



Eugenia SAN SEGUNDO FERNÁNDEZ (2023). *La fonética forense. Nuevos retos y nuevas líneas de investigación*. Octaedro. 166 pp. ISBN: 978-84-19506-48-1.

La fonética forense es una disciplina emergente dentro de la lingüística aplicada con un impacto significativo en el ámbito legal. *La fonética forense: nuevos retos y nuevas líneas de investigación* (2023), publicado por Eugenia San Segundo Fernández, científica titular del CSIC en el área de fonética experimental y aplicada, ofrece una panorámica integral sobre este campo a través de su evolución, sus aplicaciones y los principales desafíos actuales.

El libro está estructurado en siete capítulos que abordan aspectos fundamentales de la disciplina lingüística. El primer capítulo, «Introducción: ¿qué es la fonética forense?», presenta una definición clara de la fonética forense y sus principales aplicaciones dentro del contexto legal. Se destaca su papel en la identificación de hablantes y en la evaluación de autenticidad de grabaciones. Asimismo, se presenta una visión general sobre la fonética como disciplina, explicando sus tres principales áreas de estudio: la fonética articulatoria, que estudia cómo se producen los sonidos del habla a nivel fisiológico; la fonética acústica, encargada de analizar la señal sonora resultante de la producción del habla; y la fonética perceptiva, enfocada en el estudio de la recepción de los sonidos por parte de los hablantes. Además, la autora hace énfasis en el carácter interdisciplinar de la fonética y cómo se integra con otras ciencias experimentales y computacionales.

A lo largo del capítulo se subraya el crecimiento de la fonética forense como un campo con gran aplicabilidad en el ámbito judicial, donde se emplea en casos de autenticación de grabaciones, comparación de voces y transcripción de audios forenses. Asimismo, se aborda la discusión terminológica sobre el uso del término *forense*, del latín *fórum*, en lugar de *judicial* o *legal*. Aunque

algunos investigadores han utilizado la expresión *fonética judicial* (especialmente en España y en el mundo francófono), *fonética forense* es el término más aceptado, en consonancia con el de *lingüística forense*. Además, esta disciplina cuenta con respaldo institucional gracias a la creación en 1991 de la International Association for Forensic Phonetics and Acoustics (IAFPA). San Segundo explica que, aunque la fonética forense es una subdisciplina de la lingüística forense, ha adquirido un estatus propio debido a la especificidad de sus métodos y técnicas, enfocándose en labores como la identificación de hablantes, el análisis de autenticidad de grabaciones, la comparación de textos para atribuir autoría, el análisis del discurso jurídico, etc.

La segunda sección del libro, «Principales campos de actuación de la fonética forense», es fundamental para comprender el papel de la disciplina en la resolución de crímenes y en la administración de justicia, así como su creciente importancia en el contexto de la inteligencia artificial y la ciberseguridad. En este sentido, la fonética forense se divide en cinco grandes áreas de aplicación. El primero de ellos es la elaboración del perfil fonético de un hablante desconocido. Similar a un perfil lingüístico en el análisis de textos, un fonetista forense puede construir el llamado *pasaporte vocal* a partir de una grabación. Los principales rasgos que se pueden determinar incluyen el sexo del hablante, la edad aproximada, la modalidad de habla y el sociolecto. De igual modo, se puede describir a grandes rasgos el estado de salud del informante a partir del diagnóstico de posibles patologías vocales o trastornos del habla, así como la presencia de acento nativo o extranjero. Este tipo de perfil es útil en las investigaciones previas, ya que permite acotar la lista de sospechosos en casos como secuestros o amenazas. Sin embargo, hay limitaciones en la información que puede obtenerse, como el peso exacto o los rasgos psicológicos del hablante. En este sentido, la autora ejemplifica la importancia de esta labor con el secuestro de Anabel Segura, ya que a partir de una grabación de voz emitida en el programa televisivo *Quién sabe dónde* (1995) se produjeron numerosas llamadas, y una de ellas identificaba la voz del secuestrador.



La segunda área de actuación de la fonética forense se centra en la determinación de la autenticidad de una grabación, cuyo objetivo primigenio es «determinar si una grabación ha sido editada o manipulada de algún modo» (p. 24). En este tipo de procedimientos se pone en valor el carácter interdisciplinar, ya que los peritajes de esta índole requieren de la labor de diversos profesionales. El método más empleado en este tipo de trabajos es el *análisis de la frecuencia de la red eléctrica* (Grigoras, 2005; Cooper, 2009), que «se trata de una técnica forense que consiste en comparar, por un lado, los cambios de frecuencia en la red eléctrica que existen de fondo en cualquier grabación con, por otro lado, los registros históricos de alta precisión de cambios de frecuencia de la red eléctrica que se tienen en una base de datos» (p. 25).

Dentro de los mecanismos para determinar la autenticidad de una grabación, Cicres (2011) señala tres perspectivas: la física, vinculada con el ya mencionado análisis frecuencial de la red eléctrica; la auditiva, fundamentada en la escucha por parte del lingüista; y la acústica, centrada en el estudio de las representaciones gráficas del sonido (oscilogramas y espectrogramas).

En tercer término, nos encontramos con la determinación del contenido de una grabación. Para ello, se lleva a cabo la transcripción forense de grabaciones de audio. Podemos encontrar transcripciones completas en las que se pretende reflejar fielmente lo que produce un hablante en la grabación; o bien se realiza una transcripción parcial para determinar el contenido de una sección específica del archivo de audio. Generalmente, se opta por realizar este procedimiento cuando hay disputas sobre una palabra u oración en concreto. San Segundo hace hincapié en que la transcripción siempre es ortográfica y no fonética, ya que los alfabetos fonéticos «no los conocen ni jueces ni abogados ni, en definitiva, ninguno de los interesados en solicitar este tipo de peritaje» (p. 29).

El cuarto campo de aplicación es el diseño y validación de ruedas de reconocimiento. Como señala la autora, este procedimiento es el equivalente auditivo de una rueda de reconocimiento visual, pero en lugar de identificar rostros, se identifican voces. El proceso se basa en presentar

varias muestras de habla a las víctimas y/o a los testigos. Estas pueden ser de tipo objetivo presente o ausente, lo que implica que en las primeras el sospechoso se incluye en la rueda de reconocimiento, mientras que en las segundas no. No obstante, esta metodología también se vincula con estudios perceptivos; por ejemplo, se mencionan investigaciones relacionadas con las «habilidades de reconocimiento de voces por parte de invidentes» (p. 38).

El quinto y último ámbito se enfoca en la comparación forense de hablantes. Esta es la tarea más frecuente dentro de la fonética forense y consiste en determinar si dos grabaciones pertenecen a la misma persona. Se trabaja con dos tipos de grabaciones: la grabación dubitada, donde la voz es de un hablante desconocido, generalmente relacionada con un delito; y la grabación indubitada, donde la voz es de un sospechoso de identidad conocida. Se utilizan métodos de análisis fonético-acústico, incluyendo comparación de formantes vocales, ritmo, entonación y características segmentales.

Por su parte, el tercer capítulo, «La comparación forense de hablantes: contextualización histórica», se presenta como una suerte de estado de la cuestión. En él se analiza el desarrollo histórico de la comparación de voces como prueba forense y las metodologías utilizadas en la actualidad para mejorar la precisión y fiabilidad de este procedimiento. En concreto, aborda la comparación forense de hablantes (en adelante, CFH), uno de los principales campos de la fonética forense y la tarea más comúnmente solicitada a los expertos en este ámbito.

La CFH consiste en analizar y cotejar grabaciones de voz con el objetivo de determinar si pertenecen a la misma persona. Como ya señalamos anteriormente, se distingue entre dos tipos de grabaciones: dubitada, asociada a un delito y relativa a un hablante desconocido; e indubitada, perteneciente a un sospechoso cuya identidad es conocida. El propósito de la CFH no es identificar con certeza al hablante, sino «averiguar si la muestra dubitada y las indubitadas pertenecen al mismo hablante» (p. 43). Entre los delitos más comunes en los que se aplica este tipo de cotejos se encuentran las amenazas de bombas, fraudes bancarios, secuestros, tráfico de drogas y acoso.



Una de las principales controversias en este campo es el uso del término *identificación*. Tradicionalmente, se hablaba de *identificación forense de locutor*, pero en la actualidad se prefiere *comparación forense de hablantes* para evitar la impresión de que se puede determinar con absoluta certeza la identidad de una persona solo con su voz. Este cambio de terminología se debe en parte a los estudios de Saks y Koehler (2005), quienes «sostienen que todas las ciencias forenses deberían emular el enfoque de la comparación de perfiles de ADN, donde un cambio de paradigma ya habría ocurrido a partir de los años noventa» (p. 44). En lo que respecta a los aspectos metodológicos, se define el nuevo paradigma en CFH con base científica en la estadística bayesiana, que enfatiza el uso de bases de datos de referencia y la probabilidad condicional para evaluar la fuerza de la evidencia en lugar de emitir juicios categóricos sobre la identidad de un hablante.

En otro orden de cosas, también se trata el concepto de *voiceprint* o huella vocal, una técnica basada en el análisis de espectrogramas de voz, utilizada en sus inicios para identificar hablantes. Actualmente, la mayoría de los fonetistas forenses rechazan el uso exclusivo de espectrogramas para identificación, prefiriendo métodos que combinan análisis acústico y estadístico. Este tercer bloque finaliza con *El cisma del nuevo paradigma*, haciendo referencia a la división generada en la comunidad científica a principios del siglo XXI en relación con el enfoque metodológico de la CFH. El conflicto se centró en dos vectores. Por un lado, el enfoque tradicional, que se basaba en análisis auditivo-acústicos subjetivos y consideraba la CFH como un proceso de identificación. Además, su metodología no empleaba análisis cuantitativos ni bases de datos de referencia. Por otro lado, se presenta el llamado *nuevo paradigma para la evaluación de las ciencias forenses*, fundamentado en que «el nuevo modelo de ciencia forense en el que los fonetistas empiezan a fijarse es el de la comparación de perfiles de ADN» (p. 53). Sumado a esto, se propone evaluar la evidencia en términos probabilísticos y utilizar bases de datos para contextualizar la variabilidad de la voz en la población. Metodológicamente, se emplean modelos estadísticos, especialmente el marco bayesiano, para cuantificar la fuerza de

la evidencia. En este sentido, cabe destacar que el marco bayesiano ha sido ampliamente adoptado en la ciencia forense, pero su implementación en la práctica forense internacional ha sido más lenta. Aunque la investigación académica apoya este enfoque, los peritos forenses aún encuentran obstáculos para aplicarlo en procedimientos judiciales debido a su complejidad y la falta de familiaridad de jueces y abogados con los métodos estadísticos. Según Morrison (2022), se producirá «un cambio generacional con más practicantes del nuevo paradigma, al que propone llamar, simplemente, *ciencia forense de datos*» (p. 58).

El siguiente capítulo, que lleva por título «*Aproximaciones metodológicas y parámetros más utilizados en comparación forense de hablantes*», profundiza en los métodos empleados para el cotejo de voces, explorando parámetros como la entonación, la duración de los sonidos y el timbre vocal. En primer lugar, se presentan tres enfoques principales para el cotejo de voces en la fonética forense: el método auditivo, basado en la percepción del analista experto, quien evalúa las características de la voz como el acento, la entonación y el ritmo del habla; el método acústico, que consiste en el análisis instrumental de la voz a través de software especializado, permitiendo medir parámetros como la frecuencia fundamental (F0), los formantes vocálicos y la duración de los segmentos de habla; y el método semiautomático, una combinación de los métodos anteriores con la intervención de herramientas tecnológicas avanzadas. En segundo lugar, también se mencionan los parámetros fonéticos empleados en la comparación de hablantes, los cuales se complementan en su análisis, ya que «se pueden analizar perceptiva y acústicamente» (p. 84), la frecuencia fundamental (F0), la frecuencia de los formantes (F1, F2, F3, etc.), la calidad de la voz, parámetros rítmicos, pausas y disfluencias, entre otros.

El quinto capítulo, «*La expresión de resultados en comparación forense de hablantes: una introducción al marco bayesiano de evaluación de la evidencia forense con datos fonéticos*», retoma la introducción realizada en el tercer epígrafe sobre el modelo bayesiano y profundiza en su uso, haciendo especial hincapié en su objetividad y transparencia a la hora de tratar pruebas fónicas. Este enfoque metodológico, ampliamente aceptado para la eva-

luación de evidencia en la CFH, se fundamenta en el concepto de razón de verosimilitud (*Likelihood Ratio*, LR) como herramienta clave para cuantificar la probabilidad de que dos muestras de voz pertenezcan a la misma persona. Como ya se ha comentado con anterioridad, el objetivo principal de la comparación forense de hablantes no es identificar a un individuo, sino estimar la probabilidad de que dos grabaciones de voz pertenezcan o no a la misma persona. Para esto, el marco bayesiano propone evaluar la evidencia considerando dos hipótesis: hipótesis del mismo locutor (Hml), ambas muestras provienen del mismo hablante; e hipótesis de diferente locutor (Hdl), las muestras pertenecen a hablantes distintos.

La razón de verosimilitud (LR) se calcula con la siguiente fórmula:

$$LR = \frac{P(E|Hml)}{P(E|Hdl)}$$

En ella, E representa la evidencia obtenida de los análisis fonéticos y acústicos. Si el LR es mayor que 1, la evidencia apoya la hipótesis de que ambas muestras provienen del mismo hablante; si es menor que 1, apoya la hipótesis de diferente hablante. Dado que el método bayesiano debe proporcionar índices de error, también se han desarrollado métricas para evaluar la precisión del sistema de comparación de voces. Entre ellas destacan: la función de coste de la razón de verosimilitud logarítmica (Clfr), que cuantifica la capacidad de discriminación del sistema; y el gráfico Tippett, que representa visualmente la distribución de los valores de LR para muestras de un mismo hablante y de hablantes distintos.

Para ilustrar de manera práctica el uso del marco, se presenta un caso donde se analizan trayectorias formánticas de vocales en muestras de voz. Se explica que estas características acústicas pueden cuantificarse mediante análisis estadísticos avanzados y aplicarse a sistemas de validación como los utilizados en la comparación forense de ADN. En primer lugar, se combinaron los *scores* mediante el Naïve Bayes, lo que evidencia que «las variables son independientes; es decir, que no están correlacionadas» (p. 101). En segundo término, se empleó un análisis de regresión logística, que «no solo sirve para fusionar, sino también para calibrar» (p. 102).

La penúltima sección, «Los límites de la variación inter- e intralocutor», destaca la importancia de seleccionar parámetros fonéticos adecuados para la CFH, centrándose en cómo este hecho afecta de manera directa a la identificación de hablantes en el contexto forense. Además, se subrayan los desafíos que enfrentan los sistemas automáticos y la pericia forense en la identificación vocal.

Antes que nada, para que un parámetro fonético sea útil en la CFH, debe cumplir con una serie de criterios, según Wolf (1972) y Nolan (1983): alta variabilidad interlocutor, el parámetro debe mostrar diferencias significativas entre hablantes distintos, permitiendo la identificación individual; baja variabilidad intralocutor, debe ser lo más estable posible dentro del mismo hablante a lo largo del tiempo; no ser modificable fácilmente, un parámetro robusto debe resistir intentos de disimulo o imitación de voz; ser frecuente en el habla natural, esto es, debe aparecer de manera habitual en la comunicación cotidiana; no verse afectado por el ruido o el medio de transmisión, debe ser resistente a distorsiones provocadas por factores externos, como el canal telefónico o el ambiente acústico; y ser fácilmente medible.

Asimismo, los factores que afectan la variabilidad en la voz pueden dividirse en tres grandes grupos: dependientes del hablante, que incluyen el sexo, la edad, la anatomía del tracto vocal y aspectos fisiológicos como enfermedades o el consumo de alcohol; dependientes de la conversación, como variaciones en la voz según el contexto comunicativo, hablar en situaciones formales o informales, o la presencia de ruido (efecto Lombard); y dependientes de la tecnología y el ambiente, en este caso factores como el tipo de micrófono, la calidad del audio y el canal de transmisión pueden alterar la señal acústica.

El tercer epígrafe de este capítulo está dedicado al estudio de un caso práctico de CFH, el zoo biométrico. Este concepto se emplea para clasificar a los hablantes según su comportamiento en los sistemas automáticos de reconocimiento de voz. Siguiendo este planteamiento, podemos encontrarnos con *ovejas*, hablantes estándar que no generan problemas para los sistemas; *cabras*, perfiles con gran variabilidad en su voz, lo que puede dificultar su identificación; *corderos*, individuos fácilmente imitables; y *lobos*, hablantes

con alta capacidad para imitar a otros hablantes, lo que puede generar errores en la identificación.

Por último, se mencionan los estudios de voces disimuladas. En contextos delictivos, es común que los hablantes intenten modificar su voz para ocultar su identidad. Algunos estudios han analizado cómo estos intentos afectan la identificación forense. Se ha observado que entre el 15% y el 20% de los casos en Alemania involucran algún tipo de disfraz de voz. Por otro lado, en Brasil el disimulo vocal es frecuente en secuestros. En el ámbito de las voces disimuladas, Gil y San Segundo (2013) mencionan una taxonomía tripartita: «aquellas que se centran en investigar si el disimulo es fácilmente detectable, aquellas que investigan qué tipo de transformaciones de la voz son más habituales y cómo afectan a la señal acústica, y aquellas que se preguntan hasta qué punto tales variaciones disminuyen la capacidad identificatoria de los oyentes» (pp. 120-121).

En último lugar, nos encontramos con «Nuevos retos y nuevas líneas de investigación en fonética forense» como capítulo de cierre. Concretamente, se examinan los principales desafíos que enfrenta la disciplina, incluyendo el impacto de la inteligencia artificial y los *deepfakes* en la autenticidad de la evidencia fonética.

Uno de ellos es el estudio de las voces de gemelos, un tema que ha suscitado gran interés en la fonética forense, dado que plantea dificultades para los sistemas de reconocimiento de voz. Los trabajos previos han fijado una serie de objetivos fundamentales en el desarrollo de esta labor. Primeramente, se debe explorar la influencia genética en la voz, algunas investigaciones han intentado determinar qué características vocales tienen una base genética y qué tan heredables son los rasgos fonéticos. En segundo lugar, se tiene que evaluar la capacidad de los sistemas forenses para distinguir gemelos. Con respecto a esto, se han realizado pruebas para determinar la efectividad de los parámetros fonéticos en la diferenciación de la voz de gemelos idénticos y no idénticos. Los resultados han mostrado que los gemelos idénticos (monocigóticos) presentan un mayor grado de similitud vocal, lo que representa un gran desafío para los sistemas de identificación. También se han registrado casos de fraude y suplantación de identidad en los que gemelos han utilizado sus similitudes

vocales para engañar sistemas de seguridad, como ocurrió en un caso del banco HSBC, «cuando el sistema de reconocimiento de voz fue atacado por el gemelo de un cliente» (p. 127).

Por otro lado, es innegable lo que ha supuesto para la sociedad mundial el avance en términos de inteligencia artificial, que, entre otros asuntos, ha dado lugar a los *deepfakes* de voz, es decir, clonaciones artificiales de voces que pueden imitar con precisión a cualquier hablante. Esto representa un riesgo significativo en la autenticación biométrica y en el ámbito forense, ya que puede comprometer la validez de las pruebas de voz. Esto entraña una serie de riesgos y desafíos a los que los lingüistas forenses deben hacer frente, como el uso fraudulento en suplantación de identidad, la manipulación de audios para difamación o desinformación o la posibilidad de alterar pruebas en el ámbito legal. No obstante, podemos encontrar aplicaciones positivas, dado que en ocasiones se ha utilizado para ayudar a personas con enfermedades neurodegenerativas a recuperar su voz o a mejorar los asistentes virtuales de accesibilidad.

El libro destaca por su enfoque multidisciplinar, pues combina aspectos lingüísticos, estadísticos y tecnológicos para ofrecer un análisis exhaustivo de la fonética forense. La autora presenta una revisión bibliográfica detallada y expone ejemplos prácticos que ilustran la aplicación de los diversos métodos planteados a lo largo de los epígrafes.

Esta obra constituye un recurso fundamental para investigadores y profesionales en lingüística aplicada, criminología y ciencias forenses. Su enfoque didáctico permite que tanto especialistas como estudiantes en formación comprendan los aspectos técnicos y metodológicos de la disciplina. Con todo, este trabajo se ha consagrado como un referente bibliográfico, una guía necesaria para comprender los diferentes conceptos y herramientas que se abordan en el ámbito de la fonética forense.

## BIBLIOGRAFÍA

- CICRES, Jordi (2011). Transcripció i autenticació de gravacions en contextos judicials. *LSC-Llengua, societat i comunicació*, 9, 26-32.

- COOPER, Alan J. (2009). An automated approach to the Electric Network Frequency (ENF) criterion: theory and practice. *International Journal of Speech, Language & the Law*, 16 (2). <http://dx.doi.org/10.1558/ijssl.v16i2.193>.
- GIL, Juan y SAN SEGUNDO, Eugenia (2013). El disimulo de la calidad de voz en fonética judicial: un estudio perceptivo para un caso de pinzamiento de nariz. En Azucena Penas (Ed.), *Panorama de la Fonética Española Actual* (pp. 321-366). Arco Libros.
- GRIGORAS, Catalin (2005). Digital audio recording análisis—the electric network frequency criterion. *International Journal of Speech Language and the Law*, 12 (1), 63-76. <http://dx.doi.org/10.1558/sll.2005.12.1.63>.
- MORRISON, Geoffrey S. (2022). Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science. *Forensic Science International: Synergy*, 100-270.
- NOLAN, Francis (1983). *The phonetic bases of speaker recognition*. Cambridge University Press. [http://dx.doi.org/10.1016/0167-6393\(87\)90039-2](http://dx.doi.org/10.1016/0167-6393(87)90039-2).
- SAKS, Michael J. y KOEHLER, Jonathan J. (2005). The coming paradigm shift in forensic identification science. *Science*, 309 (5736), 892-895.
- WOLF, J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America*, 51 (6B), 2044-2056.

Humberto GARCÍA PÉREZ

<https://orcid.org/0009-0002-1893-9078>

Universidad de Las Palmas de Gran Canaria

Las Palmas de Gran Canaria, España

DOI: <https://doi.org/10.25145/j.refull.2025.51.29>

