughi mirabilmente tecnica di base ed espressione del proprio talento trasformando in arte un prodotto del quale egli stesso sovente non sa dare spiegazione in senso strettamente pedagogico proponendo strumenti empirici nel tentativo di elaborare una pedagogia dell'imitazione.

Recentemente un contributo pratico allo studio dell'imitazione è giunto da un doppiatore pesarese, Alberto Pagnotta che sul proprio canale YouTube nel 2018 (11) ha introdotto un format in cui mostra, sotto forma di tutorial, i procedimenti che segue per prepararsi a riprodurre una voce target. Tali procedimenti consistono in una fase iniziale di esposizione alla voce target, con più ascolti al fine di identificare gli aspetti definiti “costanti” e quelli definiti “variabili”; successivamente l'imitatore esegue un processo di immedesimazione nel personaggio da riprodurre, immaginando di possederne le caratteristiche fisiche. L'ultimo passaggio consiste nei tentativi di recitare varie battute precise del personaggio target, eventualmente inserite in una scala che va dalle battute soggettivamente più semplici da eseguire per l'imitatore in questione fino a quelle che risultano più forzate e complicate. Questo contributo dal web ci aiuta a comprendere, sebbene in termini ovviamente non scientifici, soprattutto un aspetto: il lato intrapersonale; si tratta di un aspetto saliente del processo di imitazione di un imitatore professionista. Tale contributo, pur nella sua semplicità, possiamo considerarlo un elemento, certamente ancora molto contenuto, a sostegno di un piccolo primo passo verso una pedagogia dell'imitazione vocale artistica.

Nel 2019, sempre su YouTube, Claudio Loretta (8), imitatore professionista attivo anche in radio e TV, ha realizzato un format simile, di nome Imitalent: una playlist di video in cui analizza la voce di vari personaggi televisivi noti sviscerandone tutti gli elementi che riesce a cogliere e di cui poi si serve per realizzare la sua imitazione (ad esempio cadenze regionali, postura ed espressioni ricorrenti).

2. IMPRONTA VOCALE E SUA IMITAZIONE

Di seguito si riporta il lavoro di P. Castellini et al. (2) relativo ad una ricognizione delle conoscenze rispetto all'impronta vocale e la sua imitazione attraverso l'individuazione di parametri strumentali che caratterizzano per l'appunto la mimesi vocale.

L'impronta vocale può essere definita come un insieme delle caratteristiche misurabili della voce umana che dovrebbe identificare inequivocabilmente un individuo; tali caratteristiche basate sulla configurazione fisica del sistema rino-faringeo-laringeo del parlatore possono

essere espresse come una formula matematica. Il termine impronta vocale viene utilizzato per fare riferimento ai campioni registrati per questo scopo, alla formula derivata ed alla sua rappresentazione grafica.

In origine il concetto di impronta vocale, in analogia con le più note impronte digitali anch'esse utilizzate a scopo identificativo individuale, venne coniato per indicare un particolare tipo di tracciato spettrografico, nello specifico quello a contorni di isointensità; ciò accadeva al termine della seconda guerra mondiale successivamente all'entrata sul mercato di un'apparecchiatura, il sonografo, apparecchio per l'appunto utilizzato inizialmente per attività belliche e fu in tale circostanza che nacque quello che oggi chiamiamo riconoscimento vocale. Lo spionaggio degli Stati Uniti aveva, infatti, appurato che ogni divisione germanica affidava le trasmissioni radio sempre allo stesso operatore; considerando che l'esatta provenienza della trasmissione poteva essere determinata con le tecniche di radiogoniometria, era chiaro che il riconoscimento dell'operatore consentisse di localizzare con assoluta esattezza la posizione della divisione di appartenenza e conoscere con notevole precisione lo spostamento delle truppe con tutto quello che ne può derivare in ambito bellico.

Nel periodo postbellico la maggior parte della popolarità dell'apparecchio continuò ad essere assicurata dalla sua caratteristica di strumento di riconoscimento della voce con finalità medico-legali. Nel corso degli anni, naturalmente, l'avvento e lo sviluppo della tecnologia digitale ha contribuito a modificare radicalmente la possibilità della metodica e lo stesso sonogramma ha assunto una posizione diversa nel panorama del riconoscimento del parlante. Il sonogramma, nell'ultimo decennio, nella scala dei metodi del riconoscimento del parlante risulta appartenere alla categoria dei metodi soggettivi e nell'ambito di quest'ultimi viene collocato più avanti rispetto al semplice ascolto ma al di sotto dei metodi oggettivi parametrici cosiddetti semiautomatici (6).

Ma dunque è proprio vero che l'impronta vocale possa essere considerata alla stregua dell'impronta digitale? L'imitazione di una voce può forzare il concetto di impronta vocale? Gli imitatori a quali fattori si riferiscono per rendere la loro imitazione più simile possibile al modello/impronta vocale?

Dal punto di vista della letteratura scientifica sono fondamentalmente due i filoni di studio che nel tempo si sono sviluppati: quello che riguarda il settore dell'imitazione vocale e quello del riconoscimento del parlante. Tali filoni possono costituire due entità di studio con temi di ricerca che si incrociano.

Nel campo dell'imitazione, si rileva che gli studi effettuati sono sostanzialmente di confronto

tra imitatore e traccia originale e/o tra imitatore e diverse tracce/parole “neutre” per valutare le caratteristiche qualitative e quantitative tipiche della voce imitata. In particolare sono stati studiati quali aspetti della voce imitata determinino maggior successo nei confronti del pubblico e su quali aspetti l'imitatore faccia leva per assomigliare maggiormente al suo modello. La recente revisione della letteratura internazionale, da noi effettuata, mostra che gli studi che esplorano l'imitazione vocale non sono numerosi e la maggior parte sono stati svolti dalla ricercatrice Elisabeth Zetterholm del Dipartimento di Linguistica e Fonetica dell'Università di Lund (Svezia). Nei suoi studi ritroviamo diverse modalità di esperimento coinvolgenti sia un'imitazione singola (18) sia imitazioni multiple (21) svolte da un imitatore professionista di cui vengono analizzati gli aspetti riguardanti la frequenza fondamentale, la durata e lo spettro sonoro in riferimento alla voce naturale dell'imitatore, alla voce dello stesso durante l'imitazione e alle varie voci target delle persone imitate.

In entrambi gli studi citati viene rilevato quanto la prosodia sia la qualità principale che rende l'imitazione calzante ed il fatto che l'imitatore effettui un'esagerazione delle caratteristiche del target, in accordo con quanto affermato da Laver nel 1994 (9) il quale riferisce che l'imitazione sia un processo stereotipato, non una copia esatta e confermato anche dagli studi successivi.

Nel secondo studio citato di E. Zetterholm (21) viene, inoltre, dimostrato che anche analizzando e confrontando una singola parola neutra l'imitazione risulta comunque riuscita; viene pure rilevata l'importanza della variazione dell'intonazione e dello speech style (tempo + ritmo + articolazione + continuità) per realizzare un'imitazione di qualità. In tale studio è stato inoltre dimostrato che le esagerazioni e le pronunce non perfette dei segmenti vocali, quando è stato valutato un intero discorso, non hanno inciso in modo determinante nell'impressione generale.

In uno studio successivo (17) vengono esaminate una voce imitata e quella dell'imitatore al fine di comprendere come quest'ultimo modifichi la propria voce ed il comportamento vocale per avvicinarsi alla voce bersaglio. Sono presentati i risultati preliminari delle analisi uditive e acustiche che si focalizzano sullo speech style e sugli schemi di intonazione. L'impressione uditiva è confermata con l'analisi acustica. I risultati di questo studio concordano con i risultati del test di percezione concernenti l'importanza della qualità della voce, altezza tonale e intonazione per una imitazione della voce di successo.

Un altro studio (16) ha dimostrato che anche due imitatori diversi impegnati nell'imitazione di voci diverse selezionano le medesime caratteristiche dei target per realizzare le loro imitazioni. È stato, inoltre, evidenziato che le abilità percettive di un imitatore professionista

nel riconoscere tali caratteristiche siano particolarmente sviluppate rispetto ad un gruppo di controllo (12).

Le imitazioni relative allo studio sopra ricordato di E. Zetterholm (16) risultano comunque tra loro diverse probabilmente per via delle differenze organiche che tuttavia non impediscono la realizzazione di imitazioni di successo, come confermato da un ulteriore studio della stessa ricercatrice (19) in cui un imitatore professionista riesce a cambiare la propria frequenza fondamentale, il proprio dialetto e la propria intonazione col risultato di modificare anche la durata che, come precedentemente sottolineato, è uno dei principali parametri di valutazione di un'imitazione riuscita. Il modello di durata è fortemente influenzato dalle abitudini fonetiche dell'oratore. Un quesito che resta irrisolto è quanto sia individuale il modello di durata.

Riguardo un ulteriore approccio strumentale inerente lo studio delle imitazioni vocali, si ricorda il lavoro di Carolyn McGettigan e colleghi (10) che ha introdotto l'utilizzo di fMRI nel controllo dell'esecuzione delle imitazioni vocali. In tale studio, sono stati valutati i correlati neurali del cambiamento vocale controllato in modo volontario in soggetti di lingua inglese, nessuno dei quali era doppiatore né imitatore professionista, i quali eseguono imitazioni simili ai tipi di imitazione vocale utilizzati nella conversazione quotidiana, ad esempio, nel riportare il discorso di altri durante una narrazione. L'obiettivo dello studio è stato individuare i sistemi neurali che supportano i cambiamenti nel modo in cui il discorso è articolato, senza variarne il contenuto linguistico. Lo studio, consistente in neuroimaging, evidenzia che la modificazione delle caratteristiche dell'espressione vocale, senza cambiare il contenuto linguistico del discorso, recluta principalmente l'insula anteriore sinistra e la corteccia frontale inferiore; inoltre viene rilevato che le regioni temporali superiori hanno mostrato una maggiore attivazione e una maggiore connettività positiva con i siti frontali di pianificazione del discorso durante l'emulazione di identità vocali specifiche.

2.1 IMITAZIONE E RICONOSCIMENTO VOCALE

Come specificato in un successivo studio di E. Zetterholm (20) è stato dimostrato che la qualità della voce dipende sia da strutture organiche immutabili sia da comportamenti personali appresi e mutevoli che influenzano la parlata e che costituiscono gli elementi principali da riconoscere e riprodurre da parte dell'imitatore.

L'imitatore, quindi, deve essere cosciente non solo delle caratteristiche “di gruppo” come il

dialetto sociale o regionale ma anche delle caratteristiche personali nella parlata come la pronuncia e l'articolazione. È noto, dunque, che fra le persone vi siano differenze organiche immutabili ma nonostante questo un imitatore abile ad adottare un ampio range di configurazioni articolatorie e fonetiche può raggiungere la voce target. Questo suggerisce che anche i sistemi di riconoscimento vocale potrebbero essere vulnerabili nei confronti delle imitazioni e ciò esemplifica quanto sopra prospettato in merito al possibile filo comune che collega i due temi: per l'appunto imitazione e riconoscimento vocale.

Quest'ultima considerazione apre un capitolo particolare, inerente un tipo di ricerca, che è andata via via sviluppandosi rappresentata dal tema, complesso quanto delicato (basti pensare all'ambito forense), dell'identificazione del parlatore tema per niente disgiunto dalle possibili ingerenze date dall'imitazione vocale.

Per chiarezza distinguiamo il riconoscimento vocale dal riconoscimento del parlatore: il riconoscimento vocale, riconoscimento delle parole, si basa su un sistema di conversione del segnale audio-vocale in testo leggibile dal computer o dai sistemi digitali in generale; esso spazia da applicazioni semplici come il riconoscimento di parole isolate relative ad un singolo parlatore ad applicativi più complessi quali il riconoscimento di qualsiasi parola indipendentemente dal parlatore; il riconoscimento vocale, dal punto di vista pratico, offre numerosi vantaggi per le applicazioni che da tempo, in diversi settori, sono entrate a far parte dell'uso comune quali per esempio la refertazione vocale, la dettatura digitale.

Altro è il riconoscimento del parlatore (Speaker Recognition) tema che può essere suddiviso in due categorie di applicazioni: verifica del parlatore (Speaker Verification) e identificazione del parlatore (Speaker Identification). Nella verifica del parlatore la persona si presenta tramite un suo identificativo di qualche tipo (smart card, nome utente, ecc.) e l'applicazione verifica se la voce di chi sta parlando corrisponda a quella della sua impronta vocale memorizzata nel sistema. Se risulta che la voce di chi parla e l'impronta sono simili, cioè se il valore di somiglianza supera una certa soglia, si ritiene che la verifica abbia successo e la persona è autenticata. In una applicazione di identificazione del parlatore invece, la persona che parla non si presenta e si deve misurare il grado di somiglianza della sua voce rispetto ad un insieme di impronte vocali disponibili.

Come ricordano alcuni autori (3) l'identificazione può avvenire in un insieme aperto (l'utente non è stato catalogato in precedenza) e quindi il risultato della verifica può essere che la voce non corrisponda a nessuna di quelle note al sistema, oppure l'identificazione può avvenire in un sistema chiuso (tutti i possibili utenti del sistema sono già stati catalogati) e quindi la voce della persona dovrà corrispondere ad una di quelle catalogate. Il riconoscimento di un

parlante incognito è dunque effettuato confrontando la voce ascoltata con un insieme di modelli che contengono le caratteristiche vocali dei parlanti di interesse. Questi modelli sono quelli comunemente denominati impronte vocali o voiceprint.

Gli autori ricordano che un sistema di riconoscimento automatico del parlante prevede pertanto tre funzionalità:

-la fase di addestramento (creazione delle impronte vocali) propedeutica a tutte le successive operazioni. A tale scopo il soggetto fornisce al sistema un certo numero di campioni della propria voce per consentirgli di apprendere le caratteristiche peculiari. Le impronte vocali potranno essere utilizzate per le successive fasi di riconoscimento;

-la verifica della voce del parlante: il soggetto dichiara, in modo implicito o esplicito, la sua identità; il sistema recupera l'impronta vocale associata all'identità dichiarata e, a seguito del confronto di quest'ultima con uno o più campioni della voce del soggetto, ne autentica o meno l'identità;

-l'identificazione della voce del parlante: il sistema recupera l'insieme delle impronte vocali dei soggetti che si intendono identificare; la voce del parlante sotto esame è confrontata con le impronte di soggetti noti e l'ipotesi di associazione più probabile è proposta come risultato. In tutto questo non va perso di vista il problema di due possibili errori: falsi negativi (mancata individuazione), falsi positivi (errata individuazione); a questo proposito gli autori affermano che i due tipi di errori sono tra loro correlati in funzione della scelta della soglia di decisione, pertanto, in base all'impiego applicativo è possibile scegliere quale tipo di errore minimizzare.

Da una parte dunque si nota il problema del riconoscimento del parlante, oggi più che mai argomento di grande attualità, per lo sviluppo di tecnologia che trova applicazione nei più disparati settori e che offre servizi basati su attivazioni a comando vocale e dall'altra lo stesso tema del riconoscimento del parlante, in continua evoluzione, nel settore forense.

In merito a quest'ultimo aspetto, come evidenziato in un report sui metodi di riconoscimento del parlante (15) è evidente quanto per le Forze dell'Ordine e della Magistratura il riconoscimento di una voce rappresenti massima importanza per la soluzione di particolari casi giudiziari, si pensi per esempio alle rivendicazioni di attentati, alle estorsioni, richieste di riscatto. Ciò implica un meccanismo molto complesso e delicato al cui interno l'esperto deve osservare determinati protocolli, conoscere dettagliatamente le tecniche analitiche e di indagine, l'impiego dei materiali e della letteratura internazionale in materia.

L'autore, rispetto alle relazioni tecniche fornite in merito dagli esperti, evidenzia che la posizione dominante in Europa è dettata dall'ENFSI (European Network of Forensic

Science) che si propone di presentare i dati in versione statistica-bayesiana in modo che risultino esplicite sia la probabilità di riconoscimento (cioè il grado di compatibilità tra la traccia anonima e la medesima caratteristica del sospettato) sia la probabilità di falsa identificazione (la probabilità che un'altra persona innocente possa produrre analoga traccia) ciò che viene definito con il nome di likelihood ratio. L'autore nel suo studio fa presente che le scienze forensi, in tutti i propri settori, si stanno orientando verso questo formato di risposta, malgrado il processo di trasformazione non sia semplice, stanti le difficoltà di ordine culturale (in quanto gli esperti anziani sono tendenzialmente restii all'idea di adeguarsi ad un moderno concetto di “statistica forense” che richiede, oltre allo studio, anche un cambiamento metodologico) e soprattutto pratico, perché non tutte le discipline hanno sviluppato efficaci strumenti matematici e, cosa più grave, approntato database in grado di rappresentare adeguatamente i fenomeni di studio adeguandoli ad una popolazione di riferimento significativa.

L'articolo in oggetto, a cui si rimanda, espone e confronta i metodi esistenti, utilizzati in Europa e nel mondo da quelli tradizionali ai più recenti.

Lo studio evidenzia che per la comparazione della voce ai fini forensi coesistono, di fatto, tecniche sia linguistiche sia acustico-strumentali. Le seconde hanno condotto all'elaborazione di programmi di analisi di tipo semi-automatico, che sono di ausilio all'operatore nella identificazione, o totalmente automatici, in cui l'operatore introduce esclusivamente il file audio e imposta il database di confronto; il sistema elabora l'informazione e fornisce direttamente il risultato calcolando un rapporto di verosimiglianza. L'autore ricorda che lo stato dell'arte sulle tecniche di riconoscimento del parlatore riflette la sensibilità internazionale sull'argomento; alcuni Paesi, quali Italia, Francia e Spagna, colpiti fin dagli anni Settanta dal terrorismo sono stati tra i pionieri della ricerca nel settore; altri quali gli stessi Stati Uniti hanno iniziato solo recentemente, investendo ingenti risorse per la mutata sensibilità, causata anche dai noti fatti dell'11 settembre 2001. Stante la multidisciplinarietà del settore, la ricerca in molti Paesi è incentrata sulle interazioni e sulle possibili integrazioni tra diverse tecniche per ottenere la maggior quantità di dati indipendenti utili alla risoluzione del problema comparazione. Benché tutte le tecniche abbiano dei limiti nell'applicabilità, l'autore ricorda tuttavia che questi sono diversi; alcune caratteristiche del segnale possono precludere un tipo di analisi ma non un altro e viceversa; inoltre, quando è possibile l'esecuzione di tutte le metodiche note, si possono ottenere risultati decisamente più precisi.

Nella stragrande maggioranza degli Istituti di Scienze Forensi si applica un approccio ibrido

(denominato combined), che tiene conto sia degli aspetti strumentali che di quelli linguistici. Il problema è solo quello di amalgamare con cognizione di causa i dati, valutando opportunamente quelli indipendenti da quelli che non lo sono e non possono, quindi, essere contemplati più volte. L'autore fa presente che chi utilizza sistemi semi-automatici o automatici (con database di riferimento) è in grado di fornire la probabilità di falsa identificazione o la likelihood ratio, esprimendo valori che però non tengono in considerazione le risultanze emerse dagli esami linguistici che sono rimaste di carattere qualitativo. Peraltro nella maggior parte dei laboratori europei, proprio per la diversa estrazione culturale, gli esami linguistici sono effettuati da persone differenti, con professionalità diverse rispetto a chi esegue le misure sulle frequenze di emissione vocalica. Inoltre, problema non indifferente, molte tecniche non sono esportabili fuori dai confini nazionali, in quanto richiedono una precisa e attenta conoscenza della lingua. Conclude che il confronto tra le caratteristiche e le prestazioni mostra chiaramente come non vi sia un metodo migliore in senso assoluto. Sono in corso attività di ricerca per integrare i sistemi in modo da ottenere risultati migliori in un unico responso; lo sforzo attuale rimane dunque quello di valutare le possibili integrazioni tra le tecniche, pesando scientificamente o meglio statisticamente i contributi offerti dall'impiego di ogni singolo parametro in maniera da poter affrontare, di volta in volta, ogni specifico caso reale ritagliando la soluzione più completa. Rispetto al riconoscimento biometrico che fa riferimento alle tecniche automatiche per riconoscimento dell'identità dell'individuo basate sull'uso di caratteristiche fisiologiche o comportamentali distintive F. Beritelli, in una sua presentazione (1) ben delinea gli indici biometrici derivanti appunto dalle caratteristiche biometriche sostanzialmente rappresentate da tre tipologie: comportamentali (uniche ma variabili quali la calligrafia o lo stile di battitura), fisiologiche (uniche, invarianti quali il DNA, l'impronta digitale, la retina, il volto, la mano), ibride (la voce, i suoni cardiaci). L'autore, fra l'altro mostra l'attendibilità di alcuni indici biometrici quali impronta vocale (naturale Hi-Fi e da segnale telefonico), impronta digitale, firma e retina rispetto alle possibili falsa esclusione e falsa attribuzione. Per quanto riguarda l'impronta vocale, rispetto alla falsa attribuzione viene evidenziato che l'impronta vocale mostra una percentuale dello 0,1/0,4% a seconda che si tratti rispettivamente di voce naturale Hi-Fi o di segnale telefonico a fronte di una percentuale di falsa attribuzione dello 0,001% per quanto riguarda l'impronta digitale.

Altri autori (6) affermano che tenuto conto della natura stessa del parlato, della sua intrinseca variabilità anche nelle emissioni di un medesimo parlante, elementi che come noto rendono complesso il riconoscimento del parlante, non è corretto paragonare le tracce vocali con le

impronte digitali che non variano nel tempo e sono diverse nelle diverse persone. Gli autori, tuttavia, ritengono che anche il riconoscimento del parlante possa fornire un utile contributo alla determinazione della identità.

Tornando al filo comune che lega i due temi, imitazione e riconoscimento vocale, non solo è stato dimostrato che l'imitazione vocale di alta qualità può indurre in errore l'ascoltatore (13) ma E. Zetterholm nel suo studio del 2006 (20) sopra citato afferma che le evidenze che lei riporta nel suo lavoro suggerisce che i servizi di sicurezza, protetti da sistemi di verifica vocale, è possibile che siano vulnerabili alle imitazioni della voce vera del cliente. L'autrice sottolinea che sarà quindi importante conoscere quanto i sistemi siano sensibili e che cosa possa essere fatto per migliorare la loro immunità nei confronti di questo tipo di frodi. L'autrice riferisce in merito ad uno studio (4) in cui è stata esaminata l'abilità di uno speaker spontaneo, o di imitatori professionisti, di allenare la propria voce nei confronti di quella di uno speaker scelto a caso. La frequenza di falsa accettazione da parte del sistema era significativamente più elevata quando l'imitatore si era allenato per l'imitazione rispetto a prima che l'allenamento avesse avuto luogo. L'autrice prosegue facendo presente che ciò ha portato alla conclusione che l'imitazione umana possa essere una minaccia per la verifica vocale del cliente e che ci si potrebbe attendere che anche la robustezza dei sistemi automatici possa essere minacciata da voci simili fra loro.

I risultati di questi lavori fanno riflettere non poco considerando che il riconoscimento del parlante (in particolare la verifica dell'identità del parlante) indipendentemente dalle applicazioni in campo forense, costituisce oggi un aspetto di grande attualità in cui uno degli obiettivi su cui si punta molto l'attenzione è quello volto alle implicazioni relative alle applicazioni di tipo commerciale. Oggi, infatti, in automatico con applicazioni integrate e in rete di telefonia e informatica possiamo utilizzare una serie di funzioni che notoriamente si basano su comando vocale; tali funzioni, genericamente, possono essere raggruppate in due categorie: una può non necessariamente prevedere l'autenticazione del parlante (voce non specifica) e l'altra invece necessita di autenticazione certa del parlante (voce specifica).

Alla prima categoria sono per esempio ascrivibili applicazioni tipiche della domotica (tutti i sistemi connessi della casa) quali alzare e abbassare le tapparelle, regolare la temperatura delle stanze, attivare elettrodomestici o far partire segnali di allarme che si attivano appunto proprio grazie a comando vocale.

Alla seconda categoria appartiene la possibilità di attivare con il comando vocale funzioni che necessitano di autenticazione certa del parlante quali aprire la porta di casa, accedere telefonicamente ad informazioni riservate, effettuare operazioni finanziarie, ripristinare un

PIN, gestire un account, gestire dispositivi di sblocco. In tali casi è evidente l'imprescindibile necessità che ad essere riconosciuto sia proprio quell'individuo e non un altro.

Se dunque l'identificazione risultasse errata è palese quali gravi conseguenze ne potrebbero derivare, da qui, pertanto, la necessità di rendere il meccanismo di identificazione più sicuro possibile.

Attualmente, quando con la propria voce si attivano meccanismi che consentono operazioni molto personali, ad esempio lo sblocco di un sistema protetto, non risultando le identificazioni vocali del tutto sicure, ci si basa sull'autenticazione a più fattori (per esempio biometria vocale congiunta ad un secondo e non raramente anche ad un terzo fattore di autenticazione) proprio per aumentare il grado di sicurezza.

Sono noti servizi che si avvalgono di un percorso che nell'ordine prevedono: confronto della voce del chiamante con le impronte vocali memorizzate durante la procedura di iscrizione del cliente, in presenza di corrispondenza si concede l'accesso del cliente al sistema a cui segue avvio del secondo fattore di autenticazione (per esempio richiesta di due o più elementi che riguardano il cliente quali un oggetto, una data di un anniversario).

M. Farrùs (5) riporta in una review in modo articolato le conoscenze attuali riguardo il camuffamento vocale, indicato come *vocal disguise*, che più frequentemente portano all'errore nel riconoscimento automatico del parlatore. Nella review viene citata la tabella di Rodman del 1998 in cui il camuffamento vocale viene classificato in quattro tipi: non-deliberato non-elettronico (includente voce roca, modifiche emozionali della voce, intossicazione alcolica, invecchiamento), non-deliberato elettronico (riferito alla degradazione o distorsione della registrazione per inconvenienti tecnici, ad esempio problematiche legate allo stato dei microfoni), deliberato non-elettronico (imitazioni e modifica volontaria dello stile di parlata), deliberato elettronico (volontaria modifica della traccia registrata). Questa indagine affronta la scarsa connessione tra le caratteristiche modificate durante un camuffamento vocale e le caratteristiche utilizzate nell'altoparlante automatico dei sistemi di riconoscimento, che influiscono sulla robustezza del sistema. Un esempio significativo è quello delle conversioni vocali: i sistemi di conversione vocale si basano solo sullo spettro della voce, senza tener conto della prosodia e di altre dimensioni linguistiche, mentre il mimetismo vocale fa un uso significativo dell'alterazione della prosodia. I cambiamenti dialettali e altri camuffamenti si basano principalmente su fonetica, prosodica e alterazioni lessicali. Pertanto, non ci si aspetterebbe di trovare a priori un sistema automatico spettrale in grado di riconoscere la voce di un parlante secondo l'accento o il dialetto, a meno che l'accento e le caratteristiche dialettali non si riflettano nello

spettro della voce, che non è usuale. Pertanto, una comprensione più ampia di quali caratteristiche siano alterate nel mimetismo e nel camuffamento vocale in generale sarà sicuramente d'aiuto per migliorare la progettazione dei sistemi di riconoscimento automatico. Inoltre l'aggiunta di caratteristiche alterate basate sul camuffamento in questi sistemi aumenterà le loro prestazioni, sia in termini di precisione che di robustezza.

Si ricorda, infine, lo sviluppo crescente di sistemi digitali per produrre, simulare e trasformare voci; fra essi riportiamo a titolo di esempio Lyrebird, algoritmo di una omonima startup canadese del quale non disponiamo di studi pubblicati in letteratura e che, secondo quanto recensito da T. Toniutti (14), è in grado di replicare, con grande accuratezza, timbri umani specifici, modificando intonazione e velocità, semplicemente analizzando un frammento audio della fonte sonora originaria.

L'algoritmo di Lyrebird utilizza l'intelligenza artificiale per riprodurre il timbro che gli viene sottoposto e riesce a farlo in tempi rapidi, circa un minuto per analizzare i dati e restituire un risultato. L'azienda inoltre sostiene che il software sia in grado di conferire al timbro clonato un "valore emozionale", modulando i discorsi sintetizzati con diverse intonazioni a seconda dell'intenzione che si vuole dare al contenuto.

A seguito di ciò, tale clone può parlare al posto nostro semplicemente scrivendo su una tastiera ciò che si desidera fargli dire. Nella recensione vengono evidenziati i potenziali rischi dell'utilizzo di tale tecnologia, dalla possibilità di realizzare documenti audio con dichiarazioni false alla sostituzione di intercettazioni reali con voci replicate, ipotizzando inoltre che il miglioramento di tale tecnologia porterà ad invalidare le registrazioni vocali come prove nei processi ed elementi probatori in genere. La recensione sostiene inoltre la probabilità che la voce sintetica un giorno possa completamente sostituire quella umana.

CONCLUSIONI

Come si evince dalla revisione della letteratura nazionale ed internazionale il numero di lavori nel settore della mimesi vocale è molto esiguo.

Gli studi mostrano che gli imitatori costruiscono il loro prodotto artistico plasmando elementi vocali quali altezza tonale, intonazione, articolazione e fluency coniugandoli con altri elementi del parlato (e non solo) fra i quali scelta del vocabolario, discorsive tematiche, scelte sintattiche, manierismi.

Un'imitazione di successo, pertanto, è il risultato della combinazione di strategie di

convergenza attuate dall'imitatore che adatta se stesso alle caratteristiche globali del target mettendo a servizio il proprio talento nell'acuità percettiva e la propria flessibilità vocale.

La letteratura mostra inoltre le connessioni fra imitazione-impronta vocale e riconoscimento vocale mostrandoci un aspetto con implicazioni che ci porta al di là dell'imitazione in senso stretto.

Lo stato dell'arte relativo all'impronta vocale e alla sua imitazione ci introduce infatti in una ulteriore dimensione con risvolti pratici tutt'altro che trascurabili, basti pensare alle implicazioni medico-legali dove per esempio il riconoscimento del parlatore rappresenta un prezioso rilevamento piuttosto che la tutela della sicurezza. La minaccia crescente di attacchi alla sicurezza, lascia facilmente immaginare che nel prossimo futuro sarà particolarmente utile approfondire se le caratteristiche qualitative della voce imitata possano per esempio ingannare la tecnologia della quale oggi facciamo ampio uso che dovrebbe riconoscere il parlatore e quali di tali caratteristiche nel caso abbiano questo potere in modo da poter provvedere a potenziare i livelli di sicurezza.

BIBLIOGRAFIA

1. Beritelli F., *Perizie foniche in ambito forense*. Dipartimento Ingegneria Elettrica Elettronica e Informatica DEEI Università di Catania 2013.
2. Castellini P., Bernardini A., Santomauro V., Poggi G., Travalca Cupillo B., *L'impronta vocale e la sua imitazione: stato dell'arte – Foniatria e Logopedia 2.0: come la tecnologia cambia le professioni foniatria e logopedica – Relazione ufficiale SIFEL*, Franco Angeli Editore, 2017.
3. Colibro D.E., Ferraris A., Piovano L., Scarpato L., Vair C., *La macchina riconosce la voce*. I quaderni di Teléma, 2007.
4. Elenius D. Harmning, *ett hot mot talarverifieringssystem?* (in Swedish). Master thesis, TMH, KTH, Stockholm, 2001.
5. Farrùs M., *Voice disguise in automatic speaker recognition*. ACM Computing Surveys 51 (4): 1-22, 2018.
6. Federico A., Paoloni A., *Riconoscimento del parlante*. I quaderni di Teléma, 2007.
7. Fussi F., Atti digitali del 10° Corso Internazionale Foniatria e Logopedia “La Voce Artistica”, Ravenna, 2015. <https://www.voceartistica.it/>
8. Laretta C., https://www.youtube.com/channel/UCwn3FlrkN0xPTewqPk_Ixl
9. Laver J., *Principles of phonetics*. Cambridge University Press, Cambridge.
10. McGettigan C., Eisner F., Agnew Z.K., Manly T., Wisbey D., Scott S.K., *T'ain't what you say, it's the way that you say it--left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations*. J Cogn Neurosci. 2013 Nov;25(11):1875-86. doi: 10.1162/jocn_a_00427. Epub2013 May 22.
11. Pagnotta A., <https://www.youtube.com/user/AlbertPagnos88>
12. Revis J., De Looze C., Giovanni A., *Vocal flexibility and prosodic strategies in a professional*

impersonators. Journal of Voice, 27(4), 2013, pp. 524.e23-524.e31 2013 The Voice Foundation.

13. Schlichting F., Sullivan K.P.H., *The imitated voice – a problem for voice line-ups?*. Forensic Linguistics, 4(1), 1997, pp. 148-165.
14. Toniutti T., <https://www.repubblica.it/tecnologia/sicurezza/2017/04/26/news/lyrebird-163934385/>
15. Zavattaro D., *Report sui metodi di riconoscimento del parlatore*. Rassegna dell'Arma dei Carabinieri n. 3 / 4 luglio-dicembre 2006.
16. Zetterholm E., *A comparative survey of phonetic features of two impersonators* *Proceeding of Fonetik*. TMH-QPSR, 44(1), 2002A, pp. 129-132.
17. Zetterholm E., *Impersonation – reproduction of speech*. Working Papers, Lund University, Dept. Of Linguistics, Vol. 49. 2001.
18. Zetterholm E., *Impersonation: a phonetic case study of the imitation of a voice*. Working Papers, Lund University, Dept. Of Linguistics, Vol. 46, 1997.
19. Zetterholm E., *Intonation pattern and duration differences in imitated speech*. Laboratoire Parole et Langage, CNRS / Université de Provence, Aix-en-Provence (France) 2002B.
20. Zetterholm E., *Same speaker – different voices A study of one impersonator and some of his different imitations*. *Proceeding of the 11th Australian International Conference of Speech Science & Technology*, ed. Paul Warren & Catherine I. Watson. University of Auckland, New Zealand. December 6-8, 2006.
21. Zetterholm E., *The significance of phonetics in voice imitation*. In *Proceedings of the Eight Australian International Conference on Speech Science and Technology* pages 342-347 publisher Australian Speech Science and Technology Association 2000.