



Universidad de Valladolid



CSIC



UNED

Homenaje a Antonio Quilis

Madrid, 19 de abril de 2005

Distinguido colega y amigo:

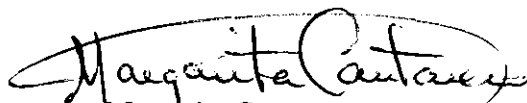
Le enviamos las pruebas de imprenta del trabajo con el que va a contribuir al Homenaje al profesor Antonio Quilis, con el ruego de que las reenvíe corregidas antes del 20 de mayo de 2005. Si no las hemos recibido una vez vencido ese plazo, entenderemos que no necesitan corrección. La revisión de las segundas pruebas será tarea de la comisión organizadora del Homenaje.

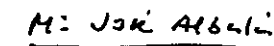
Le rogamos devuelva las pruebas a la siguiente dirección:

Margarita Cantarero
Instituto de la Lengua Española
Consejo Superior de Investigaciones Científicas
Duque de Medinaceli, 6
28014 Madrid

Le agradecemos sinceramente su colaboración

Con todo afecto, en nombre de la comisión,


Margarita Cantarero


María José Albalá

FILOLOGÍA Y LINGÜÍSTICA
ESTUDIOS OFRECIDOS
A ANTONIO QUILIS

Volumen II

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA
UNIVERSIDAD DE VALLADOLID
MADRID, 2005

HACIA UNA SINERGIA METODOLÓGICA EN LA
IDENTIFICACIÓN DE LOCUTORES

RAMON CERDÀ
Universitat de Barcelona

MIREIA FARRÚS
JAVIER HERNANDO
Universitat Politècnica de Catalunya

THE NOTE TAKER. [...] And how are all your people down at Selsey?
THE BYSTANDER. [suspiciously] Who told you my people come from Selsey?

THE NOTE TAKER. Never you mind. They did. [To the girl] How do you come to be up so far east? You were born in Lisson Grove.

THE FLOWER GIRL. [appalled] Oh, what harm is there in my leaving Lisson Grove?

[...]

THE SARCASTIC BYSTANDER. [...] Do you know where I come from?

THE NOTE TAKER. [promptly] Hoxton.

Titterings. Popular interest in the note taker's performance increases.

THE SARCASTIC ONE. [amazed] Well, who said I didn't? Bly me! You know everything, you do.

[...]

THE BYSTANDER. You take us for dirt under your feet, don't you? Catch you taking liberties with a gentleman!

THE SARCASTIC BYSTANDER. Yes: tell HIM where he come from if you want to go fortune-telling.

THE NOTE TAKER. Cheltenham, Harrow, Cambridge, and India.

THE GENTLEMAN. Quite right.

[...]

THE GENTLEMAN. [...] How do you do, if I may ask?

THE NOTE TAKER. Simply phonetics. The science of speech.

George Bernard Shaw, *Pigmalion*, Act I.

1. PRESENTACIÓN

1.1. Este fulgurante comienzo de la obra *Pigmalión* provocó, y provoca, el asombro de cualquier espectador medio, pero también del más avezado de los fonetistas. Ahí es nada, determinar por la pronunciación de la gente de dónde procede y por dónde ha vivido supone un control absoluto de muchísimas variables lingüísticas y, lo que es más inverosímil todavía, de su concatenación en el tiempo. No creemos que el autor conociera a nadie capaz de realizar semejante proeza tal como la describe allí¹.

De lo que no cabe la menor duda es de que la fonética —la ciencia del habla, como se define en la obra de Shaw— sirve para satisfacer una gran cantidad de necesidades prácticas. Enumerando muy en general sólo las principales, podemos mencionar la de enseñar la pronunciación de una lengua extranjera, perfeccionar la pronunciación de la lengua propia con fines artísticos, didácticos o terapéuticos, crear sintetizadores de voz para suministrar información, dialogada o no, o bien analizar los parámetros de un usuario para múltiples aplicaciones, como sistemas de dictado o de verificación y determinación de la identidad personal.

1.2. Aquí nos ocuparemos de esta última, la determinación de la identidad de una fuente de voz con fines forenses, es decir, la identificación de un locutor entre distintos candidatos posibles como autor de unas expresiones pronunciadas en forma espontánea. Como el escenario habitual o típico en el que se dilucida este planteamiento es policial y judicial en causas generalmente penales (por ejemplo, en un delito de narcotráfico donde se trata de determinar si la voz grabada por la policía a partir de conversaciones telefónicas pertenece o no a la persona imputada), existen diversas perspectivas de observación y, en consecuencia, distintas sensibilidades y aun distintas interpretaciones. Recurriendo a una imagen muy simplificada, hagamos un breve resumen de los principales grupos estamentales que concurren en los encuentros científicos sobre acústica forense:

1) Jueces y magistrados. Constituyen la cima de la pirámide, en el sentido de que todos los demás estamentos tratan de convencerles sobre la verosimilitud de sus respectivos cometidos. A juzgar por la jurisprudencia que existe y por declaraciones explícitas, puede afirmarse que ni siquiera el grado máximo de certeza en una identificación de locutor tiene para ellos efectos decisivos si no concurren circunstancias probatorias adicionales. Perciben, por tanto, un in-

¹ En realidad Bernard Shaw se inspiró en la personalidad de su amigo Henry Sweet (1845-1912), conspicuo fundador de la escuela fonética británica, precursora del estructuralismo y del análisis prosódico. De esta obra existe, por lo demás, una conocida versión cinematográfica de 1964 en clave musical, *My fair lady*, dirigida por George Cukor e interpretada por Rex Harrison y Audrey Hepburn en los principales papeles.

discutible factor de incertidumbre que no se presenta ante otros tipos de pruebas periciales consideradas concluyentes, como el ADN, las huellas digitales o la balística.

2) Abogados y asesores jurídicos. Por lo común, desempeñan una labor relativamente marginal en esta clase de asuntos en el sentido de que no suelen dominar —lo mismo que sucede con los jueces y magistrados— ninguno de los aspectos técnicos de las pruebas periciales. Peor aún, se quejan con razón de que, al menos por el momento, en España apenas existen laboratorios alternativos o independientes donde practicar contrapruebas o simplemente verificaciones distintas a las policiales.

3) Médicos y cirujanos. La mayor parte de sus contribuciones consisten en mostrar las alteraciones que suelen experimentar las emisiones fónicas a causa de distintos trastornos y patologías que repercuten sobre la respiración o sobre el funcionamiento de la laringe y los órganos supraglóticos.

4) Fonetistas y lingüistas. La formación más típica de quienes intervienen en estos asuntos es la fonética y la dialectología. Su perspectiva gira en torno a la interpretación lingüística de las emisiones sobre todo a partir de la transcripción de datos espectrográficos. Para ser precisos, dicho en términos técnicos, su cometido consiste en realizar, con la ayuda de distintas técnicas, la transcripción fonética más estrecha posible no sólo para establecer adscripción dialectal del locutor —como en *Pigmalión*—, sino para distinguirlo e identificarlo entre sus propios paisanos.

5) Agentes de seguridad. En España al menos, todavía son prácticamente los únicos que cuentan con personal dedicado en exclusiva a los análisis periciales de identificación de voz y similares. Sus técnicas, sin embargo, suelen apoyarse muy prioritariamente en los procedimientos metodológicos de los ingenieros de telecomunicaciones.

6) Ingenieros de telecomunicaciones. En las reuniones de especialistas son los que pisan más fuerte, dicho sea de una manera gráfica, frente al resto de concurrentes, bajo el supuesto de que sus procedimientos son objetivos porque trabajan sobre las características físicas del sonido y, por tanto, no están supeditados a valoraciones subjetivas. La fiabilidad —o, mejor, la verosimilitud— de sus aportaciones, típicamente presentadas en estimaciones estadísticas, alcanza los niveles relativos más elevados.

Entre estos grupos estamentales existen evidentes diferencias que admiten distintas agrupaciones. En primer lugar, hay que distinguir entre el estamento jurídico, 1 y 2, y el estamento científico, los demás, y, dentro de este último, más complejo, el de quienes suelen trabajar sobre generalidades, 3, y el de quienes lo hacen sobre datos concretos, grabaciones, el resto. Finalmente, entre estos últimos se distinguen al menos quienes se apoyan preferentemente, o casi, sobre datos de alto nivel (dialectos, expresiones y realizaciones fonológicas), 4, y quienes lo hacen exclusivamente, o casi, sobre datos de bajo nivel (distribución de frecuencias, intensidades, duraciones...), 5 y 6. Otra clasificación permite distinguir todavía entre los estamentos que tienden a arrojar

dudas, cuando no sombras, sobre los resultados —1, 2 y 3— y los que intentan arrojar luz al menos sobre la fiabilidad de los métodos, el resto.

En cualquier caso, lo más importante es que los encuentros de trabajo de especialistas en acústica forense se materializan en cuatro o cinco tipos de discurso coexistentes que se comportan, al menos en lo fundamental, como si fuesen compartimentos estancos conceptuales, dicho sea simplificando otra vez.

Ahora bien, como los aspectos técnicos y científicos se circunscriben básicamente a dos apartados —que hemos distribuido entre alto y bajo nivel, es decir, entre la interpretación de datos lingüísticos y la interpretación de datos físicos a partir de registros de emisiones de voz—, en lo que sigue vamos a exponer las características predominantes de ambas perspectivas con el fin de sugerir algunas premisas para su cooperación óptima, más o menos integral, y la obtención consecuente de resultados que permitan alcanzar valores máximos de certeza al menos entre los tres últimos estamentos.

2. RESUMEN EVALUATIVO DE LAS ESTRATEGIAS TÍPICAS DEL LINGÜISTA

2.1. Enfoque epistemológico

Desde su habitual perspectiva, el lingüista especializado en fonética —o, tal vez mejor, el fonetista clásico especializado en lingüística— nunca atiende a datos que no sean interpretables lingüísticamente. En rigor, podemos definir la parte más genuina de su tarea cuando analiza, pongamos, un espectrograma, diciendo que se adentra en la maraña aparente de rasgos para localizar y aislar aquellos que identifican acústicamente la información lingüística casi siempre en torno a una noción de fonema o entidad muy similar. Dicho aún de otro modo, trata de leer el espectrograma como si fuese una escritura alternativa. Y sólo a partir de aquí puede, desde luego, decidir entre varios cometidos, de mayor a menor cobertura teórica:

1) Si la pronunciación de partida es considerada lingüísticamente 'modélica', puede establecer las características acústicas de los fonemas de una lengua tanto aislados como en secuencia, así como de sus entornos enunciativos.

2) Si el modelo caracteriza un dialecto —en cualquiera de los sentidos que cabe atribuir al término—, puede establecer lo mismo del apartado anterior para ese dialecto.

3) Y si el modelo caracteriza un hablante —por algún tipo de notoriedad: un literato, un político, un actor, etc.—, puede igualmente establecer lo mismo para ese hablante.

En apariencia al menos, es evidente que los resultados de estas indagaciones pueden ser comparados con los de otras indagaciones homogéneas y sa-

car múltiples conclusiones de todo ello. Pero la experiencia demuestra que los lingüistas se pierden con frecuencia ante comparaciones de esta índole por la sencilla razón de que, aun a pesar de su enfoque, encuentran dificultades para discernir lo esencial de lo circunstancial. Comparando los esquemas vocálicos de dos o más dialectos extraídos a partir del cómputo de los dos primeros formantes, no es raro, por ejemplo, que interpreten como si fuesen 'lingüísticas' las diferencias que observan entre las distancias geométricas relativas o la situación total o parcial respecto de las coordenadas frecuenciales (Cerdà, 2000). Es decir, entienden como dialectales, por ejemplo, las diferencias de separación media entre áreas de dispersión vocálica cuando en realidad no hay dos hablantes que coincidan al respecto. Y es que sin alcanzar un número estadísticamente significativo de comprobaciones, es difícil interpretar como puramente individuales o no las enormes diferencias que puede haber en las emisiones fónicas hasta de los miembros de una misma familia, todos ellos hablantes monolingües de un mismo y único dialecto —como ya sucediera en el trabajo clásico de Rousselot (1891).

Peores todavía son los estudios sobre rasgos suprasegmentales, donde las referencias son intangibles o están como mínimo difuminadas en el conjunto de datos. En efecto, todos los rasgos que comportan modulaciones suprasegmentales de la voz son el resultado de un conjunto de factores integrados, algunos fonológicos (como la oposición entre distintos perfiles de pregunta; p. ej., ¿Vienes? / ¿Cuándo vienes?), otros pragmáticos (como la oposición, dentro de la afirmación *Sí*, entre el asentimiento a lo que se dice, [sí], y la réplica a una negativa, [si:]), otros dialectológicos (p. ej., la entonación mexicana), otros psicológicos (cuando se pide algo, pongamos, con insistencia) y otros más, más o menos automáticos, que resultan a menudo de combinaciones por lo común inextricables de estos factores. Véase un ejemplo de ello en Sosa (1999).

En realidad, la lingüística moderna procede de una conquista epistemológica alcanzada durante la primera mitad del siglo pasado precisamente cuando se dilucidó qué datos, entre todos, son cruciales para la interpretación lingüística y se aplicaron a la noción de 'fonema' (Cerdà, 1996). Gracias a la comprobación combinada de la espectrografía y la síntesis del lenguaje, pudo establecerse que los primeros formantes caracterizaban las vocales, los silencios y las transiciones formánticas conformaban las consonantes oclusivas, los ruidos eran propios de las sibilantes y así hasta catalogar lo más esencial, acústicamente, de cada fonema². A partir de esta premisa, el cometido del fonetista lingüista consistía, pues, en identificar, aislar y definir estos datos de-

² No hay más que recordar obras clásicas como Jakobson, Fant y Halle (1955), los trabajos de P. Delattre o el binarismo de R. Jakobson.

jando el resto al margen como algo propio de las características individuales del hablante y de las condiciones concretas de la emisión.

Bien, pues, aunque al menos desde entonces los fonetistas lingüistas han tenido la convicción de que un sólo hablante nunca pronuncia exactamente del mismo modo una misma emisión —pongamos, *mesa, mesa, mesa...*—, siempre que pueda medirse con suficiente detalle, lo cierto es que no siempre, ni mucho menos, han llegado a tener en cuenta que las características individuales del hablante y de las condiciones concretas de la emisión se hallan *igualmente* presentes en los datos que caracterizan los fonemas.

Lo malo es que, aunque esté convencido de que los datos individualizadores son efectivamente omnipresentes, al menos el fonetista lingüista tradicional no dispone de recursos ni teóricos ni metodológicos para aislarlos y luego, según convenga, desecharlos o trabajar con ellos (Cerdà y Veyrat, 2000; Cerdà, Muñiz y Veyrat, 2003).

2.2. Enfoque lingüístico

En Quilis (1999) se encuentra sin duda la exposición sumaria más clara y explícita de la labor que realiza el fonetista lingüista ante la tesitura de determinar si una voz indubitada grabada es la misma o no a la que se encuentra en una grabación previa única o múltiple. Ante todo insiste en la multitud de despropósitos de todo tipo que dificultan esa labor; algunos técnicos e involuntarios, como la baja calidad de la señal debido a sistemas de grabación deficientes, capaces de distorsionar considerablemente los datos al escuchar la grabación en un sistema diferente al de registro; y otros producidos por una mala concepción de cómo optimizar las pruebas, en especial cuando se exige al imputado que repita determinadas frases por medio de la lectura sin advertir que la lectura misma introduce numerosas pautas fonológicas y fonéticas a menudo decisivamente distintas a las del habla espontánea.

En un ámbito ya más propiamente lingüístico, también destaca las grandes diferencias que sobre la pronunciación de las palabras impone asimismo no sólo el contexto en que aparecen, sino también el ritmo, el tempo, el énfasis y multitud de factores externos como el cansancio, el estado de ánimo o el ansia de fingimiento del emisor.

A partir de aquí la labor del fonetista lingüista orientada hacia la identificación del locutor se distribuye en dos niveles básicos de trabajo metodológicamente solidarios: el espectrográfico y el dialectológico. En el primero, trata de caracterizar las pautas acústicas más estables de su realización fónica a partir de la frecuencia del fundamental, el timbre vocálico diferencial entre vocales tónicas y átonas, la frecuencia de inicio de [s], la caracterización del sonido esvarabático y otros rasgos especialmente significativos para alcanzar

una adscripción dialectal precisa y, a ser posible, perfilar una imagen del habla máximamente individualizada.

La indagación dialectológica discurre en un nivel interpretativo superior pues comporta la identificación directa de datos lingüísticos de toda índole —léxicos, morfológicos, sintácticos...— con el propósito de situar al locutor en un ámbito tan reducido como se pueda desde el punto de vista de su procedencia, los lugares donde haya vivido, el nivel cultural o la profesión (en una auténtica réplica al fonetista de *Pigmalión*). A ello se suman todavía otras exploraciones de un máximo poder identificativo desde el punto de vista del lingüista: los datos paralingüísticos que acompañan toda emisión espontánea, en especial las expresiones fáticas y expletivas, es decir, respectivamente, aquellos sonidos, palabras o expresiones que sirven para abrir un canal de comunicación (*¡Oye!*, *¡Eh..!*) o para mantenerlo abierto (*Ya, Sí, Claro, ¿Verdad?...?*) y los que sirven para poner de relieve una actitud o un dato comunicativo (*¡Hombre! ¿Qué dices? ¡No te digo yo! ¡Ni hablar! ¡Jol!, ¡Toma ya!...*).

De todo ello se desprenden varias consideraciones:

- 1) La labor del fonetista lingüista sólo se ejerce cabalmente a partir de la comprensión de los datos comunicativos de las grabaciones. Todo ruido, distorsión o deficiencia de cualquier tipo menoscaba de un modo esencial la capacidad de conseguir una evaluación útil.
- 2) En teoría, hasta en la mejor de las caracterizaciones espectrográficas y dialectológicas, el fonetista lingüista no podrá pronunciarse de un modo inequívoco sobre la identificación precisa de un sólo locutor, pues por muy precisos que sean los datos individualizadores es casi imposible que no afecten, al menos presuntamente, a otros hablantes de su mismo entorno dialectal, social, cultural, profesional...
- 3) En la práctica, sin embargo, no se plantea nunca la necesidad estricta de caracterizar individualmente a un hablante entre sus congéneres más próximos, sino de establecer las probabilidades de que dos grabaciones pertenezcan o no a un mismo locutor, lo que ya concede un grado relativamente importante de fiabilidad metodológica a la labor del fonetista lingüista... en función, eso sí, de la calidad de las grabaciones.

Una conclusión parece imponerse clamorosamente. Y es que es urgente sustituir la memorización abierta, de tan extraordinariamente extensa, a que se ve sometido el fonetista lingüista por bancos de datos donde se clasifiquen y caractericen todos los rasgos dialectales (geográficos, sociales, culturales, profesionales...) por lenguas, de modo que puedan cruzarse rasgos de cualquier procedencia hasta conseguir un alto nivel de identificación individual, tal vez, casi sin duda, decisivamente alto.

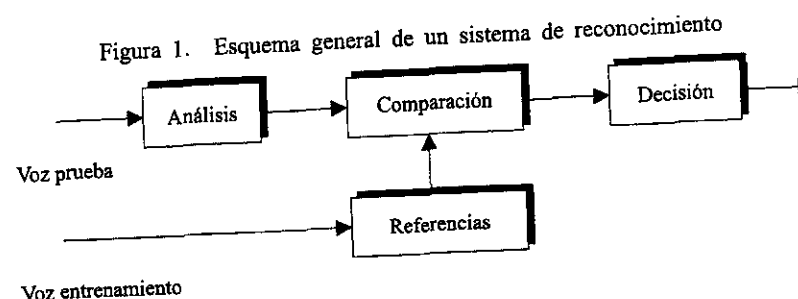
Actualmente no hay —no puede haber— óbice para igualar y hasta superar (salvo, acaso, en la cronología relativa de los datos) al sagaz fonetista que vislumbraba Bernard Shaw.

3. RESUMEN EVALUATIVO DE LAS ESTRATEGIAS TÍPICAS DEL INGENIERO

En el ámbito de la ingeniería, el interés por la tecnología del reconocimiento automático de personas por su voz ha aumentado fuertemente en los últimos años debido a la rápida convergencia de las telecomunicaciones y la informática, al uso creciente de interfaces orales en las comunicaciones hombre-máquina y a la necesidad de manejar voz en las aplicaciones multimedia. Existe un amplio rango de aplicaciones: controles de acceso a recintos físicos o virtuales, máquinas y contenidos; transacciones financieras y comerciales en general; indexado de audio de reuniones y programas de radio y televisión; por supuesto, investigación policial y resoluciones judiciales... En las aplicaciones comerciales, el reconocimiento oral de personas obtiene resultados plenamente aceptables y tiene evidentes ventajas sobre los métodos clásicos basados en llaves, tarjetas y claves, que pueden ser olvidadas, perdidas o robadas, y sobre otros métodos de identificación biométrica, como el rostro, la huella digital o el ADN, que precisan de dispositivos más sofisticados. En cuanto a las aplicaciones forenses, hay que hacer notar que los resultados proporcionados por la ingeniería han mejorado notablemente, pero son insuficientes para resolver por sí solos todos los casos.

Como hemos dicho antes, ello es debido a que el método de producción del habla es complejo y la señal de voz depende de las características físicas del locutor, tales como las dimensiones de las cuerdas y el tracto vocales, el entorno y el estado emocional, los hábitos fonatorios y lingüísticos aprendidos... Por ello, hay una gran variabilidad en la señal de voz entre locutores y, más importante, en la voz del mismo locutor en diferentes ocasiones. Otra pronunciación del mismo mensaje lingüístico emitida por la misma u otra persona puede presentar un aspecto bien distinto. Así pues, una selección cuidadosa de los parámetros acústicos es crucial para la eficacia de los sistemas automáticos de reconocimiento. Los parámetros usados por estos sistemas deben discriminar entre locutores y ser transparentes a los cambios producidos en un mismo locutor (variabilidad intra-locutor), ser fácilmente medibles a partir de la señal de voz, ser invariantes con el tiempo, no ser susceptibles de imitación por impostores. En general, como decíamos en 1.2, suele distinguirse entre rasgos de alto y de bajo nivel. Los rasgos de alto nivel, como dialecto, léxico, entonación, duración, etcétera, son usados frecuentemente por las personas para identificar a otras. Se han hecho esfuerzos para incorporar estos rasgos en los sistemas automáticos, pero no se ha tenido éxito debido a la dificultad de cuantificarlos. Por ello, los sistemas operativos usan parámetros de bajo nivel, tales como el tono, las magnitudes espectrales, las frecuencias de los formantes, los perfiles de energía, etc., derivados de medidas acústicas sobre la señal de voz (Campbell, 1997).

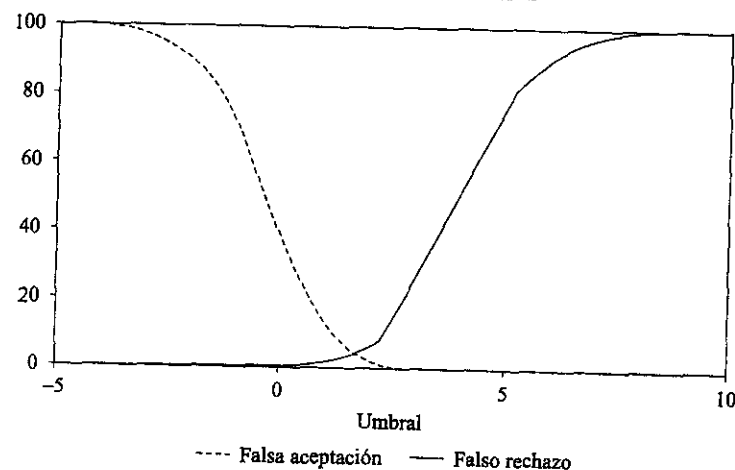
A partir de estos parámetros, el modo general de operación consiste en generar unos patrones o modelos estadísticos a partir de señales de voz de un locutor determinado. El objetivo es que el sistema aprenda las características específicas del locutor en cuestión, así como los márgenes de variabilidad de cada parámetro en estudio. Los parámetros se calculan mediante técnicas de estimación espectral para todas las señales disponibles en intervalos de algunos milisegundos y la decisión se establece mediante medidas de distancia entre señales especialmente diseñadas para esta aplicación. El proceso se ha representado gráficamente en la figura 1. Se trata de un método científico en el sentido de que permite que los experimentos y pruebas sean repetibles, es decir, no dependen de la persona que los realiza y los resultados son siempre los mismos.



Dependiendo de la aplicación, se distinguen dos tareas específicas en el reconocimiento de locutores: identificación y verificación. En identificación de locutores, el objetivo es determinar cuál de un grupo de voces conocidas encaja mejor con la muestra de voz de entrada. En verificación, el objetivo es determinar a partir de una muestra de voz si una persona es quien dice ser, que suele ser la situación habitual en las aplicaciones forenses. Esta tarea requiere distinguir la voz del locutor que reclama ser reconocido frente a un grupo potencialmente grande de voces desconocidas para el sistema. Para ello, se ha de ajustar un umbral de decisión en función del compromiso entre rechazar las elocuciones del locutor verdadero (errores de falso rechazo) y aceptar las elocuciones de locutores falsos (errores de falsa aceptación), ver figura 2.

Se distingue también entre sistemas independientes o dependientes del texto. En un sistema independiente del texto, el mensaje es completamente libre. En un sistema dependiente del texto, el locutor ha de pronunciar un mensaje lingüístico predeterminado, cuyo conocimiento permite mejorar las prestaciones. En estos sistemas se emplea frecuentemente una secuencia aleatoria

Figura 2. Curva típica de las tasas de falsa aceptación y falso rechazo en función del umbral de decisión



de dígitos o palabras predefinidas, como prevención del uso de grabaciones de audio por determinados intrusos. Mientras que los sistemas de texto fijo son principalmente usados en aplicaciones de control de acceso donde el cliente es cooperante, los sistemas de texto libre son necesarios en aplicaciones policiales en que el usuario es no cooperante y, a menudo, no es consciente de la tarea.

Los resultados experimentales indican que el mayor factor limitante de las prestaciones de los sistemas actuales son las distorsiones introducidas por el canal de transmisión, el ruido ambiente y la variabilidad del micrófono. También pueden ser importantes ligeras enfermedades o molestias respiratorias, e incluso el estado emocional del hablante... (Rosenberg, 1976).

4. EL LINGÜISTA VISTO POR EL INGENIERO

Los métodos de identificación de locutor desarrollados durante los últimos años por los ingenieros no han logrado una fiabilidad total, pero sí unos niveles muy altos de aceptación. Sin embargo, sus análisis se centran, como ya se ha comentado anteriormente, en los datos de bajo nivel. Si bien recientemente se ha prestado mayor atención al estudio de la entonación en el ámbito del procesado de voz en general, su aplicación principal es la de mejorar sistemas de síntesis de voz, y no tanto de reconocimiento.

Así pues, hay que tener en cuenta que los lingüistas exploran características que los ingenieros pasan por alto. La subjetividad, aunque *a priori* sea un factor negativo en cualquier tipo de análisis, puede resultar imprescindible para complementar algunas carencias de los sistemas automáticos. La tarea del lingüista es descriptiva y manual, pero, ¿acaso en la traducción no es necesaria la supervisión del traductor humano para suplir las carencias y dificultades de los traductores automáticos?

Los espectrogramas han sido la herramienta esencial del lingüista para el análisis de los datos de bajo nivel. Sin embargo, además de ser una herramienta con poca fiabilidad debido a la gran variabilidad intra-locutor, el análisis de datos es cuantitativamente limitado, ya que un lingüista experto podría llegar a comparar algunas decenas de espectrogramas, pero no más. Es necesario, pues, que el lingüista utilice su arte no para suplir el trabajo de los ingenieros, sino para complementarlo, eso es, para concentrar su análisis en los rasgos que los sistemas automáticos todavía no pueden modelar: la entonación, las expresiones lingüísticas o los dialectos.

5. EL INGENIERO VISTO POR EL LINGÜISTA

La mejor cualidad del ingeniero respecto al lingüista es, sin duda, la objetividad. El ingeniero trabaja con resultados analíticos que permiten una comparación mucho más exhaustiva que los obtenidos por el lingüista. Sin embargo, las herramientas del ingeniero no permiten, en algunos casos, una exploración tan minuciosa como la que haría el lingüista.

Los sistemas de reconocimiento de voz utilizados por los ingenieros suponen que, en cada tramo de análisis, las características estadísticas de la señal no varían con el tiempo; esta aproximación es muy válida en la mayoría de sonidos del habla continua, pero no en todos: las oclusivas, que se caracterizan por cambios muy bruscos de sonoridad, no se pueden modelar con estos sistemas. Sin embargo, aunque esto puede dificultar el reconocimiento del habla, debido a que las oclusivas tienen una situación privilegiada a la hora de marcar y fragmentar una secuencia hablada por su característica zona de silencio, no constituye un inconveniente tan problemático en el reconocimiento del locutor, por la sencilla razón de que la ausencia de sonoridad aporta muy poca información al respecto.

Además, es bien sabido que en algunos casos de investigación, la pista decisiva se ha basado en características de alto nivel. Existen ejemplos especialmente ilustrativos al respecto, y en especial de lo importantes que resultan las expresiones expletivas —comunicativamente gratuitas, en principio— para caracterizar un estilo individualizado de habla. Así, en un caso de secuestro acaecido hace unos quince años pudieron reunirse sucesivamente gra-

baciones de una mujer con acento gallego y dos hombres, uno con deje rioplatense y otro de habla castellana indeterminada. Ante la perspectiva de una ardua identificación policial plagada de incertidumbres, alguien advirtió que los tres personajes remataban muchas frases con la coletilla *¿Entiende usted...?* A partir de ello se emitió la hipótesis de que podía tratarse de una sola persona, seguramente una mujer, que habría vivido en Galicia y en Argentina (*¿de nuevo Pigmalión!*). Cruzando estos datos entre personas con antecedentes, pudo identificarse a una mujer que satisfacía todos aquellos supuestos y que era efectivamente la autora de todas las llamadas. Así, pues, aquel dato, que nada tenía que ver con el nivel bajo del lenguaje, delató al locutor y permitió su posterior identificación.

Aun así, es indiscutible que los ingenieros alcanzan niveles muy altos de fiabilidad y pueden analizar características de bajo nivel que serían imperceptibles al oído humano. Más aún, la automatización de los sistemas permite trabajar con una gran cantidad de indicios y el análisis puede basarse en la comparación de grandes bases de datos ya existentes, tarea que resultaría inalcanzable para el lingüista. Así pues, el trabajo del ingeniero resulta, en la mayoría de los casos, totalmente indispensable. Es más, no parece en absoluto descabellado esperar que en un futuro seguramente próximo alcance un nivel de seguridad y de automatización como el que exhiben las comprobaciones dactilares.

6. PROPUESTAS SINÉRGICAS

Aunque tradicionalmente los lingüistas y los ingenieros han trabajado de una forma individual y con sus propias herramientas, en la actualidad es harto evidente la necesidad de que cooperen entre sí, sobre todo en el ámbito de la acústica forense que tratamos aquí. Interdisciplinariedad sí, por supuesto, pero, a estas alturas, hace falta un paso más. Habiendo reconocido la necesidad de una sinergia metodológica de ingenieros y lingüistas, debemos preguntarnos cómo debe ser esa sinergia, y qué características debe tener esta cooperación tan reivindicada, qué conocimientos deben compartir y con qué conocimientos y métodos pueden complementarse las dos partes especializadas.

En un primer paso, es precisa una puesta en común de conocimientos y de terminología utilizada. Después, hay que determinar cuáles son las herramientas que son únicamente usadas por los lingüistas y podrían ayudar a los ingenieros en sus análisis y viceversa, así como los métodos y herramientas que pueden utilizar las dos partes indistintamente. Algo así sólo es posible creando un marco académico que integre todos los aspectos pertinentes y sus respectivos fundamentos teóricos y prácticos.

Mucho más aún. Como decía el propio Antonio Quilis, habría que imitar el célebre Real Instituto de Tecnología (KTH) de Estocolmo que, tiempo atrás, dirigía su amigo Gunnar Fant, integrando ingenieros, fonetistas, lingüistas, logopedas, otorrinos, músicos e... inventores³. Sobre todo, gente con imaginación, perspectivas interdisciplinares y espíritu de cooperación. Es la única forma, reconocía con razón indiscutible, de concebir, diseñar y construir efectivamente cualquiera de los artefactos, o los sistemas, cada vez más precisos que se utilizan en la experimentación actual. Es decir, siendo capaz de percibir los problemas en cualquier sitio y hallar, o anticipar, su solución en cualquier otro.

Es lo que ofrece también el Instituto de Policía Científica y Criminología en Lausanne⁴, pionero en su envidiable programa de estudios forenses, fundado en 1909, donde se integran matemáticas, informática, física, derecho, química, estadística, fotografía, microscopía, análisis instrumental, medicina legal, criminología y a largo etcétera, y que, junto con la Universidad de Strathclyde (Escocia), ofrece una titulación modélica con estas características.

Las diferentes perspectivas deben reunirse en un nuevo enfoque de la organización de los estudios universitarios que cubran todos los ámbitos forenses, tomando ejemplo de los casos mencionados. Sólo así, la acústica forense tendrá un sentido completo. Entretanto, es urgente que los lingüistas y los ingenieros definan un protocolo de actuación conjunta, sistematizando y estipulando la metodología que habrá de utilizarse en cada caso. Como decíamos antes, el lingüista debe ser capaz de complementar sus conclusiones con las del ingeniero y concentrarse en el análisis de aquellos aspectos inasequibles a los sistemas automáticos, como, por ejemplo, los expletivos, los dejes dialectales o las pautas entonativas. También debe ser tarea del lingüista aplicar la sistemática utilizada por los ingenieros no sólo en la teoría, sino también en sus aplicaciones, con el fin de optimizar sus herramientas. Y ante todo, también, el ingeniero no debe olvidar que la lingüística existe.

BIBLIOGRAFÍA

- CAMPBELL, J.P. (1997): "Speaker recognition: A tutorial", *Proceedings of the IEEE*, vol. 85, págs. 1437-1462.
- CERDÀ, R. (1996): "La función del sonido", en C. MARTÍN VIDE (ed.), *Elementos de lingüística*, Barcelona, Octaedro, págs. 129-169.
- (2000): "Letras, sonidos y su organización", en M. ALVAR (comp.), *Introducción a la lingüística española*, Barcelona, Ed. Ariel, págs. 107-137.

³ <http://www.speech.kth.se/>

⁴ <http://www2.unil.ch/droit/esc.html>

- y M. VEYRAT (2000): "Reflexiones y paradojas teóricas en torno a la unicidad de la voz", *Libro de Actas del I Congreso de la Sociedad Española de Acústica Forense*, Madrid, Universidad Politécnica de Madrid, págs. 135-144.
- MUÑOZ, C. y M. VEYRAT (2003): "Lo esencial y lo accesorio en la composición individual de la voz. Propuesta para su identificación acústica", *Actas del II Congreso de la Sociedad Española de Acústica Forense*, Barcelona, Centre de Tecnologies i Aplicacions del Llenguatge i la Parla, Universitat Politècnica de Catalunya, págs. 71-78.
- HERNANDO, J. (2001): "Reconocimiento automático de locutores", *Ciencia & Tecnología del siglo XXI*, Barcelona, Tibidabo Edicions, vol. 2, págs. 23-26.
- JAKOBSON, R., G. FANT y M. HALLE (1955): *Preliminaries to speech analysis. The distinctive features and their correlates*, Cambridge, Mass., MIT.
- QUILIS, A. (1999): "El reconocimiento de la voz en la investigación judicial. La experiencia del lingüista", *Lengua y discurso. Estudios dedicados al Profesor Vidal Lamíquiz*, Madrid, Arco Libros.
- ROSENBERG, A. (1976): "Automatic speaker verification: A review", *Proceedings of the IEEE*, vol. 64, págs. 475-487.
- ROUSSELOT, P. (1891): *Les modifications phonétiques du langage étudiées dans les patois d'une famille de Cellefrouin (Charentes)*, Paris, H. Didier.
- SOSA, J.M. (1999): *La entonación del español. Su estructura fónica, variabilidad y dialectología*, Madrid, Cátedra.