

La identificación de la voz humana

ÁNGEL RODRÍGUEZ BRAVO

Departamento de Comunicación Audiovisual y Publicidad
de la Universidad Autónoma de Barcelona

87

1. INTRODUCCIÓN

La gran popularidad de todo tipo de tecnologías del audio ha facilitado enormemente la comunicación mediante la voz a través del teléfono convencional, el radioteléfono, los contestadores automáticos, los casetes, los ordenadores personales, los minidiscos, etc. Como consecuencia directa de esto, hoy es relativamente fácil conseguir registros electromagnéticos de conversaciones entre personas relacionadas con delitos graves. Entonces, surgen frecuentemente situaciones policiales como la siguiente: «Se ha conseguido una grabación determinada de una voz y se ha comprobado posteriormente, sin ninguna duda, que su fuente fue una persona que cometió el delito que se está investigando».

El problema que se plantea, pues, es la existencia de un registro en cinta magnética de la voz del delincuente (voz no dubitativa), que se podría transformar en una prueba concluyente de culpabilidad o de inocencia de las personas presuntamente implicadas en el caso. No obstante, para que esta grabación se pueda utilizar como prueba judicial hay que demostrar sin riesgo de error que alguna de las voces de los sospechosos (voces dubitativas) coincide exactamente con la de la grabación.

A menudo descubrimos en nuestra vida cotidiana que somos capaces de reconocer a alguien solamente si le oímos hablar. Así, la experiencia auditiva humana parece indicar que, ciertamente, un locutor se podría identificar a partir del sonido de su voz si se dispone de unos instrumentos técnicos determinados y de los conocimientos científicos adecuados.

Esta intuición, ¿es acertada? ¿Disponemos realmente de los instrumentos y los conocimientos que permitirían la identificación de un sospechoso a través del estudio de su voz? Y si esta respuesta fuese afirmativa, ¿cuáles son estos instrumentos y en qué tipo de conocimientos científicos se deben apoyar para ser utilizados?

El objetivo de este artículo será dar algunas respuestas a estas preguntas que, en realidad, constituyen un territorio común a la mayoría de los problemas científicos que tienen como objeto de estudio la voz humana.

2. LOS INSTRUMENTOS DISPONIBLES

El desarrollo vertiginoso de la informática durante los últimos veinte años ha dado un impulso muy fuerte a la acústica aplicada. Como consecuencia, las disciplinas como la fonología, la psicoacústica, la fonética experimental o la lingüística computacional han avanzado mucho. Ahora disponemos en cualquier laboratorio de análisis y tratamiento del habla de instrumentos muy potentes que posibilitan el desarrollo de un tipo de trabajos que antes eran impensables. Ciertamente, esto ha hecho avanzar mucho los conocimientos sobre la voz.

No obstante, la potencia de los instrumentos de análisis acústico ha avanzado, en realidad, mucho más rápidamente que el conocimiento básico sobre los objetos de estudio que es posible analizar. Lo que quiero decir es que el desarrollo de la tecnología de análisis comporta por sí mismo el conocimiento de aquello que es analizable. Hoy cualquier sistema de análisis acústico es capaz de lanzar miles de datos fiables a partir de unos pocos segundos de grabación de una voz humana; pero este cúmulo de información no es en absoluto directamente interpretable. Para el estudio de la voz, la mayor parte de esta información todavía es críptica o, en el mejor de los casos, sumamente confusa. Y su esfuerzo se centra todavía en buscar modelos que le permitan procesarla y utilizarla adecuadamente.

Es cierto que actualmente se comercializan en el mercado numerosos sistemas de análisis acústico y que son de fácil acceso; podemos citar algunos que ya son clásicos y de alto nivel, como el reciente Pulse de Brüel & Kjaer, el CSL de Kay Elemetrics, o el Mac Speech Lab de Apple. También hay sistemas más especializados en parámetros concretos, como el Signaix desarrollado en el Laboratoire de Langage et Parole de la Universidad de Aix-en-Provence. En el mercado hay muchos otros, siempre con la misma base científica de trabajo, pero con unas prestaciones y unos precios muy diferentes. Y también es cierto que cualquiera de estos instrumentos se puede utilizar para resolver problemas de identificación de la voz humana. Pero debemos tener en cuenta que todos son equipos de aplicación general al análisis acústico en sus más diversas facetas.

En varias universidades se está trabajando directamente con el desarrollo de sistemas automáticos para el reconocimiento del locutor; en el caso de España podríamos citar las investigaciones que se están haciendo en el Departamento de Matemática Aplicada de la Universidad de Murcia, en el Departamento de Teoría de la Señal y Comunicaciones de la Universidad Politécnica de Cataluña, o en el Departamento de Ingeniería Audiovisual y Comunicaciones de la Universidad Politécnica de Madrid. No obstante, los resultados que se han obtenido hasta ahora son muy parciales. En realidad no existen todavía sistemas comerciales orientados a resolver un problema tan concreto como la identificación del locutor, que proporcionen resultados de identificación satisfactorios con cualquier tipo de

voz. Podemos encontrar sistemas mejor o peor adaptados a la medida de unos u otros parámetros, o relativamente especializados para el análisis del ruido industrial o de la voz; pero los que son realmente fiables son siempre instrumentos de análisis genérico que requieren amplios conocimientos tanto de acústica como del fenómeno concreto que se está analizando para que puedan ser explotados adecuadamente.

3. LOS CONOCIMIENTOS NECESARIOS

Como he dicho antes, aunque las investigaciones sobre la voz humana han tenido un impulso muy grande en las últimas décadas, el habla se ha ido revelando progresivamente como un objeto de estudio con una complejidad extraordinaria y que sólo está resuelto muy parcialmente.

En estos momentos, la investigación aplicada al habla se orienta mayoritariamente a la síntesis y el reconocimiento automático, buscando el inmenso campo de aplicación que se prevé para este tipo de sistemas en la telefonía, la traducción automática y la comunicación hombre-máquina (máquinas que escriban al dictado, máquinas que hablen, máquinas que entiendan el habla, etc.).

Esta orientación tiene, a mi entender, un problema importante desde el punto de vista de la investigación policial. Para este tipo de aplicaciones es mucho más interesante el estudio de los factores que son comunes a todas las voces, y no investigar y buscar las características que son específicas y únicas para cada locutor concreto. Es decir, la perspectiva de investigación sobre lingüística aplicada que es actualmente mayoritaria avanza justo en el sentido contrario al que necesitaría la identificación de locutores. Además, veo que éste no es un problema coyuntural que dependa de una moda pasajera. La lingüística misma es, en su sentido más profundo, una disciplina orientada a la localización y la formalización de las estructuras sonoras comunes que comparten los hablantes para comunicarse.

Así, ni la lingüística ni sus disciplinas derivadas —fonética experimental, fonología, sociolingüística, etc.— tienden a observar y sistematizar los rasgos acústicos que caracterizan la individualidad del locutor (que son los que necesita la investigación policial), sino que les ocupa precisamente lo contrario. Es decir, lo que es intercambiable, lo que comparten todos los hablantes.

Es indiscutible que muchos conocimientos desarrollados desde la fonética experimental son directamente aplicables al problema de la identificación del locutor. Un ejemplo muy claro sobre esta cuestión es toda la investigación llevada a cabo en torno al análisis del espectro medio a largo plazo o SMLT (cfr. HARMEGNIES, 1992). Evidentemente, también es aplicable a la identificación el conocimiento sobre ciertas características concretas de la composición espectral de los fonemas, o toda la información ya disponible sobre rasgos acústicos sociolingüísticos. No obstante, como veremos detalladamente un poco más adelante, para la acústica forense este tipo de conocimientos genéricos todavía es insuficiente.

4. TÉCNICAS APLICABLES A LA IDENTIFICACIÓN DEL LOCUTOR

Ciertamente, existen una serie de procedimientos de análisis acústico que son aplicables como punto de apoyo científico para decidir si una voz dubitativa determinada se emitió o no por un locutor concreto. A continuación, revisaremos a grandes rasgos las corrientes de estudio que proporcionan instrumentos útiles para la identificación a través de la voz.

Podríamos establecer tres grandes líneas de trabajo al entorno de la voz que ya han sido utilizadas con bastante regularidad y que aportan soluciones útiles para la identificación del locutor:

- el análisis sociolingüístico
- el análisis manual de zonas puntuales del espectro
- el análisis automatizado global del espectro

4.1 El análisis sociolingüístico

Este procedimiento se basa en la localización y la sistematización de los rasgos acústicos y articulatorios que están asociados a la forma de hablar de los diferentes grupos a los que pertenece el individuo, como el medio geográfico, el origen social, el grado de cultura, la profesión, etc. Aparte de los rasgos estandarizados que se configuran en cada lengua, en cada zona geográfica, en unas profesiones determinadas (por ejemplo, en el *singing formant* o formante del canto), en los diferentes dialectos, etc., aparecen unas características sonoras específicas que sólo son localizables en este grupo concreto. Estos caracteres acústicos sociolingüísticos son observables en las formas entonativas, en el espectro de unas vocales determinadas, en formas concretas de articular algunas consonantes, etc.

Obviamente, el uso adecuado de este tipo de información por parte de un experto permite observar técnicamente los rasgos sociolingüísticos de la voz no dubitativa y determinar si éstos aparecen en las voces dubitativas. Es evidente que el procedimiento es muy útil para descartar las voces en que no están presentes los rasgos sonoros que se están buscando. Pero no hay suficientes garantías únicamente a partir de este tipo de análisis para afirmar inequívocamente que el propietario de la voz en el que hemos localizado las coincidencias es quien emitió la voz no dubitativa. El propio carácter de la metodología con que se trabaja parte de la premisa que estos rasgos acústicos no son específicos de un solo individuo, sino que son característicos de un grupo determinado. Por tanto, el resultado de este tipo de análisis no nos identifica de una forma específica al sujeto, porque sabemos positivamente que hay muchos otros con esas mismas características en su voz.

Así, el procedimiento sólo es útil, en realidad, como apoyo complementario de la globalidad de la información sobre el caso.

4.2 El análisis manual de zonas puntuales del espectro

Este procedimiento consiste en estudiar detalladamente unas partes muy concretas de un fragmento de la voz no dubitativa que sabemos que aportan información específica del sujeto. Es el caso, por ejemplo, de la «distancia» entre los dos primeros formantes dentro de los márgenes de variabilidad de cada una de las vocales. El dato concreto que aporta información significativa es el que se obtiene cuando se resta la frecuencia del primer formante a la del segundo, en sonidos vocálicos concretos, es decir, utilizando el índice ($F_2 - F_1$). De acuerdo con la información de la que dispongo en estos momentos, este tipo de datos se debe estudiar comparando siempre entre la misma vocal de las mismas palabras. Esta información refleja las posiciones internas de la boca específicas del locutor.

El procedimiento de la distancia entre formantes ha demostrado su utilidad para identificar a actores que estaban construyendo personajes diferentes, es decir, que enmascaraban y manipulaban su voz voluntariamente. Dicho de otra manera, los resultados del análisis estadístico de los datos que se obtienen cuando se calcula el índice ($F_2 - F_1$) vocal por vocal muestran que al contrastar el resultado del análisis acústico de locutores diferentes, aparecían diferencias significativas. Contrariamente, cuando la contrastación se hacía entre las diferentes interpretaciones de un mismo locutor, el índice ($F_2 - F_1$) no reflejaba diferencias significativas (cfr. RODRÍGUEZ, 1989). Este índice también lo han utilizado en Italia otros investigadores y han obtenido resultados muy parecidos.

De cualquier manera, hay que tener en cuenta que el índice ($F_2 - F_1$) no fue eficaz en el 100 % de los casos. Se estudiaron ocho interpretaciones de cuatro actores diferentes (dos interpretaciones por actor) y el procedimiento se mostró eficaz para identificar al locutor en todas las combinaciones comparativas posibles, menos la que contrastaba las dos interpretaciones de una de las locutoras. En este caso, la diferencia entre las dos versiones fue lo suficientemente buena como para engañar al método. En resumen, el margen de acierto del índice era del 96,4 %. Ciertamente muy elevado si tenemos en cuenta el tipo de prueba a la que se había sometido.

Es perfectamente posible encontrar otros índices de este tipo y utilizarlos combinados, lo que reforzaría las garantías de los resultados. De todas formas, considero que este tipo de trabajo requiere líneas de investigación básica que sean específicas, es decir, le hacen falta investigaciones que tengan la identificación del locutor como objetivo concreto, en lugar de recoger únicamente los resultados complementarios de otros proyectos que trabajan con objetivos esencialmente lingüísticos.

4.3 El análisis automatizado global del espectro

Éste es, sin duda, el procedimiento que tiene más adeptos entre los ingenieros. Se basa en obtener de forma automática una serie de análisis espectrales de los que se harán promedios. Se llevan a cabo, por ejemplo, unos cien análisis por segundo de la composición de frecuencias y se obtienen aproximadamente cuatrocientos valores frecuenciales en cada espectro (estas cifras dependen del método de análisis que se aplique). El análisis se hace durante unos cuantos segundos

de grabación y, posteriormente, se suman y se hacen los promedios de todos los análisis para obtener, finalmente, un espectro medio. La similitud o disimilitud entre los respectivos espectros medios de dos voces diferentes es el índice que se utiliza para estudiar si estas voces pertenecen o no al mismo locutor. Para estudiar cuantitativamente el grado de similitud a partir de los espectros medios, se suele utilizar la distancia euclidiana. Por otro lado, hay investigadores que han desarrollado sus propios índices para objetivar la comparación.

El espectro medio a largo plazo se ha revelado, efectivamente, como un índice muy útil para estudiar aspectos de la voz no dependientes de la lengua. Este procedimiento se ha aplicado para estudiar la influencia en el habla de los estados emocionales, de los rasgos de personalidad, del sexo y la edad, de problemas fisiológicos como la sordera, de las patologías de la voz, etc., y se obtienen buenos resultados (cfr. HARMÉGNIES, 1992). Naturalmente, también se ha utilizado para estudiar la individualidad vocal y se han conseguido unas tasas de acierto en la identificación del locutor entre el 90 % y el 95 % (cfr. FURUI, 1978; TOSI, 1977 y 1979).

La base teórica de este tipo de aproximación al problema es la convicción de que cada locutor tiene unas dimensiones específicas e invariables del tracto vocal que se reflejan constantemente en el espectro acústico de su voz. Si esto fuese cierto, sería suficiente encontrar el método adecuado de análisis espectral y aplicarlo en un fragmento suficientemente largo de la voz, para poner de relieve en el espectro medio estos caracteres sonoros específicos del individuo. De hecho, la ingeniería del tratamiento de señal persigue desde hace mucho tiempo desarrollar un sistema automático basado en este tipo de procedimiento, que permita comparar dos voces para tener como respuesta inequívoca si pertenecen o no al mismo locutor. No obstante, según la información que tengo de esto, ninguno de los sistemas desarrollados hasta ahora ha conseguido rebajar las tasas de error, que ya hace una década que siguen manteniéndose sistemáticamente entre el 5 % y el 10 %.

Desde mi punto de vista, este procedimiento arrastra un error fundamental en su base conceptual: ignora que todo el tracto vocal es globalmente blando, elástico y enormemente dinámico, y que puede ser alterado de una forma importante en todas sus zonas por factores psicológicos, emocionales y de estrés, y, además, por la propia voluntad del locutor si éste es lo bastante hábil. Es decir, el análisis automatizado global del espectro ignora que la configuración del tracto vocal puede resultar muy modificada para un mismo individuo con una facilidad relativa.

5. EL PROBLEMA POLICIAL ANTE EL PROBLEMA CIENTÍFICO

Después de haber revisado las tres grandes corrientes científicas de análisis del habla aplicables a la identificación, creo que es importante poner de relieve las diferencias que hay entre la concepción científica y la concepción policial del problema de la identificación del locutor a partir de la voz.

Desde el punto de vista del método científico, un procedimiento de análisis que dé un grado de acierto en la identificación del locutor de alrededor del 95 % se suele considerar muy satisfactorio, porque supone una demostración clara de que las

hipótesis de las que partía la investigación han sido contrastadas y, por tanto, no ha habido errores fundamentales. Por ejemplo, si lo que se estudia es el índice ($F_2 - F_1$), un resultado de este tipo significa que, efectivamente, este parámetro proporciona la información que buscamos aunque haya algunos factores no controlados que nos distorsionan el resultado con un error del 5 %. El estudio se considera, pues, finalizado y las conclusiones aceptan como científicamente válido el índice localizado.

En cambio, desde el punto de vista de la aplicación policial, las condiciones en que se encuentra este índice todavía son bastante insatisfactorias. Los resultados obtenidos suponen que de cada cien procesos judiciales en que se aplique esta prueba, cinco personas podrían ser acusadas y condenadas injustamente. Un margen de error altísimo para basarse en él exclusivamente. A esto hay que añadir que mientras que la calidad acústica de las muestras con que se trabaja en un laboratorio científico suele ser excelente (voces sin ruidos y con una alta gama dinámica y espectral), las muestras que se pueden conseguir en las investigaciones policiales suelen ser de muy baja calidad, grabadas en malas condiciones y mezcladas con muchos ruidos. Este cambio radical en las condiciones de las muestras analizadas aumenta con facilidad el margen de error en la identificación a un 15 % o un 20 %, lo que convierte el índice que, en principio, era científicamente válido en totalmente inaceptable desde el punto de vista pericial.

No obstante, las malas condiciones de la muestra en cuanto a ruidos es un problema clásico de la ingeniería acústica; hay varias líneas de investigación que trabajan en sistemas especializados en mejorar la identificación de la voz a partir del reconocimiento y la sustracción de los ruidos que puedan estar mezclados (cfr. ORTEGA-GARCÍA y GONZÁLEZ-RODRÍGUEZ, 1997).

De todas formas, las pruebas periciales que se utilizan en los casos en que se recurre a la identificación a través de la voz siguen basándose en el juicio de expertos porque todavía no hay ningún protocolo sistemático de análisis acústico que se pueda aceptar de forma universal. El problema surge, entonces, cuando el dictamen de un perito se contrapone al de otro que —a partir de la misma prueba— afirma justamente lo contrario.

En resumen, la aplicación policial de los procedimientos de análisis acústico a la identificación, por el hecho de que tiene unos objetivos muy concretos y específicos, exige unas garantías de fiabilidad que están muy por encima de las que se aplican habitualmente en los proyectos de investigación científica básica. Evidentemente, esto supone que si se quiere conseguir el tipo de resultados que requiere la investigación policial sea imprescindible desarrollar proyectos de investigación con sus objetivos explícitamente orientados a las necesidades de la acústica forense.

6. LA VARIABILIDAD DEL HABLA

Antes de exponer las conclusiones a las que me conducen los planteamientos de este trabajo, me parece que hay que hacer una última reflexión global sobre los problemas objetivos con que se enfrenta la búsqueda de métodos de análisis para conseguir la identificación del locutor.

Si queremos obtener unos resultados realmente aplicables a la identificación, me parece que es imprescindible que tanto los servicios policiales como los investigadores científicos y técnicos tomen conciencia de que nos enfrentamos a dos problemas de mucha envergadura:

- a) la propia tendencia natural de la investigación con base lingüística al estudio de los rasgos acústicos compartidos y comunes, en lugar de los rasgos peculiares del individuo;
- b) la dificultad objetiva que implica la extraordinaria variabilidad acústica del habla.

94

Ya he hablado más arriba del primer problema, por esto en este apartado nos dedicaremos al segundo, es decir, a la variabilidad.

La voz humana actúa configurando formas acústicas desde un doble sistema de formas sonoras que evolucionan paralelamente en el tiempo, y que articulan dos flujos de información complementarios: por un lado, el arbitrario y normalizado del sistema lingüístico y, por otro, el del sistema de la expresión fonostésica, compuesto por índices acústicos relacionados, por ejemplo, con el sexo, la edad, la configuración física, la personalidad y el estado emocional del locutor, o sobre sus mismos movimientos y todo lo que el hablante describe oralmente (cfr. RODRÍGUEZ, 1989).

Así, por ejemplo, imaginémonos que hemos conseguido grabar la voz de un delincuente interviniendo una llamada telefónica. Al hablar, ciertamente, se construirán formas sonoras basadas en la estructura fonemática y léxica de la lengua. Y estas formas fonemáticas estarán organizadas globalmente con unas variaciones tonales y de intensidad sistematizadas, también, por los códigos lingüísticos. No obstante, toda estructura lingüística resulta siempre alterada, matizada e, incluso, muy modificada por una serie de condiciones previas que no dependen nada de la misma lengua.

Ahora supongamos que ese delincuente cuando hacía esta llamada, por alguna razón que desconocemos, estaba furioso. Esto significa que durante aquella locución habrá quedado modificada su manera habitual de hablar.

La agresividad hace que el tono y la intensidad de la voz sean mucho más altos de lo normal, la tensión articularia también aumenta bastante más que de costumbre y la duración de las pausas se acorta. Además, la agresividad modifica el tracto vocal y aumenta la tensión muscular, lo que influye sobre el espectro acústico, potencia las frecuencias altas y reduce su armonicidad. Como resultado, las variaciones acústicas que dependen del estado emocional del locutor (expresión fonostésica) alteran todos los rasgos sonoros lingüísticos y se produce una modificación en la voz nada previsible desde una perspectiva estrictamente fonética.

Ahora supongamos que una grabación de este tipo es nuestra voz de referencia, es decir, la voz no dubitativa que hemos de comparar con las voces de los sospechosos. Aunque uno de estos sospechosos fuese realmente el criminal, es muy difícil que durante las nuevas grabaciones de su voz éste se halle en unas condiciones emocionales parecidas a las que tenía cuando hizo la llamada que

estamos utilizando como prueba. Por tanto, por el hecho de que una voz emocionada puede resultar muy alterada en todas sus dimensiones sonoras, nos enfrentaríamos a un problema de variabilidad acústica que dificultaría muchísimo la identificación.

En suma, la acústica del habla se configura con dos sistemas paralelos, complementarios y perfectamente articulados. En primer lugar, el hablante selecciona su tono medio o modulador, la intensidad media de su voz y la colocación global de su tracto vocal en función de la situación comunicativa en la que se encuentra. Y a partir de aquí empieza una labor paralela de construcción de formas sonoras según el sistema de la lengua que, a medida que se van construyendo, se van desviando acústicamente por el sistema de la expresión fonoestésica en uno u otro sentido, de acuerdo con el estado emocional del hablante, de su voluntad expresiva, de la actitud del receptor, etc.

Hay que añadir, además, que todos estos rasgos acústicos que, en principio, están determinados por la situación contextual, son también susceptibles de ser alterados por voluntad del propio locutor si éste está entrenado adecuadamente. Así, hay que tener en cuenta que todo hablante está capacitado fisiológicamente para alterar su voz, variando a voluntad su tono medio, su grado de armonicidad, sus inflexiones tonales, su tensión articulatoria, etc., y así se mostraría como más o menos agresivo, joven, viejo, alegre, triste, afectuoso, aniñado, femenino...

Esta gran variabilidad hace que el sonido de la voz, al ser recogido por un micrófono ya fuera de la boca, sea una masa acústica compacta e interdependiente, que configura un tipo de señal en la que es extraordinariamente difícil encontrar alguna característica que no haya sido modificada varias veces durante su recorrido del tracto vocal. Como consecuencia, no podemos ignorar que cualquier característica de la señal, que suponemos dependiente únicamente de un locutor, siempre puede llegar a ser alterada por alguna de las muchas posibilidades de modificación del aparato fonador.

7. CONCLUSIONES

En primer lugar, se concluye que, si bien es perfectamente posible basarse en instrumentos de análisis y en conocimientos de acústica y de lingüística para la identificación a través de la voz, todavía no existen protocolos de análisis ni sistemas automáticos que puedan hacer una identificación con suficiente confianza y fiabilidad en todos los casos.

En segundo lugar, es importante indicar que la gran variabilidad acústica del habla no supone que la identificación a través de la voz no sea posible, sino que exige una nueva perspectiva de trabajo.

Aunque anteriormente afirmaba que no hay parámetros estables en la voz humana, mi experiencia como educador en el campo de la locución también me ha enseñado que, en realidad, no es nada fácil adquirir la capacidad de cambiar la calidad acústica de la voz y el estilo de la locución. Aunque se consiga un nivel de dominio importante del aparato fonador, siempre olvidamos alterar ciertos rasgos

y éstos permanecen constantes la mayoría de nuestras locuciones. El problema es que estos rasgos acústicos estables varían de un locutor a otro. Mientras que en un locutor lo que permanece invariable es una cierta inflexión tonal característica, en otro es la posición un poco elevada del dorso de la lengua (palatalización), y en un tercero, la duración máxima de las vocales.

Es decir, según mi opinión, el problema de la identificación de la voz se debe tratar mucho más como un problema de identificación de *estilo*, o sea, como un problema más próximo al de saber si una pintura ha sido o no pintada por un artista determinado, que no al de asignarle un propietario a una huella digital o a una secuencia de ADN. El sonido de la voz no es el fragmento físico de un individuo producido por su programación genética y sobre el que él no tiene ningún control racional, sino que es el simple resultado de un trabajo expresivo sobre las formas sonoras del habla, en el que el sujeto emisor influye conscientemente. Y hay que tratarlo como tal.

Personalmente, pienso que efectivamente sí que es posible resolver el problema de la identificación analizando acústicamente la voz. No obstante, estoy convencido de que para hacerlo se debe tener en cuenta la multidimensionalidad de la materia fónica y trabajar con un modelo considerablemente complejo. Este modelo debería tener en cuenta conocimientos de fonética, sociolingüística, psicoacústica, acústica de las emociones, etc., y llevar a cabo análisis tanto de parámetros sonoros segmentales como suprasegmentales, así como contemplar dimensiones como la entonación, los ataques y las caídas. Es decir, el modelo debería encontrar un procedimiento de análisis estadístico de n variables acústicas, que hayan sido seleccionadas en varias investigaciones previas. El modelo desembocaría, finalmente, en un *protocolo de análisis acústico-estadístico*, que se tendría que poner a prueba analizando grabaciones de locutores expertos que hayan modificado su voz y analizando voces en las que se haya destruido técnicamente una parte de su información sonora.

En esta línea, actualmente está en marcha un proyecto interuniversitario en el que colabora también el Servicio de Policía Judicial de la Guardia Civil; el objetivo es la puesta a punto de un método para verificar sistemas automáticos de identificación por la voz (cfr. ORTEGA-GARCÍA y otros, 1998). La base del proyecto es construir un corpus muy amplio con voces no dubitativas, que contenga y sistematice muchas variaciones del habla y de la calidad de grabación en cada locutor.

En suma, considero que la solución estará en el diseño de un protocolo de análisis múltiple con un número sistemático de controles acústicos suficientemente amplio y diversificado como para que no pudiese ser «engañado» por manipulaciones técnicas ni por la capacidad expresiva de algún locutor. Lógicamente, este protocolo exigiría previamente unas condiciones muy concretas de calidad acústica para que las grabaciones pudiesen ser sometidas a la prueba.

Finalmente, pienso que las instituciones interesadas en aplicar el análisis de la voz humana a la investigación de delitos deben ser conscientes de que, mayoritariamente, las líneas actuales de investigación no avanzan en absoluto en el camino que requiere la identificación.

Así como a la lingüística y a la ingeniería les suele resultar mucho más intere-

sante y rentable el estudio de los factores que son comunes a todas las voces que investigar y formalizar las características acústicas de la individualidad, el conocimiento y la tecnología tienden a avanzar justamente en el sentido contrario al que necesita la identificación de locutores. De hecho, cuando se consiguen conocimientos o instrumentos aplicables a la identificación por la voz, éstos suelen resultar insatisfactorios para la investigación policial precisamente porque sólo son conclusiones complementarias. Es decir, porque son resultados que, en realidad, provienen de proyectos con objetivos claramente orientados a lo que hoy se llama *las industrias de la lengua*.

Como investigador y como conocedor del rendimiento real del método científico, sé que sólo es posible encontrar la solución de un problema de conocimiento cuando los objetivos de la investigación definen con total claridad y exactitud qué es lo que se busca. Y sé también que la exploración científica sólo obtiene algo que no está buscando cuando de manera fortuita e inesperada interviene el azar.

Considero, en consecuencia, que si se quiere conseguir un avance real en la investigación de delitos aplicando el análisis de la voz, es imprescindible que las instituciones interesadas presten apoyo, estimulen y proporcionen los recursos económicos necesarios para financiar proyectos que tengan la identificación del locutor como un objetivo único, específico y concreto.

BIBLIOGRAFÍA

- FURUI, S. «Effects of Long-Term spectral Variability on Speaker recognition». *J.A.S.A.*, 64, suppl. 1, papel NN28, 183, 1978.
- HARMEGNIES, B. «Les sources de variation du spectre a long terme de parole: revue de la literature». *Acoustique Canadienne*, 20 (2), pág. 9-35, 1992.
- ORTEGA-GARCÍA, J.; MARRERO-AGUIAR, V.; GONZÁLEZ-RODRÍGUEZ, J.; DÍAZ-GÓMEZ, J.; GARCÍA-GIMÉNEZ, R.; LUCENA-MOLINA, J.; SÁNCHEZ MOLERO, J.A.G^a: «Speaker Recognition Oriented 'AUMADA' Large Speech Corpus». En: First International Conference on Language Resources and Evaluation, Granada, mayo de 1998.
- ORTEGA-GARCÍA, J. Y GONZÁLEZ-RODRÍGUEZ, J. «Providing Single and Multi-channel Acoustical Robustness to Speaker Identification Systems». IEE International Conference on Acoustics Speech and Signal Processing. ISASSP-97, vol. II, pág. 107-110, Munich, abril de 1977.
- RODRÍGUEZ, A. *La construcción de una voz radiofónica*. Tesis doctoral. Departamento de Comunicación Audiovisual y Publicidad. Universidad Autónoma de Barcelona, 1989.
- TOSI, O.; PISANI, R.; DUBES, R.; JAIM, A. «An Objective Method of Voice Identification». En: International Phonetic Sciences Congress, Miami, 1977.
- TOSI, O. *Voice identification: Theory and legal Applications*. Baltimore, University Park Press, 1979.