DOI: 10.17469/O2101AISV000031

ANDREA PAOLONI, MASSIMILIANO TODISCO

Calcolo del rapporto di verisimiglianza tra parlanti in prove soggettive di ascolto

Voice identification using measurements of physical parameters, generally made through instrumental techniques, combines a significant performance, especially in the case of automatic methods, with an easy control of the data obtained. However, the possibility of working in this field with subjective investigation criteria on the basis of human listening should not be underestimated. In the present work we illustrate a method for obtaining a numerical value based on the likelihood, for subjective tests in speaker recognition tasks. This method was finally validated through a number of experiments made by two different types of listener: non-experts and technical experts.

1. *Introduzione*

È esperienza comune riconoscere una persona dalla voce quando la ascoltiamo parlare senza vederla, perché in un ambiente diverso o perché la ascoltiamo attraverso il telefono. Peraltro, il riconoscimento vocale dell'identità di una persona basato sul mero ascolto è stato l'unico disponibile per molti millenni ed è stato adottato, nelle sue diverse connotazione in numerosi casi giudiziari, anche celebri[1].

In ambito forense però questa metodica di riconoscimento del parlante è stata sostanzialmente abbandonata in favore dell'identificazione della voce tramite parametri acustici, in quanto queste misure erano in grado di supportare il parere dell'esperto con dati numerici verificabili (Paoloni, 1997; 2014). Inoltre queste metodiche uniscono al vantaggio della parziale o totale automazione dei processi operativi una notevole velocità di esecuzione delle misure, soprattutto nel caso dei metodi automatici.

Per quanto detto, sembrerebbe logico l'abbandono dell'indagine uditiva a favore delle tecniche automatiche. Vogliamo però richiamare l'attenzione sul fatto che, allo attuale della conoscenza, anche le più sofisticate metodologie automatiche (Campbell, 2014) (Drygajlo, 2012) non utilizzano, al fine del riconoscimento del parlante, tutte le caratteristiche della sua voce, caratteristiche che nella loro interezza, sono valutate dall'orecchio e dal sistema di percezione di un ascoltatore e a volte costituiscono un elemento decisivo di identificazione. Le considerazioni sopra riportate suggeriscono quindi l'opportunità di utilizzare entrambe le metodologie. Nel presente lavoro illustriamo una via per ottenere, anche dalle prove soggettive, un dato numerico verificabile nella forma oggi ritenuta migliore, ossia nel rapporto di verisimiglianza. Il rapporto di veri-

[1] La voce del sequestratore del figlio del famoso trasvolatore Lindbergh fu da lui riconosciuta in Tribunale come quella che aveva chiamato Condon al cimitero la notte della consegna del riscatto.

simiglianza (LR – Likelihood Ratio) è il rapporto tra la similarità dei campioni a confronto, ovvero quanto sia simile il campione del noto rispetto a quello dell'anonimo, e la tipicità, ovvero quanto il campione all'esame sia comune nella popolazione di riferimento. Supponendo che i valori di similarità siano dello stesso ordine di grandezza, il rapporto dipende da quanto il campione sia comune all'interno della popolazione.

Il presente lavoro è organizzato nel seguente modo: nel paragrafo 2 verranno illustrate i criteri per le prove soggettive di ascolto; nel paragrafo 3 la metrica utilizzata e la definizione dell'esperimento; i risultati ottenuti sono riportati nel paragrafo 4 e infine il paragrafo 5 presenterà le conclusioni.

2. *Le prove soggettive di ascolto*

Nell'identificazione del parlante attraverso l'ascolto si possono definire due diverse situazioni. La prima è quando la voce da identificare è nota all'ascoltatore (*familiar voices*) e sfrutta la capacità dell'uomo di memorizzare archetipi della caratteristiche peculiari di un suono, sia esso una voce, sia esso una musica. Questa capacità, che è quella che ci consente di riconoscere la voce dei nostri congiunti e dei nostri amici, è quella che secondo la letteratura fornisce le migliori prestazioni, con un tasso di riconoscimento superiore al 90%.

La seconda è quella dagli esperti negli esperimenti di "ricognizione" e si basa sul confronto tra due o più voci non note all'ascoltatore (*unfamiliar voices*) rese contemporaneamente disponibili da un sistema di registrazione audio. In questo caso l'ascoltatore confronta coppie di voci in successione (o anche contemporaneamente), cercando elementi di diversità o di somiglianza. Questa metodica ha prestazioni lievemente inferiori alla precedente, con un tasso di riconoscimento intorno all'85% (Nolan, 1983). Nel caso che la situazione sia quella di un esperto estraneo ai fatti (e non può perciò trattarsi di voce familiare) le metodologie più frequentemente impiegate per il confronto uditivo delle voci possono essere ricondotte ai seguenti due criteri.

Un primo criterio basato su un reiterato ascolto, da parte dell'esperto, dei campioni di voce in esame al fine di individuare eventuali elementi di natura linguistica, fonatoria o acustica comuni alle due voci. L'esperto, sulla base degli elementi recepiti, esprimerà un giudizio sulla attribuzione o meno ad uno stesso parlatore delle voci ascoltate. Un secondo criterio è quello basato sul confronto delle voci effettuato da una squadra di ascoltatori, anche non esperti. Il materiale fonico in questo caso è costituito da un insieme di voci comprendenti la voce da identificare, le voci anonime ed eventualmente alcune voci estranee (popolazione di riferimento), prelevate da parlatori aventi caratteristiche fonatorie simili a quelle delle voci in esame. Questi campioni vengono poi ascoltati a coppie che saranno, se presentate tutte agli ascoltatori, $M(M\text{-}1)/2$, dove M è il numero di campioni al confronto. Ciascun operatore dopo l'ascolto di ogni coppia dovrà esprimere un giudizio di similarità o meno delle voci a confronto. L'elaborazione statistica dei giudizi espressi dagli ascoltatori consente di giungere a conclusioni di tipo sostanzialmente qualitativo (Anil, 2005). È tuttavia possibile, utilizzando questo secondo metodo, stimare un rapporto di verisimiglianza (LR) che sarà dato dal rapporto tra il

valore medio della somiglianza attribuita alla coppia o alle coppie voce anonima-voce del sospettato (voce nota) e il valore medio della somiglianza attribuita alle coppie voce del sospettato-voci della popolazione di riferimento.

3. *Definizione dell'esperimento*

La nostra preferenza, tra i metodi di ascolto citati, va all'ultimo dei criteri illustrati che si rifà ad una collaudata metodica utilizzata nelle valutazioni di qualità in campo telefonometrico. Questo metodo consente la stima della verisimiglianza relativa alla particolare prova eseguita e pertanto consente una valutazione dell'attendibilità del risultato rispettando i criteri richiesti in ambito giudiziario. Per verificare quali valori di LR siano ottenibili con questo tipo di prove, è stato proposto un test disponibile in rete che mette a confronto 14 campioni di popolazione (p_1,p_2,...,p_{14}) + 2 campioni anonimi (a_1,a_2) con 2 campioni della voce nota (n_1,n_2) per un totale di 32 confronti divisi in due sessioni di test (16 + 16). L'ascoltatore potrà ascoltare quante volte desidera ciascuno dei 32 confronti ed esprimerà, per ciascuno di esso, un giudizio di somiglianza a 7 livelli: somiglianza 0 (nessuna), somiglianza 1 (minima), somiglianza 2 (limitata), somiglianza 3 (qualche somiglianza), somiglianza 4 (abbastanza simile) somiglianza 5 (molto simile), somiglianza 6 (identità). La *LR* (rapporto di verisimiglianza) è calcolata secondo la formula

$$LR = \frac{1}{K}\sum_{k=1}^{K} lr_k$$

dove *K* è il numero di ascoltatori e *lr* è dato da

$$lr = \frac{\dfrac{1}{N_n N_a}\displaystyle\sum_{i=1}^{N_n}\sum_{j=1}^{N_a} S(n_i, a_j)}{\dfrac{1}{N_n N_p}\displaystyle\sum_{i=1}^{N_n}\sum_{j=1}^{N_p} S(n_i, p_j)}$$

dove *S* è il valore di somiglianza tra le coppie di voci, N_n, N_a e N_p sono, rispettivamente, il numero totale delle voci note, delle voci anonime e delle voci della popolazione di riferimento. In particolare, nel nostro esperimento $K = 26$, $N_n = 2$, $N_a = 2$ e $N_p = 14$.

Si è ritenuto altresì utile verificare se le prestazioni di ascoltatori esperti fossero migliori di quelle di soggetti che non hanno particolare addestramento nel campo riconoscimento delle voci e pertanto il test è stato sottoposto a 9 esperti e 17 non esperti.

4. *Risultati*

I risultati delle misure soggettive d'identificazione del parlante sono riportati nella Tabella 1 e 2.

Nella Tabella 1 sono riportate le LR di 9 esperti, mentre nella Tabella 2 quelle di 17 non esperti.

Nella Tabella 3 sono invece riportare le medie e le varianze relative ai due esperimenti, a1 e a2, degli esperti e non esperti.

Tabella 1 - *Likelihood Ratio dei 9 Esperti*

		ESPERTI																
		n1	n2	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	LR
E1	a1	2	0	1	0	0	0	0	0	1	2	0	1	0	0	0	0	5.60
	a2	0	2	0	0	0	1	0	1	1	0	1	0	1	1	0	2	3.50
E2	a1	6	0	1	0	1	0	1	0	0	1	0	1	0	0	0	1	14.00
	a2	0	6	1	0	0	0	0	0	0	1	0	0	0	0	0	1	28.00
E3	a1	3	0	2	0	0	0	0	0	0	4	0	3	0	0	0	0	4.67
	a2	0	5	3	0	0	3	1	2	4	4	0	2	1	3	0	3	2.69
E4	a1	4	4	4	0	1	3	0	1	2	2	0	1	0	0	0	4	3.11
	a2	0	4	2	0	5	5	1	1	3	2	2	1	1	5	0	2	1.87
E5	a1	4	5	5	5	0	2	1	1	0	0	0	2	1	0	0	0	3.29
	a2	4	6	3	6	6	5	1	5	1	2	5	5	4	2	1	5	1.65
E6	a1	3	0	1	1	0	0	1	0	0	3	0	1	0	0	0	0	6.00
	a2	0	4	0	1	0	0	0	0	1	1	1	0	0	1	0	0	11.20
E7	a1	4	0	4	1	0	0	0	0	0	0	0	1	0	1	0	1	7.00
	a2	0	5	5	1	0	5	5	4	4	0	1	2	1	4	0	0	2.19
E8	a1	4	1	1	0	0	0	0	1	0	2	0	0	0	0	0	0	14.00
	a2	0	4	0	0	5	1	0	-1	1	1	0	0	1	5	0	0	4.31
E9	a1	5	0	3	2	0	0	5	0	0	0	0	4	1	0	0	0	4.67
	a2	0	5	0	1	5	4	0	4	5	0	1	1	2	3	0	1	2.59

Tabella 2 - *Likelihood Ratio dei 17 non Esperti*

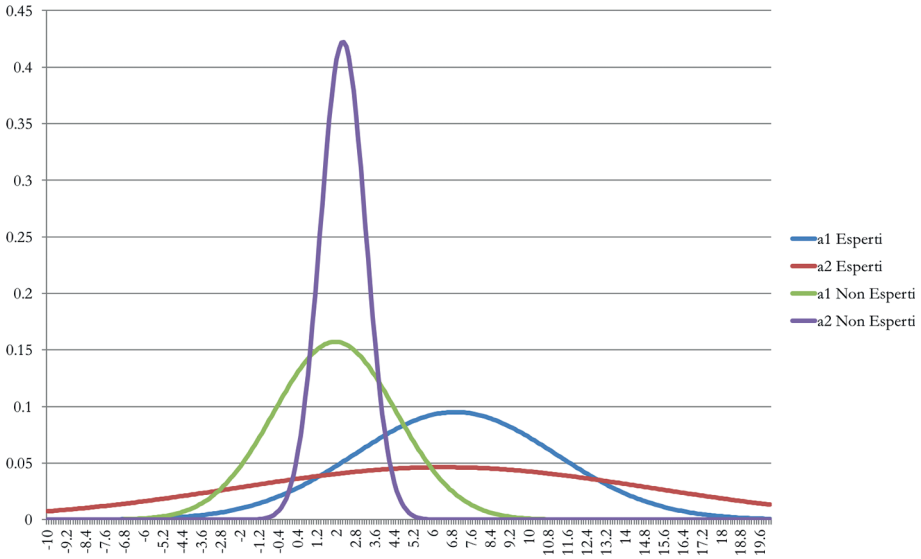
		NON ESPERTI																
		n1	n2	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	LR
I1	a1	6	6	5	0	3	4	3	3	4	4	0	1	0	2	0	0	2.90
	a2	2	6	2	0	1	6	0	-1	6	0	0	3	1	5	0	1	3.50
I2	a1	1	0	3	1	4	2	0	4	4	5	3	0	5	4	1	5	0.34
	a2	4	4	4	4	5	1	2	3	1	1	0	0	4	0	1	1	2.07
I3	a1	5	0	5	0	0	0	0	0	0	2	0	0	0	0	0	0	10.00
	a2	0	5	2	0	2	1	1	2	0	4	0	0	0	5	0	0	4.12
I4	a1	1	0	3	2	1	0	1	0	1	3	2	1	1	1	1	1	0.78
	a2	1	4	4	1	3	3	2	2	1	4	1	1	2	2	0	0	2.15
I5	a1	2	4	5	5	0	2	4	2	5	5	1	2	1	0	2	3	0.76
	a2	4	3	5	2	2	4	0	3	5	3	2	5	4	6	4	5	0.84
I6	a1	6	4	6	0	6	4	0	4	2	5	0	5	4	6	0	6	1.75
	a2	0	5	2	0	6	0	4	4	6	4	2	5	3	5	0	5	1.52
I7	a1	1	2	1	2	1	2	6	1	2	4	4	3	2	2	2	3	1.20
	a2	3	6	2	3	4	5	1	1	4	2	2	3	0	2	3	3	2.40
I8	a1	0	4	4	0	0	0	0	0	2	0	1	4	0	0	0	4	0.00
	a2	0	2	4	2	3	3	0	4	3	1	3	4	3	0	0	1	0.90
I9	a1	2	3	1	2	0	0	0	2	1	1	3	3	1	0	0	3	1.65
	a2	2	3	3	2	3	3	1	2	0	1	3	1	0	1	0	2	1.91
I10	a1	4	3	0	2	1	0	5	1	1	3	1	1	0	0	0	0	3.73
	a2	0	6	5	1	5	5	4	1	4	3	0	2	1	1	2	0	2.47
I11	a1	2	2	4	1	3	5	1	2	1	2	1	4	4	2	1	4	0.80
	a2	2	5	4	2	5	4	6	5	5	2	3	4	2	4	1	4	1.37
I12	a1	6	0	6	0	0	0	4	0	1	6	0	1	0	0	1	1	4.20
	a2	0	5	1	0	3	6	0	1	5	2	5	5	0	6	0	5	1.79
I13	a1	1	0	1	0	0	0	0	0	0	3	0	0	0	0	0	2	2.33
	a2	0	3	1	0	2	3	1	0	2	0	0	1	0	2	0	1	3.23
I14	a1	4	0	6	6	1	0	1	3	1	3	1	3	0	2	0	6	1.70
	a2	0	6	4	1	5	6	3	4	4	0	3	2	5	4	2	3	1.83
I15	a1	0	0	0	0	5	1	1	0	1	5	0	5	1	1	5	4	0.00
	a2	2	6	1	0	4	5	5	1	6	0	1	1	0	5	3	4	2.33
I16	a1	2	4	2	2	1	2	4	1	2	4	4	3	2	2	2	3	0.82
	a2	2	5	2	1	3	3	1	1	4	2	2	3	0	2	3	3	2.33
I17	a1	0	0	4	0	0	0	0	0	0	0	0	4	0	0	0	0	0.00
	a2	0	6	6	0	1	0	0	0	0	5	0	5	0	6	0	0	3.65

Tabella 3 - *Medie e le varianze relative ai due esperimenti, a1 e a2, degli esperti e non esperti*

	MEAN ESPERTI	STD ESPERTI	MEAN NON ESPERTI	STD NON ESPERTI
a1	6.93	4.19	1.94	2.54
a2	6.44	8.60	2.26	0.94

In Figura 1 sono stati altresì riportati i valori dei test d'identificazione del parlante a1 e a2 eseguiti da esperti e non esperti

Figura 1 - *Risultati dei test d'identificazione del parlante a1 e a2 eseguiti da esperti e non esperti*



5. *Conclusioni*

Obiettivo del presente lavoro era proporre un metodo che consenta di oggettivizzare i risultati delle prove soggettive d'ascolto utilizzando il rapporto di verisimiglianza così come avviene per i sistemi automatici di riconoscimento del parlante. L'esperimento ha permesso anche di verificare che le prestazioni dell'esperto sono senz'altro migliori di quelle di soggetti che non abbiano particolare addestramento nel campo riconoscimento delle voci. Per quanto attiene ai valori di LR, che confermano l'identità delle voci di interesse, si osserva una compressione della scala, nel senso che è lecito ritenersi che anche in casi evidenti, ossia di grande somiglianza delle voci a confronto e di significativa differenza di quelle della popolazione di riferimento, difficilmente si riuscirà a ottenere valori di verisimiglianza elevati in quanto un soggetto tende a non fornire dati che escludano la somiglianza delle voci "diverse", ponendo tale valore a zero, e dare il valore massimo alle voci provenienti alla stessa sorgente. Il fatto quindi che il rapporto di verisimiglianza abbia valori

così compressi potrebbe far ritenere opportuno considerare l'impiego di una scala logaritmica come in altre fenomenologie percettive.

Bibliografia

Anil, A., Dessimoz, D., Botti, F. & Drygajlo, A. (2005). "Aural and Automatic Forensic Speaker Recognition in Mismatched Conditions". In *The International Journal of Speech, Language and the Law*, vol. 12, 214-234.

Campbell, J.P. (2014). Speaker Recognition for Forensic Applications. Keynote Address at Odyssey 2014: *The Speaker and Language Recognition Workshop*. Joensuu, Finland.

Drygajlo, A. (2007). "Forensic Automatic Speaker Recognition". In *IEEE Signal Processing Magazine*, 24 (2), 132-135.

Drygajlo, A. (2012). Automatic Speaker Recognition for Forensic Case Assessment and Interpretation. In Neustein, A., Patil, H.A. (Eds.), *Forensic Speaker Recognition, Law Enforcement and Counter-Terrorism*. Berlin: Springer, 21-39.

Falcone, M., De Sario, N. (1994). "A PC speaker Identification System for Forensic Use: IDEM". In *ESCA Workshop on Automatic Speaker Recognition, Identification, and Verificatio*, Martigny, Switzerland.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Nolan, F. (1997). Speaker recognition and forensic phonetics. In Hardcastle, W., Laver, J. (Eds.), *A Handbook of Phonetic Science*. Oxford: Blackwell, 744-767.

Koenig, B (1993). "Selected Topics in Forensic Voice Identification". In *Crime Laboratory Digest*, vol. 20, n. 4, 78-81.

Paoloni, A. (1997). Il riconoscimento del parlatore. In *Detective&Crime Magazine / Criminalistica - Le indagini fonetiche*.

Paoloni, A. (2014). Sul riconoscimento del parlante in ambito forense. In *Sicurezza e Giustizia*, Anno IV, n. 3.

Paoloni, A. (2003). Note sul riconoscimento del parlante nelle applicazioni forensi con particolare riferimento al metodo parametrico IDEM. In *Riv. Italiana di Acustica*, vol. 27 n. 3-4.

Rose, P. (2002). *Forensic Speaker Identification*. Taylor & Francis.

Wolf, J.J. (1972). "Efficient acoustic parameters for speaker recognition". In *J.A.S.A.*, vol. 51, n. 6, 2044-2056.