

## **Desarrollo y validación de un protocolo para la identificación perceptiva forense en ruedas de reconocimiento de voz**

### *Development and validation of a protocol for forensic perceptual identification on speech recognition wheels*

#### **Autores:**

*Ángel Rodríguez Bravo; Patricia Lázaro Pernias; Norminanda Montoya Vilar; Josep M<sup>a</sup> Blanco Pont; Dolors Bernadas Suñé; Daniel Tena Parera; Ludovico Longhi; Josep Manel Oliver Comes.*

***LAICOM** (Laboratorio de Análisis Instrumental de la Comunicación)  
Departamento de Comunicación Audiovisual y Publicidad II,  
Universidad Autónoma de Barcelona), 2003.*

**RESUMEN:** En este informe científico se exponen **completos** los resultados finales obtenidos en el proyecto de investigación “*IDENTIFICACIÓN PERCEPTIVA DE LOCUTORES PARA LA ACÚSTICA FORENSE*” financiado por el Ministerio de Ciencia y Tecnología (DGCYT PB98-0862). Estos resultados fueron publicados parcialmente en Actas del II Congreso de la Sociedad Española de Acústica Forense<sup>1</sup>

El objetivo de esta investigación es experimentar y validar un protocolo sistemático para realizar tests de discriminación auditiva en procedimientos de Ruedas de Reconocimiento de Voz (RRV). La investigación se ha realizado siguiendo el método de análisis instrumental<sup>2</sup> y se ha desarrollado en 3 etapas: 1) búsqueda y formalización de las variables fundamentales que actúan en la memoria auditiva para el reconocimiento de una voz; 2) definición de las condiciones óptimas para las variables /duración del discurso/, /dependencia del texto/ y /estado emocional del hablante/ para estimular la memoria auditiva y 3) estudio de la influencia de un periodo de latencia y del estrés emocional en el reconocimiento de voces. Para el desarrollo del estudio se han aplicado diversos métodos de investigación cualitativa y dos procedimientos experimentales.

**ABSTRACT:** In this scientific report we expose the final results obtained in the research project “*PERCEPTIVE SPEAKER IDENTIFICATION FOR THE FORENSIC ACOUSTICS*” funded by the Science and Technology Department (DGCYT PB98-0862).

The aim of this investigation is to experiment and validate a systematic protocol to elaborate acoustics discrimination tests in a voice identity parade. The investigation has been carried out in 3 stages: 1) search and settle down of the fundamental variables that work in the acoustics memory to recognize a voice; 2) definition of the peak conditions for the variables/ length of the speech/ dependence on the text/ and/ emotional condition of the speaker/ to stimulate the acoustics memory and 3) Study of the influence of a latent period and of the emotional stress in the recognition of voices. To develop this study, sundries qualitative investigation methods have been applied and two experimental procedures.

## **1. OBJETIVOS DE LA INVESTIGACIÓN**

Hasta el momento sólo las Ruedas de Reconocimiento Visual han sido tratadas y desarrolladas en profundidad, y gozan de mayor reconocimiento y fiabilidad como pruebas en causas judiciales. Los principales problemas aún sin resolver a los que deben enfrentarse las Ruedas de Reconocimiento de Voz son, precisamente, la ausencia de un criterio claro para el diseño de estos procedimientos, la dificultad de extrapolar los resultados de la experimentación en laboratorio a las circunstancias reales y el desconocimiento del proceso perceptivo de la codificación de la información auditiva y su recuperación.

El estudio de cualquier fenómeno que implique la percepción y fijación en la memoria de ciertos rasgos sonoros de una voz y su posterior reconocimiento e identificación, debe contemplar la descripción de los procesos vinculados al almacenamiento de esos estímulos, es decir, averiguar las características de la memoria auditiva o ecoico-vocal. Mediante este tipo de investigación es posible optimizar el diseño de las ruedas de reconocimiento de voz de modo que tanto su preparación como su diseño final consigan el máximo estímulo de la memoria de los testigos auditivos y la optimización de la eficiencia del método.

El objetivo final de esta investigación es desarrollar y validar un protocolo sistemático y eficiente para realizar tests de discriminación auditiva en procedimientos de Ruedas de Reconocimiento de Voz (RRV). La consecución de este objetivo supondrá la culminación de un proceso de estudio en el que se están buscando nuevos conocimientos sobre los parámetros acústicos que explota el oído humano en el proceso perceptivo de reconocimiento de las formas sonoras que caracterizan una voz que ya ha sido escuchada con anterioridad, y sobre las estrategias perceptivas que permiten discriminar esa voz de entre otras emitidas por locutores diversos.

## **2. ALGUNOS ANTECEDENTES**

El trabajo previo desarrollado tanto de forma individual por algunos de los integrantes del equipo LAICOM, como las investigaciones del propio equipo, han consolidado un conocimiento importante de esos procesos de percepción auditiva ligados a la expresión sonora, que recoge planteamientos que van del análisis acústico a la psicoacústica y la psicología de la percepción, sin olvidar los conocimientos sobre lenguaje audiovisual [3]. La tarea del equipo en el desarrollo de esta investigación se ha centrado en la localización de las variables fundamentales que actúan en la memoria ecoico-vocal y, una vez localizadas, en buscar sus condiciones óptimas para estimular el recuerdo y el reconocimiento de las voces.

Coinciden los psicólogos en considerar “la memoria como un complicado sistema de procesamiento de la información que opera a través de procesos de almacenamiento, codificación, construcción, reconstrucción y recuperación de la información” [4]. Pero la interpretación de esos complicados procesos ya no es tan coincidente. Así, son varios los modelos propuestos para tratar de explicar cómo se desarrollan las tareas de recuperación y recuerdo de información. En todos ellos es necesario el conocimiento de los tipos de memoria según el plazo de permanencia de la información y se distingue entre memoria a largo plazo (MLP) y memoria a corto plazo (MCP), con relación al rendimiento de la memoria, determinado por la profundidad de procesamiento de la información [5]. De entre los modelos de recuperación cabe citar el de Memoria Asociativa: “el proceso de recuperación comienza cuando un conjunto de claves se coloca en la MCP y activan un conjunto de información almacenada en la MLP” [6]; y la Recuperación mediante Esquemas, según el cual los contenidos que tienen más probabilidades de recordarse son aquellos relacionados con los

esquemas utilizados por el sujeto durante la recuperación [7]. Estos modelos teóricos son los que han orientado el trabajo prospectivo del equipo.

### ***3. PRIMERA ETAPA: ESTUDIO CUALITATIVO DE LAS VARIABLES QUE ACTÚAN EN LA MEMORIA ECOICO-VOCAL***

Aún teniendo en cuenta los modelos que explican los procesos de recuperación y recuerdo, en el contexto concreto de la memoria ecoico-vocal, el problema fundamental es la falta de conocimientos específicos sobre las variables que intervienen y actúan en la memoria. Cuando una persona es víctima (o incluso testigo) de un delito, ya sea robo, violación, amenaza telefónica, etc., y no ve a su agresor, sino que solo le oye hablar, toda la información relacionada con éste que puede llegar a percibir y que luego podrá recordar se relaciona con los rasgos sonoros de su voz, sus palabras, su tensión emocional. El conocimiento de qué rasgos sonoros se fijan en la memoria y por qué esos y no otros, además de la localización y formalización de otras variables fundamentales relacionadas con la influencia de la situación delictiva, el contexto sonoro y temporal e incluso las propias capacidades perceptivas de la víctima devienen esenciales para la comprensión de los mecanismos de la memoria ecoico-vocal.

La Metodología<sup>8</sup> de la primera fase de trabajo de la investigación que presentamos ha consistido en el desarrollo de una tarea prospectiva para facilitar esos conocimientos previos sobre los rasgos sonoros de una voz que se fijan en la memoria de un oyente y que después le permiten reconocerla e identificarla entre otras. Una segunda finalidad de esta parte del estudio ha sido localizar otras variables relacionadas con el contexto de audición que pueden influir en estos ejercicios de memorización y recuerdo. Sólo a partir de un conocimiento preciso de qué variables devienen esenciales en los procesos de reactivación de la memoria estaremos en condiciones de iniciar el trabajo de diseño de un protocolo válido para las RRV. Los instrumentos metodológicos en los que se ha apoyado esta prospección han sido de tipo cualitativo, en concreto entrevistas en profundidad a víctimas de delitos que solo han oído y no han visto a su agresor, y grupos de discusión entre oyentes sometidos a la audición de una voz sin identificar.

#### ***3.1 Entrevistas de exploración cognitiva de la memoria ecoico-vocal***

En psicología forense se ha desarrollado una técnica de entrevista que se fundamenta en modelos cognitivos de memoria y que se basa en la estimulación del recuerdo a partir de distintos procedimientos. Este tipo de entrevista es conocido como Entrevista Cognitiva [9]. Las tácticas y técnicas de la entrevista cognitiva empleada en la psicología forense han sido recogidas y complementadas en la propuesta de entrevista de exploración cognitiva de la memoria ecoico-vocal desarrollada como primer instrumento de trabajo cualitativo de la investigación que presentamos. Esta entrevista en profundidad, cuyo objetivo principal es la localización, descripción y formalización de las variables que actúan en la memoria de una víctima de un delito, se desarrolló en tres grandes bloques conceptuales:

- la exploración de las capacidades del entrevistado, con el fin de conocer las capacidades audio-memorísticas de la víctima.
- la reconstrucción del registro sonoro que ha quedado en la memoria del entrevistado en relación con el delito, para estudiar el grado de precisión y los puntos débiles y fuertes de la memoria ecoico-vocal de la situación del delito.

- orientación del entrevistado hacia una valoración crítica de las RRV (en caso de haber sido sometido a alguna rueda).

A partir de estos tres grandes bloques se trabaja en detalle cuestiones tales como:

- valoración de la propia capacidad de recuerdo de voces, tanto en los contenidos como en las formas: capacidad de memoria global, capacidad de memoria auditiva, capacidad de descripción, capacidad de discriminación sonora. Esta autovaloración se somete a control a partir de la solicitud de recuerdo y reconstrucción de voces de famosos.
- influencia de la situación emocional en el recuerdo: se pide una reconstrucción de las circunstancias para generar una situación emocional próxima al hecho delictivo de manera que se promueva con mayor facilidad el recuerdo.
- influencia del contexto sonoro durante la agresión: recuerdo de los sonidos del contexto.
- influencias temporales: tiempo total del suceso, tiempo de recepción del estímulo (cuanto rato se oye hablar al agresor), tiempo pasado desde la agresión, tiempo transcurrido entre la agresión y la RRV.
- variables relacionadas con la voz del agresor: sus características físicas, sus características emocionales, sus rasgos sonoros, su forma de hablar, etc.

Las entrevistas de exploración de la memoria ecoico-vocal se realizaron a víctimas reales de agresiones (robo y abuso sexual) contactadas a través del Departamento de Acústica Forense de la Comisaría General de Policía Científica del Cuerpo Nacional de Policía. El protocolo seguido contempla la presencia de dos investigadores para cada una de las entrevistas realizadas. En el caso de la víctima por abusos sexuales se considera adecuado que el equipo desplazado esté formado por dos investigadoras, para facilitar la relación con la víctima.

Las principales conclusiones sobre las variables fundamentales que este método reveló que deben tenerse en cuenta en las etapas posteriores de la investigación, sobre todo en aquellas experimentales, son las siguientes:

- las víctimas muestran un alto grado de memoria, tanto general (el suceso, su contexto, etc.), como auditiva: son capaces de describir con precisión la voz de su agresor y sus palabras;
- a pesar de ello, los entrevistados se mostraron extremadamente desconfiados de su propia capacidad para el recuerdo de voces;
- utilizar como voces de control y de referencia común entre el entrevistador y el testigo auditivo voces de famosos (o de cualquier otro tipo de locutor) que sean conocidas por ambos se reveló como extraordinariamente útil y eficiente;
- la inducción al recuerdo, al pedir la narración de los hechos, favorece la memoria;
- las víctimas consideran esencial para la identificación de la voz del agresor, que en las Ruedas de Reconocimiento de Voz donde se reproduzca el contexto sonoro de la agresión y que las voces que integren la rueda repitan las mismas palabras y la misma actitud y tensión emocional del agresor. Sin embargo, la investigación experimental posterior muestra la dependencia del texto como una variable útil solamente en algunos casos;
- las víctimas insisten en la necesidad de reducir al máximo el intervalo temporal entre la agresión y la Rueda de Reconocimiento de Voz por miedo a la pérdida de su capacidad de evocación;
- los investigadores ven la necesidad de desarrollar un instrumento para la descripción de voces que facilite una tarea pormenorizada. La entrevista propone un instrumento incipiente que recoge características propias de las voces y una escala de valoración (por ejemplo: ¿Su voz era aguda o grave? Si 0 es muy grave y 10 muy aguda, ¿cómo la calificaría?).

### 3.2 Grupos de discusión

El segundo método de análisis cualitativo aplicado en la investigación son los grupos de discusión. Los grupos de discusión se plantean simultáneamente a la primera experimentación. Los motivos son claros: es preciso disponer de un cierto número de individuos que hayan oído una misma voz y que, tras un período de latencia, puedan recordarla y hablar sobre ella. La etapa experimental será explicada en detalle en el apartado siguiente, no obstante, avanzamos que los grupos de discusión se organizaron en cuatro sesiones y participaron un total de 27 sujetos, que dos semanas antes habían asistido a una audición de un fragmento de voz de unos 4'30", la llamada "voz a reconocer".

Los objetivos generales que se plantearon con la realización de los grupos de discusión fueron:

- encontrar un procedimiento general que permita establecer las características de una voz a partir de la información que guarde en su memoria un sujeto que la haya escuchado y afirme recordarla;
- explorar qué rasgos sonoros asociados a factores fisiológicos, sociológicos, lingüísticos, semánticos, etc. se retienen en la memoria, tanto comunes a un grupo de voces (estereotipos), como específicos y diferenciales para identificar una voz concreta;
- explorar qué términos y qué conceptos facilitan que personas sin conocimientos especializados puedan describir una voz.

Así mismo, y en relación con la etapa experimental, los grupos de discusión persiguen otros objetivos concretos, específicos, que son los siguientes:

- localizar los rasgos sonoros (fisiológicos, sociológicos, lingüísticos, etc.) de la voz oída durante la audición dos semanas antes, la llamada "voz a reconocer";
- localizar las frases y términos concretos usados por la "voz a reconocer" que podrían ser útiles para identificarla.

Los grupos de discusión son conducidos siguiendo un protocolo específico desarrollado a tal efecto. Cada grupo de discusión está formado por siete sujetos (a excepción de uno, formado por seis individuos) y está moderado por un investigador. Las fases más relevantes de las sesiones son: inicio con reconstrucción de los contenidos de la historia narrada por la "voz a reconocer" (inducción a la memoria), recuerdo de las frases dichas por la "voz", discusión sobre los términos de apoyo útiles para la descripción de cualquier voz y aplicación de esos términos a la "voz a reconocer".

Las principales conclusiones obtenidas a través de los grupos de discusión se refieren a los términos de apoyo para la descripción de los rasgos sonoros de una voz. De los términos empleados no presentan dificultades aquellos referidos a: rasgos físicos, fisiológicos y orgánicos; la intensidad o volumen de la voz; tono o tonalidad; ritmo; características socioculturales y acento; estilo personal del hablante. Presentan algunas dificultades los calificativos asociados al concepto de timbre y, sobre todo, el concepto de entonación. Otras conclusiones coinciden con las obtenidas en las entrevistas de exploración cognitiva: la necesidad de la reproducción de las características de la voz y su estilo, además de sus palabras para facilitar el reconocimiento.

Los resultados de los grupos de discusión confirman una intuición que el equipo LAICOM ya venía planteándose en investigaciones anteriores [<sup>10</sup>]: la necesidad de construir una biblioteca de voces, de estereotipos vocales, que faciliten la descripción de cualquier voz a través de sus rasgos más relevantes. En muchas ocasiones los oyentes no especializados quieren describir algún rasgo sonoro que recuerdan con claridad, pero su mayor dificultad radica en no poseer un repertorio de términos, un vocabulario claro, explícito y compartido.

#### **4. SEGUNDA ETAPA: ESTUDIO EXPERIMENTAL**

Recordemos, tal y como apuntábamos anteriormente, que el objetivo final de la investigación, es el diseño y puesta a punto de un instrumento que permita aplicar con fiabilidad tests de discriminación auditiva. Tras la primera etapa de investigación cualitativa, que nos ha permitido profundizar en aquellos aspectos de la memoria ecoico-vocal relevantes para un ejercicio de recuperación como son las Ruedas de Reconocimiento de Voz, se abre una segunda fase, esta vez empírica, orientada a la experimentación concreta sobre protocolos de RRV. La experimentación deberá partir de una reconstrucción dramática, es decir, una simulación de un hecho delictivo, de manera que se proporcione al equipo investigador *víctimas, testigos y delincuentes* a los que posteriormente aplicar procedimientos de test de discriminación auditiva.

La revisión bibliográfica sobre Ruedas de Reconocimiento de Voz ha puesto de manifiesto la existencia de ciertas variables, conocidas como *variables del sistema* (en este caso recogemos la terminología de las Ruedas de Reconocimiento Visuales) [11], sobre las que no existe un acuerdo claro de aplicación. Así, cuestiones tales como la duración de los fragmentos de las voces incluidas en las ruedas, las características de los discursos portadores, la selección de las voces falsas o de contraste, el número de voces de contraste, realización de test singular o test múltiple, etc. son aspectos que han sido utilizados [12] pero escasamente comprobados a nivel experimental y sobre los que los expertos recomiendan un trabajo de mayor profundidad [13].

Por este motivo el equipo decide que antes de pasar a la experimentación con dramatización es necesario abordar una línea experimental previa que trabaje esas variables del sistema, es decir, esas variables dependientes del procedimiento de las Ruedas que no dependen ni de variables relacionadas con la víctima (habilidades perceptivas, tensión emocional) ni de variables relacionadas con la situación de la agresión (contexto sonoro, duración del estímulo, etc.). Una vez se obtengan resultados validados será posible iniciar la etapa experimental final.

##### **4.1 Primer experimento: variables del sistema**

Después de revisar con detalle los resultados de la etapa de investigación cualitativa sobre las variables del sistema que intervienen en la preparación de una Rueda de Reconocimiento de Voz, se decidió estudiar de forma experimental tres de ellas:

- la duración de cada fragmento de voz que componen las cadenas de las RRV: existe un acuerdo tácito de empleo de fragmentos de unos 30 segundos, aunque no se ha estudiado experimentalmente la duración óptima de esos fragmentos para que se produzca reconocimiento, y al mismo tiempo el fragmento no resulte excesivo. Creemos necesario determinar un Tiempo Mínimo Necesario Óptimo (TMNO), es decir, qué tiempo resulta suficiente para que cualquier testigo pueda reconocer una voz con garantías. Se trabajarán tres duraciones: 15'' – 30'' – 45'';
- la reproducción o no de la actitud emocional del sospechoso durante el hecho delictivo en las cadenas de las RRV: aunque hay total acuerdo en considerar esencial la reproducción de las características sociolingüísticas del agresor y el contexto sonoro del delito para mejorar la inducción al reconocimiento, el acuerdo es inexistente con relación a la actitud emocional con la que actuó el delincuente. Creemos necesario determinar si la reproducción del estado emocional facilita o no el reconocimiento por parte de la víctima. Se trabajará con ruedas que contengan fragmentos de voz con tensión emocional y ruedas con fragmentos sin tensión emocional.

- características del discurso portador: tampoco hay acuerdo en cómo debe ser el texto portador de las cadenas de voces de las RRV. Algunas posturas apoyan una reproducción fiel de las palabras que la víctima recuerda haber oído pronunciar a su agresor, en cambio otras posturas abogan por un texto libre, de habla espontánea. Creemos necesario determinar qué forma del discurso portador resulta más adecuada para estimular el reconocimiento. Se trabajará con ruedas que contengan textos libres y ruedas que contengan frases repetidas.

Acotadas las variables a estudiar se diseña una propuesta de experimentación. En realidad se trata de un planteamiento simple: disponer de una voz para que sea escuchada por un número estadísticamente relevante de sujetos experimentales, que posteriormente la deberán reconocer de entre una cadena de 5 secuencias de voz de 5 locutores distintos, entre las cuales está la voz del locutor a reconocer, en un test de discriminación auditiva. Es importante advertir que, dado que estas variables a estudiar dependen, por una parte, del propio procedimiento de tests y, por otra parte, de las capacidades perceptivas de cualquier individuo, se decide que esta fase experimental no contemplará el estrés emocional que una agresión o delito genera en su víctima. Esta variable, achacable a la propia víctima, será estudiada en la siguiente etapa.

#### **4.2 Procedimiento**

El proceso experimental se inicia con la selección y grabación de una voz, la llamada “voz a reconocer”. Se trata de un sujeto cualquiera que narra un suceso personal real durante el que sufrió gran tensión emocional. Esa tensión queda recogida en el “registro sonoro”, un fragmento editado de 4’30”, que será sometido a audición por diversos grupos de sujetos experimentales (unos 12 grupos de unas 60 personas cada uno). Tras la audición se realizan grupos de discusión (ver apartado anterior) con algunos de los sujetos participantes en las audiciones. Los grupos de discusión permiten obtener información sobre los rasgos sonoros de la “voz a reconocer” y sobre el contenido de su narración y las palabras empleadas. Estas descripciones se utilizan para la selección de las voces de contraste y para la preparación de los textos dependientes y las entrevistas para la obtención de habla espontánea, variables a contemplar en las ruedas.

Atendiendo a las advertencias indicadas por los expertos en Ruedas de Reconocimiento Visuales sobre el error habitual de buscar participantes para las ruedas (voces falsas o de contraste) similares al sospechoso en lugar de ajustarlos a las descripciones del testigo [14], se decide que una parte del equipo investigador actuará bajo el procedimiento “sordo”, es decir, desconocerá las características de la “voz a reconocer” de manera que tanto los grupos de discusión como la selección de las voces de contraste se realicen observando siempre las informaciones dadas por los sujetos experimentales-testigos, y no inducidos por las percepciones personales de los investigadores.

Una vez tomadas las distintas muestras de voz a los informantes: voces de contraste y voz a reconocer, en sus distintas modalidades (con / sin tensión emocional; texto dependiente / independiente), se preparan las cadenas de voces para las ruedas de reconocimiento. La combinación de las variables independientes a estudiar arroja un total de 12 ruedas distintas. Cada una de estas ruedas está formada por cinco voces (cuatro voces de contraste y la “voz a reconocer”) combinadas siempre de la misma manera. Cada rueda es sometida a uno de los grupos de sujetos experimentales que participaron en las audiciones. Los participantes escuchan dos veces la cadena de voces y evalúan si reconocen la voz que escucharon con anterioridad. Entre la primera audición y la rueda de reconocimiento se ha mantenido estable el período de latencia, que es de 45 días. Las 12 ruedas tienen el objetivo de estudiar el rendimiento de las 12

combinaciones posibles (3x2x2) de las 3 variables independientes citadas en las condiciones siguientes.-

- duración de los fragmentos de texto: 15", 30" y 45";
- dependencia de la emoción original del locutor: voz emocionada, voz tranquila,
- dependencia del texto original: dependiente, independiente.

En el experimento participaron 438 sujetos experimentales entre 15 y 49 años de edad, mayoritariamente estudiantes. Para comprobar el grado de eficiencia de cada una de las 12 combinaciones entre las distintas condiciones de las variables se recurrió a que los sujetos efectuaran dos audiciones completas de la rueda antes de pedirles una decisión final sobre cual de las 5 voces era reconocida. Inmediatamente después de cada audición, se solicitaba a los sujetos experimentales que evaluaran el grado de parecido de cada una de las voces que iban escuchando con la voz que tenían en su memoria mediante un test. La técnica utilizada fue otorgar un valor entre 1 y 5 a cada voz según el grado de parecido con la voz recordada, siendo 1 el grado de parecido menor y 5 el de parecido mayor. Así, esta operación se repetía 2 veces y, finalmente, se preguntaba a los sujetos cual de las 5 voces que habían escuchado era la que se correspondía con la del locutor que oyeron 45 días antes.

Este planteamiento nos ha permitido trabajar con dos datos diferenciados, el de la decisión inmediata y el de la decisión reflexionada después de las 2 audiciones, generando dos variables dependientes distintas: a) la que denominamos "similitud con la voz 4" -con valores de 1 a 5- (la voz situada en 4º lugar era la correcta); y b) la que denominamos como "voz reconocida", que simplemente indicaba cual de las 5 voces de la rueda era la reconocida por el sujeto y que fue recodificada como acierto (valor 1) o como error (valor 0).

### 4.3 Resultados

Para estudiar la incidencia de las variables independientes "duración", "dependencia de la emoción" y "dependencia del texto" en la eficacia del reconocimiento se utilizó la prueba del análisis de la varianza (ANOVA) cuando manejábamos la variable dependiente "similitud con la voz 4". Y se recurrió a las tablas de contingencia y a la prueba de  $\chi^2$  cuando utilizábamos la variable dependiente "voz reconocida".

Al observar los resultados desde la perspectiva de la primera audición de los sujetos experimentales; es decir utilizando la variable dependiente "similitud con la voz 4" y tomando como hipótesis que las tres variables independientes tienen incidencia en la eficacia del reconocimiento nos encontramos con el siguiente resultado:

<i>variables</i>	<i>gr. liber</i>	<i>valor de F</i>	<i>Significación</i>
EMOCIÓN	1	2.807	0,095
<b>DURACIÓN</b>	<b>2</b>	<b>3.400</b>	<b>0,034</b>
<b>TEXTO</b>	<b>1</b>	<b>6.759</b>	<b>0.010</b>
<b>EMOCION/DURACIÓN</b>	<b>2</b>	<b>3.053</b>	<b>0.048</b>
EMOCIÓN/TEXTO	1	2.651	0,104
<b>DURACIÓN/TEXTO</b>	<b>2</b>	<b>3.608</b>	<b>0,028</b>
EMOCIÓN/DURACIÓN/ TEXTO	2	1.782	0,179
Error: 421			

La tabla que presentamos refleja que se produce la hipótesis alternativa para las variables DURACIÓN, TEXTO, EMOCIÓN/DURACIÓN y DURACIÓN /TEXTO. Es decir que existe

una diferencia significativa en los resultados del reconocimiento cuando se modifican DURACION o TEXTO, pero no cuando se modifica la EMOCION. No obstante, en tanto que las tres variables siempre actúan relacionadas entre sí, hemos de señalar que la variación conjunta de EMOCION/DURACIÓN sí que genera diferencias significativas en los resultados del reconocimiento. Observamos, además, que la influencia de la variable TEXTO es claramente significativa respecto a los resultados del reconocimiento. Pero, curiosamente, al combinarse entre sí EMOCION/DURACIÓN y TEXTO configurando la relación EMOCION/DURACIÓN/TEXTO dejan ya de serlo, a pesar de que tanto esta pareja como la variable TEXTO sí que lo son. Nuestra conclusión final respecto a esta tabla es que no refleja una coherencia global, lo cual nos indica que los datos de la primera audición son todavía poco fiables y no deben ser utilizados aisladamente.

Pasamos, en consecuencia, a trabajar con los resultados finales y con la variable “voz reconocida” como variable dependiente de referencia, manejándola ya recodificada como acierto o error en el reconocimiento de la voz correcta.

Los resultados obtenidos aplicando la prueba de  $\chi^2$  para estudiar la influencia de las variables independientes generando tablas de contingencia que las toma de dos en dos son los siguientes:

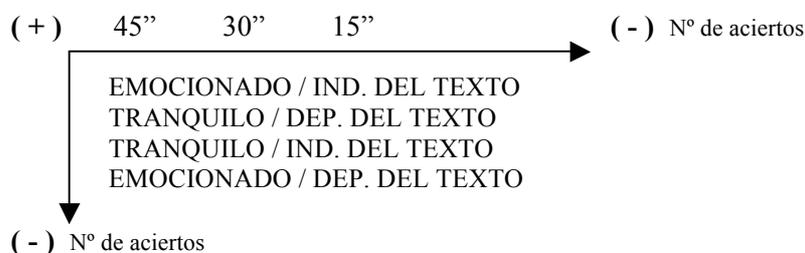
<i>variables relacionadas</i>	<i>sig. asintótica</i>
<b>IND. TEXTO / 45”</b>	<b>0,005</b>
IND. TEXTO / 30”	0,769
IND. TEXTO / 15”	0,058
DEP. TEXTO / 45”	0,918
<b>DEP. TEXTO / 30”</b>	<b>0,000</b>
DEP. TEXTO / 15”	0,945
<b>IND. TEXTO / EMOCIONADO</b>	<b>0,046</b>
IND. TEXTO / TRANQUILO	0,138
<b>DEP. TEXTO / EMOCIONADO</b>	<b>0,024</b>
DEP. TEXTO / TRANQUILO	0,307
<b>45” / EMOCIONADO</b>	<b>0,008</b>
45” / TRANQUILO	0,581
<b>30” / EMOCIONADO</b>	<b>0,048</b>
30” / TRANQUILO	0,199
15” / EMOCIONADO	0,403
15” / TRANQUILO	0,062

Los resultados de esta tabla resultan ya mucho más coherentes y podemos observar con claridad cómo la presencia de emoción en la voz es el elemento más significativo respecto al número de aciertos en la rueda de reconocimiento. La duración del texto se revela también como significativa cuando tiene una duración mínima de 45” o de 30”, pero ya no lo es cuando ésta se reduce a 15”. Y, finalmente, podemos observar, también la alta significación ( $p = 0.000$ ) que aparece cuando se asocian la dependencia del texto y una duración de 30”.

Para finalizar el análisis, estudiaremos el porcentaje de aciertos que obtienen los sujetos experimentales en función de las 12 combinaciones posibles entre distintas condiciones de las tres variables. De este modo sabremos cual es la potencia de cada combinación para mejorar el rendimiento en el número de aciertos:

<i>orden</i>	<i>Combinatoria de variables</i>	<i>% de aciertos</i>
1.	45'' / EMOCIONADO / INDEP	92,3
2.	15'' / EMOCIONADO / INDEP	84,6
3.	30'' / TRANQUILO / DEPEN	84,4
4.	15'' / EMOCIONADO / DEPEN	76,7
5.	45'' / TRANQUILO / INDEP	75,9
6.	15'' / TRANQUILO / DEPEN	75,6
7.	30'' / TRANQUILO / INDEP	73,2
8.	30'' / EMOCIONADO / INDEP	70,0
9.	45'' / TRANQUILO / DEPEN	69,2
10.	45'' / EMOCIONADO / DEPEN	68,4
11.	15'' / TRANQUILO / INDEP	56,1
12.	30'' / EMOCIONADO / DEPEN	45,9

De la observación de estas dos tablas de resultados podemos deducir las siguientes tendencias generales:



Es decir, las flechas expresan el sentido creciente o decreciente en el número de aciertos en función de como se hayan combinado las condiciones de las tres variables independientes. Así, podemos afirmar, por ejemplo, que una voz de una rueda de reconocimiento en las condiciones EMOCIONADO / IND. DEL TEXTO (*es decir, que reconstruya la emoción del locutor original y no utilice las mismas palabras que utilizó éste*) obtendrá un índice mayor de reconocimientos acertados cuando los fragmentos de voz que componen la rueda tengan una duración de 45'' que si la tienen de 30'' o de 15''. O que si las voces que componen la rueda están en las condiciones TRANQUILO / DEP. DEL TEXTO se conseguirá un mayor índice de reconocimientos acertados que si las condiciones fueran TRANQUILO / IND. DEL TEXTO.

#### **4.4 Segundo experimento: simulación realista**

Como ya hemos explicado anteriormente, la fase empírica final estuvo orientada a la experimentación concreta sobre protocolos de RRV. Se partió de la dramatización de una situación delictiva de manera que gracias a la simulación los sujetos experimentales participantes se convertían en *víctimas, testimonios y delincuentes* que podían colaborar en el desarrollo y aplicación de forma iterativa de procedimientos de ruedas de reconocimiento de voz. Es cierto que la bibliografía sobre psicología forense discute la fiabilidad de las simulaciones en laboratorio, ya que se considera que éstas no consiguen situar a la víctima en un mismo estado de estrés emocional que una agresión real [15]. Nosotros creemos que nuestra experiencia como estudiosos y profesionales de la comunicación audiovisual nos facilita la

tarea de generar “efecto de realidad” en aquello que es ficción, es decir, que las simulaciones que diseñamos tenían el objetivo de ser percibidos como hechos reales a los sujetos participantes, sobre todo teniendo en cuenta que contábamos con la colaboración de sujetos experimentales que desconocían los detalles concretos de su participación y la colaboración directa de la Policía Nacional en el experimento.

Las simulaciones implicaron a 80 sujetos experimentales que desempeñaron roles diferentes. Trabajábamos con Testigos-Víctima (TV), es decir, personas sobre la que se concretará la amenaza directa de un agresor; con Testigos-Observadores (TO), un grupo reducido de sujetos que serán testigos directos de las amenazas proferidas a la Víctima; y, finalmente, Observadores Experimentales (OE), un grupo más numeroso de sujetos a los que se les pidió que prestaran atención a la situación de simulacro, aunque sin dar mayor detalle sobre los hechos que se desarrollarían. El experimento se diseñó de modo que cada tipo de participante debía tener una intervención más o menos activa en los hechos simulados, y por tanto, su estrés emocional sería también distinto. De esta manera fue posible estudiar el papel de ese estrés en la memoria ecoico-vocal y los procesos de recuperación de información. Muy a grandes líneas, los simulacros, que se llevaron a cabo en la Escola de Policia de Catalunya (en las instalaciones que la escuela de la policía autonómica catalana tiene en Mollet del Vallès), contemplaron la simulación de un interrogatorio policial abierto al público en el que inesperadamente se producía un intento de evasión del interrogado con amenazas verbales y físicas a todos los presentes.

La simulación realista en la que se basaba el segundo experimento se realizó con éxito y el equipo de investigación, tras la recogida posterior de datos y el análisis de sus resultados llegó a las conclusiones finales, que ofrecemos a continuación.

## **5. CONCLUSIONES**

### **5.1 Conclusiones de la 1ª etapa: investigación cualitativa**

Conclusiones de la primera etapa de la investigación respecto a la entrevista de exploración cognitiva de la memoria ecoico-vocal, que pueden ser de utilidad tanto para la preparación de las ruedas de reconocimiento como para su realización:

- Las víctimas muestran un alto grado de memoria, tanto general (el suceso, su contexto, etc.) como auditiva: son capaces de describir con precisión la voz de su agresor y sus palabras.
- Comprobamos que era muy útil utilizar como voces de control y de referencia común entre el entrevistador y el testigo auditivo voces de famosos (o de cualquier otro tipo de locutor) que sean conocidas por ambos.
- La inducción al recuerdo, al pedir la narración de los hechos, favorece la memoria.
- Las víctimas consideran esencial para la identificación de la voz del agresor que en las RRV se reproduzca el contexto sonoro de la agresión y que las voces que integren la rueda repitan las mismas palabras y la misma actitud y tensión emocional del agresor. Sin embargo, la investigación experimental posterior muestra la dependencia del texto como una variable útil solamente en algunos casos.
- Las víctimas insisten en la necesidad de reducir al máximo el intervalo temporal entre la agresión y la RRV por miedo a la pérdida de su capacidad de evocación.
- Los investigadores proponen un protocolo de entrevista que ha resultado ser muy preciso, tanto para la reconstrucción del registro sonoro, como apoyo al recuerdo, reconstrucción del texto y descripción del sonido de la voz, que ponen a disposición de otros investigadores.

Siguiendo con la extracción de conclusiones en esta primera etapa y tras la utilización de los grupos de discusión surgió un listado de términos de apoyo que nos han permitido describir con mayor precisión los rasgos sonoros de una voz. Los términos que fueron identificados como útiles y funcionales son los siguientes:

- Rasgos físicos, fisiológicos y orgánicos (edad, sexo).
- Intensidad, volumen. Habría que contemplar también la variabilidad.
- Tono, tonalidad de una voz. Funciona al compararlo con una voz conocida o de famoso/sa.
- Timbre. Hay dificultad con algunos conceptos para describir la voz como por ejemplo: "mate".
- Ritmo. Velocidad de locución, pausas, silencios, etc.
- Características socioculturales. Además del acento habría que incluir el registro o tipo de vocabulario utilizado.
- Estilo personal del hablante.
- Presentan algunas dificultades los calificativos asociados al concepto de timbre, y sobre todo, el concepto de entonación.
- Los resultados de los grupos de discusión ponen de manifiesto la necesidad de disponer de una biblioteca de voces, de estereotipos vocales, de fonotipos que faciliten la descripción de cualquier voz a través de sus rasgos más relevantes.

## ***5.2 Conclusiones de la segunda etapa: primer estudio experimental***

Tras la etapa de investigación cualitativa sobre las variables del sistema que intervienen en la preparación de una RRV se decidió estudiar de forma experimental la influencia de las siguientes tres variables independientes:

- La duración de cada fragmento de voz que componen las cadenas de las RRV. Se trabajó con duraciones de 15"-30"-45".
- La reproducción o no de la actitud emocional del sospechoso durante el hecho delictivo en las cadenas de RRV.
- Las características del discurso portador. Se trabajó con ruedas que contenían texto libre y ruedas que contenían frases repetidas.

Tras desarrollar 12 RRV a 438 sujetos experimentales organizados en 12 grupos distintos y someter los datos resultantes a un análisis estadístico llegamos ya a las siguientes conclusiones:

- La presencia de emoción en la voz<sup>16</sup> es un elemento significativo respecto al número de aciertos en la RRV.
- También la duración del texto se revela como significativa cuando tiene una duración mínima de 45" o de 30", pero ya no lo es cuando ésta se reduce a 15".
- Para mejorar el rendimiento en el número de aciertos la combinación óptima es de las variables siguientes: tiempo: 45"/ voz emocionada / texto independiente. Con un 92,3 por ciento de aciertos.

La dependencia del texto en las RRV (uso de las mismas palabras que utilizó el locutor que se intenta reconocer) se ha revelado como muy eficiente cuando no hay dependencia de la emoción original.

### ***5.3 Conclusiones de la última etapa de la investigación:***

En esta última etapa de la investigación se presentan los resultados definitivos sobre la influencia del grado de estrés emocional y del tiempo de latencia en la memoria auditiva. Y se desarrolla una evaluación definitiva sobre la fiabilidad global de los protocolos que hemos experimentado.

Nos parece importante señalar que el análisis estadístico de los resultados del último experimento del estudio refleja una influencia muy débil tanto de la impresión emocional (grado de estrés) como del tiempo de latencia del recuerdo de la voz

Cuando el análisis se hace a partir de los resultados de acierto y error después de la segunda audición y de la decisión final, no aparecen en ningún momento diferencias significativas que reflejen en el índice de aciertos la influencia de las dos variables independientes (impresión emocional y tiempo de latencia). No obstante, estas diferencias sí que aparecen (aunque muy leves) cuando el análisis de los resultados de acierto/error se hacen después de la primera audición. Es decir, en un momento en que el sujeto todavía no siente la responsabilidad de tener que tomar una decisión y sabe que va a volver a escuchar todas las voces de nuevo antes de hacer la identificación perceptiva. En esa situación, observamos que tanto la impresión emocional como el tiempo de latencia influyen en el recuerdo, ambas en un sentido negativo. A mayor impresión emocional menos aciertos, y a mayor tiempo de latencia del recuerdo menos aciertos.

En cualquier caso, el resultado final vuelve a mostrarnos otra vez la alta fiabilidad de las Ruedas de Reconocimiento de Voz (RRV) si éstas están apoyadas con un protocolo sistemático que estimule el recuerdo del testigo.

#### ***Respecto a la fiabilidad global del procedimiento RRV***

Para hacer un estudio global de la fiabilidad se ha configuró un único fichero de datos con los resultados del segundo y el tercer experimento integrando las variables “edad”, “sexo”, “voz reconocida” y “aciertos”. A partir de este fichero único que recopila los 438 sujetos del primer experimento y los 71 del segundo, hemos podido analizar la influencia de distintas condiciones del protocolo de las Ruedas de Reconocimiento de Voz (RRV) sobre un total de 509 sujetos.

Tomando los datos de ambos experimentos sin hacer en ellos ningún tipo de selección ni de depuración, es decir, tomando conjuntamente todos los resultados de acierto y error en cualquiera de las condiciones estudiadas de la RRV, el análisis de las frecuencias muestra lo siguiente:

aciertos: **75,1 %**    errores: **24,9 %**

Con esto resultados podemos afirmar con una probabilidad de error inferior al 1 por 1000 que el acierto al reconocer en una RRV una voz que ha sido escuchada antes en cualquiera de las condiciones estudiadas, es un hecho que no responde al azar sino a la audición previa.

A partir de este resultado, hemos realizado sucesivas selecciones de los datos de modo que se han ido seleccionando los sujetos en las condiciones que fueron identificadas a lo largo

de estas investigaciones como más favorables a la memoria auditiva, con objeto de estudiar el índice de aciertos y errores en esas condiciones.

Las selecciones fueron las siguientes:

1. Se depuraron los datos totales seleccionando los sujetos que fueron expuestos a la RRV según los siguientes criterios:

- a) Se eliminaron los sujetos indecisos
- b) Se seleccionaron los sujetos del 1er experimento en condiciones óptimas (duración de 45", voz emocionada e independencia del texto)
- c) Se seleccionaron los sujetos del 2º experimento en condiciones óptimas (30 días de latencia)

En estas condiciones obtenemos la siguiente relación de acierto-error sobre una muestra de 71 sujetos:

Aciertos: **93,0 %** Errores: 7,0 %

2. Se separaron los datos de ambos experimentos, seleccionando para cada uno de ellos solamente los sujetos que han estado sometidos a la RRV en las condiciones identificadas como óptimas. En estas condiciones se obtuvieron los siguientes resultados:

Para el 1er experimento (duración 45", voz emocionada, independencia del texto) y eliminando los casos indecisos, el resultado de la muestra de 37 sujetos es:

Aciertos: **97,3 %** Errores: 2,7 %

Para el 2º experimento, el cual se realizó ya en las condiciones óptimas para el sistema (duración de 45", voz emocionada, independencia del texto), seleccionamos solamente los 29 casos de la segunda simulación, en la que se consiguió una verosimilitud mucho mayor tanto por parte del actor como de los policías :

Aciertos: **100 %** Errores: 0 %

Si estudiamos la relación de acierto-error para este mismo grupo de sujetos tras la primera audición aparece ya de nuevo un porcentaje de error del **13,8 %**, lo que nos indica con mucha claridad la alta fiabilidad del protocolo que se ha diseñado. Resulta evidente a lo largo de todo el análisis de los resultados que tras la segunda audición y la toma de decisión el error se reduce de manera drástica, llegando a desaparecer por completo.

#### ***5.4. Beneficios científicos***

Los beneficiarios directos de estos conocimientos serán disciplinas tales como la psicología, la acústica o la comunicación audiovisual. Pero además, la principal aportación del estudio, el protocolo validado para las RRV, supondrá un importante avance en el contexto de la investigación forense. La investigación que se presenta en este texto está proporcionando nuevos conocimientos relacionados con las formas de percepción, retención y recuperación de la información auditiva, sobre cómo se comportan el oído y la mente humana en los procesos perceptivos de reconocimiento de formas sonoras. Estas aportaciones son claramente aplicables

en campos como la propia acústica forense, las telecomunicaciones, las ingenierías de las comunicaciones y del habla y, evidentemente, la comunicación audiovisual y su industria.

El trabajo desarrollado ha permitido avanzar científicamente en las dos líneas previstas:

1. Una línea de investigación aplicada al desarrollo de un procedimiento fiable para la identificación perceptiva en la acústica forense, en la cual se han obtenido resultados específicos muy concretos;
2. Una línea de investigación metodológica que aplica los nuevos conocimientos obtenidos en este proyecto al desarrollo de la investigación en el ámbito de la comunicología en su sentido más amplio (expresividad oral, síntesis del habla, uso de la voz en la educación, en la publicidad, etc.).

En la primera línea, entendemos que el resultado más relevante de este proyecto es la constatación y la contrastación científica, en contra de toda sensación subjetiva y de toda intuición, de que la memoria sonora asociada al reconocimiento de la voz tiene una altísima fiabilidad y, por tanto puede ser utilizada como prueba policial y jurídica.

En la segunda línea, consideramos que el resultado más relevante de la investigación es que ha permitido depurar y asentar, trabajando sobre un problema real concreto, un nuevo modelo metodológico para el desarrollo de investigaciones en el ámbito de la comunicología aplicada. Este nuevo modelo permite objetivar los estudios sobre comunicación y los abre definitivamente a otros entornos técnicos e industriales.

---

<sup>1</sup> RODRÍGUEZ BRAVO, A.; LÁZARO PERNIAS, P.; MONTOYA VILAR, N.; BLANCO PONT, JOSEP M<sup>a</sup>.; BERNADAS SUÑÉ, D.; TENA PARERA, D.; LONGHI, L.; OLIVER COMES, J.M.: "Identificación perceptiva de locutores para la acústica forense." SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense, pp. 23-34 Barcelona, 10 y 11 de abril de 2003. ISBN: 84-86108-32-2.

<sup>2</sup> RODRÍGUEZ BRAVO, A: La investigación aplicada: una nueva perspectiva para los estudios de la recepción. En Anàlisi: Quaderns de comunicació i cultura. N<sup>o</sup> 30, pág. 17-36. ISSN

<sup>3</sup> RODRÍGUEZ BRAVO, Ángel: *La dimensión sonora del lenguaje audiovisual*, Paidós Papeles de Comunicación, Barcelona, 1998.

<sup>4</sup> Cfr. RUIZ VARGAS, J.M: *Psicología de la memoria*, Alianza Editorial, Madrid, 1998, pp.38.

<sup>5</sup> Op. Cit. Cfr. RUIZ VARGAS, J.M: *Psicología de la memoria*, Alianza Editorial, Madrid, 1998, pp. 164.

<sup>6</sup> Op. Cit. Cfr. RUIZ VARGAS, J.M: *Psicología de la memoria*, Alianza Editorial, Madrid, 1998, pp. 195.

<sup>7</sup> Op. Cit. Cfr. RUIZ VARGAS, J.M: *Psicología de la memoria*, Alianza Editorial, Madrid, 1998, pp. 307-309

<sup>8</sup> PATRÍCIA LÁZARO PERNIAS; ÁNGEL RODRÍGUEZ BRAVO; NORMINANDA MONTOYA VILAR, JOSEP M<sup>a</sup> BLANCO PONT; DOLORS BERNADAS SUÑÉ; DANIEL TENA PARERA; LUDOVICO LONGHI; JOSEP MANEL OLIVER COMES: un caso de comunicología aplicada: la identificación perceptiva de locutores para la acústica forense. Actas del IV Congreso Internacional Comunicación, Universidad y Sociedad del Conocimiento. Publicaciones Universidad Pontificia de Salamanca. Salamanca, 2006. ISBN 84-7299-684-0.

<sup>9</sup> Cfr. GEISELMAN, E, FISCHER, R: "La técnica de la entrevista cognitiva para víctimas y testigos de crímenes", en RASKIN, David C: *Métodos psicológicos en la investigación y pruebas criminales*, 1994.

---

<sup>10</sup> MONTOYA, N: *El papel de la voz en la publicidad audiovisual dirigida a los niños*, UAB, 1996; MONTOYA, N; LÁZARO, P; RODRÍGUEZ, A: “Estereotipos vocales de carácter en la publicidad infantil”, en *Comunicación y Pedagogía*, n.171, pp. 69-76.

<sup>11</sup> Cfr. DIGES, M; ALONSO-QUECUTY, M: *Psicología forense experimental*, Promolibro, Valencia, 1993, pp. 45.

<sup>12</sup> Cfr. BROEDERS, A.P.A.; AMELSVOORT, A.G.: “lineup Costruction for Forensic Eartwitness Identification: a Practical Approach” in *Versie 2.1*, September 2000.

<sup>13</sup> HOLLIEN, H: “Considerations of guidelines for eartwitness lineups”, en *Forensic Linguistics*, n.3 (1), 1996, pp.14-23.

<sup>14</sup> Cfr. Op. Cit. DIGES, M; ALONSO-QUECUTY, M: *Psicología forense experimental*, Promolibro, Valencia, 1993, pp. 9.

<sup>15</sup> Cfr. Op. Cit. DIGES, M; ALONSO-QUECUTY, M: *Psicología forense experimental*, Promolibro, Valencia, 1993.

<sup>16</sup> ISCA (*International Speech Communication Association*) Workshop: IRIONDO, Ignasi, GUAUS, Roger, RODRÍGUEZ, Ángel, LÁZARO, Patricia, MONTOYA, Normi, BLANCO, Josep M<sup>a</sup>, BERNADAS, Dolors, OLIVER, Josep Manel, TENA; Daniel i LONGHI, Ludovico, “Validation of an Acoustical Modelling of Emotional Expression in Spanish Using Speech Synthesis Techniques”, en COWIE, R, DOUGLAS-COWIE, E, SCHRÖDER, M (eds.): *Speech and Emotion: A Conceptual Framework for Research*, ISCA, Belfast, 2000, pp. 161-166.