Homenaje al profesor Ambrosio Rabanales
BFUCh XXXVII (1998-1999): 161-180

Rendimiento de parámetros acústico-fonéticos en el reconocimiento de hablantes

Mario Bernales Lillo, Jorge Díaz Villegas y Robert Paredes Ruminot
Universidad de la Frontera, Chile

Ricardo de Figueiredo Molina
Universidad Estatal de Campinas

Este es un trabajo de carácter interdisciplinario en el campo de la lingüística y la computación que da a conocer el desarrollo de un sistema automatizado para el reconocimiento de hablantes a través de la voz humana, partiendo de un corpus *fónico grabado por un número potencial de informantes seleccionados al azar que constituyen la muestra de la investigación. La selección de los mismos se hizo con el apoyo de un fonoaudiólogo y entre las características principales se tuvo en cuenta que fueran hablantes del dialecto del español de Chile (ciudad de Temuco); pertenecieran al nivel estándar; fueran de sexo masculino y no presentaran dificultades articulatorias.*

El registro de sus voces se realizó en dos ocasiones con un intervalo de tiempo equivalente a seis meses, con el objeto de comparar más tarde los valores fonético-acústicos de ambos registros. Las grabaciones reunieron diversas modalidades de habla, tanto espontánea como aquella proveniente de la lectura de textos (fluida, veloz, de tablas numéricas aleatorias, de combinaciones silábicas, de palabras y frases claves).

Para el análisis de los parámetros acústico-fonéticos (Fo, Formantes, ELT, FFT, Amplitud, Coeficientes de Reflexión, etc.) de cada informante se utilizó el equipo Computerized Speech Laboratory *(CSL) 4300B fabricado por la casa Kay Elemetrics. Con ellos se construyó la base de datos en un archivo con formato especial* (Software de

Identificación de Hablantes). *Luego se procedió a 'alimentar' la red neuronal para que ésta 'aprendiera' los padrones acústicos de cada hablante, optimizando los coeficientes de penalidad estadísticos de que dispone la red. Finalmente, en el proceso de identificación se ingresó la información correspondiente a un segundo análisis acústico del mismo informante (etapa predictiva) y, de este modo, la red debía entregar resultados estadísticos y probabilísticos correspondientes a cada una de las variables consideradas. Así, ella aceptó o rechazó la identificación del hablante.*

La posibilidad de utilizar los parámetros acústicos como dispositivos complementarios de identificación tendrá gran aplicación a corto plazo. Los sistemas de verificación automática de hablantes podrán ser empleados, por ejemplo, en transacciones bancarias y comerciales, como filtros para el acceso a áreas restringidas, en telefonía, como medios para identificar a personas que ponen en peligro la seguridad individual, etc.

1. INTRODUCCIÓN

1.1. Antecedentes

En esta última década existe un considerable interés por desarrollar procedimientos y abrir un campo de posibilidades dentro de la investigación fonética que permitan identificar personas a través de la voz e incorporar los resultados de estos estudios, principalmente, a los sistemas de seguridad nacional, al acceso restringido a determinadas informaciones, a las agencias oficiales de la policía, a la telefonía, a las transacciones bancarias y comerciales y a otros ámbitos insospechados hasta ahora.

Como se sabe, la identificación de hablantes se inserta en un cuadro más general, conocido en la literatura especializada, como Reconocimiento de Padrones *(Pattern-Recognition)* y puede ser considerada como un ejemplo de identificación personal biométrica, la cual es diferente a otras técnicas que se emplean en la identificación de personas inherentes al individuo, tales como, las impresiones digitales, padrones del iris, retina, estructura genética, etc. Sin embargo, la característica más importante de la señal del habla es que debe ser entendida como un fenómeno complejo la cual no solo envuelve aspectos articulatorios, sino que también incluye factores socioculturales, neurológicos, emocionales, etc. (Laver y Trudgill, 1979). Y, en consecuencia, la información útil para identificar a un hablante viaja

indirectamente en la señal del habla, es decir, en forma paralela al proceso articulatorio. En cierta forma, la información inherente al hablante puede ser vista como un *ruido* aplicado sobre el mensaje lingüístico básico.

De otro lado, existen personas que presentan características en la señal de habla que son bastante particulares. La experiencia personal de cada uno demuestra que existe una gran habilidad humana capaz de reconocer personas por la voz, muchas veces en condiciones adversas (ruidos, señales muy bajas, interferencias). El gran desafío que se presenta ahora para el científico del habla es establecer un modelo que reproduzca esa habilidad humana (sin que ello exija reproducir los mismos procesos humanos). Este desafío ha motivado, en las últimas décadas, un gran número de investigaciones en el área de la identificación de hablantes, un área que ha recibido un extraordinario impulso con el desarrollo de los sistemas de procesamiento digital de señales. Con los múltiples estudios y enfoques, la cuestión específica del reconocimiento de hablantes adquirió otro matiz y tuvo aplicaciones y subproblemas específicos, entre ellos el modelo forense (Molina, 1994; Hollien, 1990).

Varios términos y expresiones surgieron enseguida para denominar estos nuevos estudios -Identificación, Verificación, Discriminación, Autenticidad de la voz o del hablante *(speaker o talker)*- sin que esto hiciera referencia a otros aspectos diferentes y específicos. En general, hoy se acepta el término genérico Reconocimiento de Hablantes como una expresión que engloba todos los procesos de decisión (humanos o automatizados) que utilizan trozos de la señal de habla para determinar si una persona *es* o *no es* quien ha emitido el mensaje (Atal, 1976). Sin embargo, conviene saber que la discusión acerca de reconocimiento de hablantes puede ser enfocada en dos sentidos diferentes: (1) Identificación de Hablantes *(Identification)* y (2) Verificación de Hablantes *(Verification)*, según (Doddington, 1985; Braun, 1995; Broeders, 1995; etc.). El primer caso consiste en atribuir un enunciado emitido por una persona desconocida perteneciente a un grupo de **N** hablantes. Por lo tanto, si el grupo es cerrado, el proceso de decisión tiene **N** salidas posibles y **N+1** salidas si éste es abierto (incluso, existe una tercera posibilidad, que el enunciado no corresponda a ninguna persona del grupo). El segundo caso, la situación clásica de Verificación, consiste en determinar y decidir si el padrón de voz de un hablante desconocido es suficientemente semejante al padrón de referencia. Entonces, aquí el proceso de decisión se reduce a una elección binaria que ofrece solo dos salidas posible. En resumen, en la Verificación se exige apenas una comparación de padrones, independientemente del tamaño de la población.

Cabe señalar que el modelo forense frecuentemente aparece asociado al paradigma de Identificación, aunque nada impide, de acuerdo con la situación forense típica que este modelo también se asocie con la Verificación, como explica Doddington (1985), cuando se trata de grupos muy numerosos. En la mayor parte de los casos forenses lo que se espera del perito es una decisión binaria: la voz cuestionada *es* o *no es* la voz del sospechoso.

En este sentido, el paradigma de Verificación de Hablantes resulta más fácil si se piensa, por ejemplo, en procedimientos automatizados. En estas aplicaciones el informante hasta es capaz de colaborar, al contrario de lo que sucede en la Identificación (en el modelo forense), donde el informante examinado eventualmente puede intentar disfrazar algún aspecto de su voz natural. En el paradigma de Verificación el intento de fraude (engaño), por ejemplo, recae en la imitación y como tal es un problema menos complicado que el disfraz o alteración de la voz. Aunque algunos padrones acústicos resulten alterados (Endres *et al.*, 1971) como el Espectro de Largo Tiempo (ELT), (Doherty 1975; Hollien e Majewski 1977), otros, como los temporales, parecen ser relativamente resistentes (Johnson *et al.*, 1984).

La investigación de laboratorio en el campo de la Identificación de Hablantes tiene algunas décadas y muestra diferentes técnicas y metodologías. Por ejemplo, Hecker (1971) reconoció tres métodos básicos: (1) a través de la escucha de cintas magnéticas; (2) por medio de sistemas automatizados; y (3) por la inspección visual de espectrogramas (llamada también *aproximación perceptual, automática* y *espectrográfica*).

La clasificación propuesta por Hecker fue aceptada como criterio de referencia básico y tuvo varios seguidores, entre los que destacan Bricker y Pruzansky (1976); Atal (1976); Rosenberg (1976); Doddington (1985); Hollien (1990). Sin embargo, Nolan (1983) en uno de sus trabajos más conocidos sobre Identificación de hablantes, discutió esta clasificación tripartita y propuso una división apenas de dos categorías que llamó: (1) *technical speaker recognition* y (2) *naive speaker recognition*. De acuerdo con su argumentación, la diferencia básica entre ambas propuestas debe ser entendida en términos de empleo y no de *técnicas analíticas* para resolver el problema, independientemente del hecho de que sean técnicas adquiridas por humanos o programadas automáticamente. Nolan, además, afirma que la división entre los métodos automáticos y el examen de espectrogramas es meramente contingente, ya que para un observador entrenado es posible realizar medidas confiables basadas en los espectrogramas, las cuales podrían servir como *input* para estrategias posteriores de decisión automática, o sea, una especie de sistema híbrido, semiautomático.

Esta especie de habilidad auditiva para identificar hablantes entre oyentes *naive* y especializados ha sido discutida y hoy se encuentran opiniones

divergentes basadas en el análisis instrumental, el cual mostró el peligro que podrían tener estas afirmaciones que algunas veces lindaron en el llamado terreno de las 'opiniones'. Si el análisis técnico-instrumental actual, especialmente el digitalizado, entregó otros resultados, ya no valía la pena mantener la división de Nolan.

Cuando este autor se refiere al examen de los espectrogramas conviene destacar la importancia de su perspectiva. En este caso, se puede establecer una división más clara entre una aproximación técnica respecto de una aproximación *naive*. En el segundo caso, la identificación será el mero resultado de la comparación visual de padrones gráficos, mientras que el examen técnico presupone un cierto nivel de estructuración. Se debe observar que hay una distinción fundamental en relación a la **habilidad auditiva** en reconocer hablantes, ya que en último término se trata de una capacidad naturalmente adquirida, inherente a los esquemas internos de representación que se manifiestan desde muy temprano (un niño de brazos reconoce la voz de su madre mucho antes de procesar la información de nivel lingüístico).

De paso, se señalará que existen algunos procedimientos experimentales en relación con la Identificación de Hablantes particulares que consideran características generales, tales como: (1) su constitución física (diferencias anatómicas, dimensiones del aparato fonador), (Lass y Davis, 1976; Lass *et al.*, 1980); (2) el sexo (variación de Fo de acuerdo con el largo de las cuerdas vocales entre hombres, mujeres y niños), (Behlau, 1984; Wu y Childers, 1991); (3) la edad (especialmente alteraciones fisiológicas a nivel de la laringe modifican la señal acústica en función de la edad), (Endres *et al.*, 1971; Hollien y Shipp, 1972).

En forma paralela, en estos experimentos de carácter perceptual para reconocer hablantes figuran: (1) el material de habla utilizado como estímulo, enfatizando los efectos de duración y contenido del mismo (Pollack *et al.*, 1954; Bricker y Pruzansky, 1966); y (2) las condiciones del canal de transmisión, demostrando la experiencia cotidiana que el reconocimiento de hablantes a través del canal telefónico impone un mayor grado de dificultades, más los ruidos o distorsiones naturales del ambiente (Künzel, 1990).

1.2 Parámetros fonético-acústicos

Los principales parámetros fonético-acústicos que se utilizaron en este trabajo son conocidos y han sido extensamente comentados en relación con el tema de la identificación de hablantes (Molina, 1994). Sin embargo,

aquí se discutirán brevemente las contribuciones más significativas sobre este punto.

En relación con los Formantes y Frecuencia fundamental *(fo)*, diversos estudios comprueban que la frecuencia de los formantes de las vocales presenta una notable variabilidad interhablantes, aunque se utilicen contextos fonéticos fijos (Peterson y Barney, 1952). Los formantes de una vocal no son estables a lo largo de todas las condiciones de producción, hay factores como la velocidad de emisión, el contexto fonético, la acentuación de las voces, etc, que pueden ejercer una influencia considerable en las características espectrales de una vocal (Lindblom, 1963).

Cabe señalar también que los resultados de algunas investigaciones demuestran que las frecuencias de los formantes pueden ser una pista importante para la identificación de un hablante. La importancia de la estructura formántica en la individualidad de la voz es destacada en el experimento de Kuwabara y Takagi (1991). En este estudio, a través de la modificación de *fo*, de los formantes y del largo de banda de las vocales en palabras *nonsense,* se constató que las alteraciones de los formantes provocaba las distorsiones más sensibles en el reconocimiento de voces que hasta eran familiares para los oyentes.

Varios experimentos han confirmado la importancia perceptual de *fo*, su variación en el habla compromete la cualidad vocálica y también la percepción del mensaje (Foulkes, 1961). Por esa razón, se afirma que los imitadores profesionales, por lo general, intentan aproximarse al *fo* medio de la persona que imitan, aunque no consigan una coincidencia exacta (Hall y Tosi ,1975).

Otro parámetro acústico importante considerado en esta oportunidad corresponde al *Coeficiente de Reflexión*, el cual, como ya se conoce, entrega valores relativos al comportamiento de las cavidades del tracto vocálico, es decir, muestra las variaciones de este filtro a través del tiempo. Para extraer los valores de este coeficiente se utiliza el LPC *(Linear Predictive Coding)* y, al mismo tiempo, al aplicar esta técnica el investigador dispone automáticamente de otras posibilidades, como son, por ejemplo, el *Peak, Length,* FFT (Fourier Fast Transform), Cesptrum, las cuales puede seleccionar a voluntad según las características del estudio.

1.3 Redes Neuronales Artificiales (RNA)

En relación con las Redes Neuronales Artificiales (RNA), éstas nacieron de las ideas de los conexionistas allá por 1943, con los trabajos de McCulloch y Pitts (Mc Culloch, y Pitts,1943), quienes, formularon un

modelo de neurona lógica que podía ser conectada en red, simulando la conectividad del cerebro humano.

El surgimiento de la RNA se localiza en los años 1982 a 1984 con los trabajos de John J. Hopfield (Hopfield, 1982; Hopfield, 1984). También hay aportes de Grossberg (Grossberg, 1980), J. A. Anderson (Anderson, 1983), Kohonen (Kohonen, 1982), Fukushima (Fukushima, 1980) y otros. A partir de 1985 se inician los estudios en otros campos de aplicación (Soluciones de serie de tiempo, Análisis de parámetros económicos, Medicina, Meteorología), los cuales robustecen el campo de las Redes Neuronales Artificiales. En cuanto a la aplicación de esta técnica de Inteligencia Artificial, este equipo de trabajo es el primero en Chile que propone utilizarla en el área de la *fonética forense*, especialmente en el modelamiento de las variables acústicas de la onda sonora.

Las Redes tienen diferentes arquitecturas, siendo posible enmarcarlas en una sola o en una combinación de las arquitecturas de Red. El aprendizaje de una Red se lleva a cabo mediante reglas, que son métodos de tipo matemático (Miller, y Reinhardt, 1990). Tenemos dos tipos de aprendizajes: Supervisado y No-Supervisado. Además, existe un método llamado Hebb que no se relaciona con los tipos de aprendizajes anteriores.

El Aprendizaje Supervisado se caracteriza porque la Red aprende en base de un conjunto de ejemplos. Puede ser modelada mediante un sistema rápido y otro lento (Iost, y Rivera, 1993). Dentro del Aprendizaje Supervisado tenemos categorías de redes: Retropropagación de Errores, Máquinas de Boltzman y otros.

El Aprendizaje No Supervisado tiene como objetivo obtener correlaciones, crear categorías y describir características comunes entre las entradas, por lo tanto no requiere de patrones entrada-salida, sino que solo de entrada. No existe un modelo general en este tipo de aprendizaje, pero existen dos modelos que se destacan: Aprendizaje Competitivo y el modelo de Kohomen.

En este trabajo se aplica solamente un tipo de aprendizaje de Red, la *feedforward* (alimentación hacia adelante) que se basa en un subconjunto de la clase de regresión no-lineal y modelos discriminatorios. Se debe notar que tal Red tiene dos capas ocultas (capa B y capa C). El número de neuronas en la capa B es igual al producto del número de todos los vectores modelo N y el número de variables de entrada M (N M), mientras el número de neuronas en la capa C es igual 2 veces al número de vectores modelo (aiNet, 1996).

1.4 La fonética y el ámbito de trabajo interdisciplinar

El proceso de producción y percepción de los sonidos articulados es estudiado teórica y experimentalmente por la fonética. Al referirse a la *experimentalidad* de la fonética, como dice Joaquín Llisterri (1991), se desea señalar que el estudio del habla en el campo de la Lingüística no se basa solo en la instrospección, sino en datos provenientes de situaciones comunicativas naturales, analizados con métodos que permitan conocer profundamente las propiedades físicas de la señal sonora, relativa a los mecanismos de producción y a los de recepción, de tal modo que las conclusiones sean extraídas de un número de datos matemático-estadísticos altamente representativos para hablantes diferentes.

Por lo tanto, todo esto supone que el fonetista tenga que aprender recursos de trabajo propios de otras disciplinas, como la fisiología, la física, la informática, el procesamiento de señales (análogas/digitalizadas), estadística, psicología, etc., con lo cual la investigación actual se torna interdisciplinar.

Un ámbito de aplicación moderno de la fonética y de importancia reciente es el de la *tecnología del habla*, la que tiene que ver con el conocimiento sobre el habla y la identificación de personas a partir de sus voces. Hoy, se aplica en varios países (Estados Unidos, Alemania, Japón, Brasil, Francia, etc.) en el ámbito judicial y constituye una especialidad denominada y conocida como *fonética forense.*

Harry Hollien (1990) dice en su libro titulado *The acoustics of crime. The new sciencie of forensic phonetics* que la primera razón que tuvo para escribir esta obra es para hablar de esta nueva subárea de la ciencia forense, cuya extensión ha sido muy rápida en esta década y es preciso conocer sus antecedentes, evolución, metodologías de trabajo y técnicas, así como su naturaleza y extensión. En un artículo más reciente (1995), este mismo autor discute cuestiones relativas a las características del hablante que se desea identificar, el posible parentesco entre los informantes, los instrumentos y sus procesos técnicos para el análisis de la señal sonora, la importancia de las grabaciones y la incidencia del ruido, la constitución de grupos interdisciplinarios (fonetistas, ingenieros de audio, psicólogos), etc.

También Braun (1995) del Bundeskriminalamt de Alemania, discute las aplicaciones de la fonética forense y las controversias que ésta ha generado en instituciones como *Voice Identification and Acoustic Analysis Subcommittee* de la *International Association for Identification*, o en la *British Association of Academic Phoneticians*, o en la *Société Française d'Acoustic*, etc.; los sistemas metodológicos generados (especialmente, la audición y la espectrografía) y las respuestas de los laboratorios. Algo similar nos

entrega el artículo de Broeders (1995), destacando en el punto 3 las aplicaciones comerciales de la fonética forense y los problemas comunes que experimenta este tipo de investigación.

De otro lado, la relevancia alcanzada por el tema se observa en los congresos o reuniones científicas de carácter internacional. Actualmente cuenta con secciones especiales, como ha sucedido, por ejemplo, en algunos congresos, como: *VIII International Conference of IAFP (International Association of Forensic Phonetics),* Alemania (1996); *The XIIIth International Congress of Phonetic Sciences*, Suecia (1995); *Eurospeech '97, 5th European Conference on Speech Communication and Technology*, Grecia (1997); *SPECOM'97, Second International Workshop «Speech and Computer»*, Rumania (1997); etc., por citar los más importantes.

Casi del mismo modo, se podría explicar la creación de la prestigiosa publicación *Forensic Linguistics. The International Journal of Speech, Language and the Lawel*, Estados Unidos (1996) o *Studies in Forensic Phonetics*, Londres (1990), las cuales contienen trabajos especializados sobre el tema, sobre síntesis de habla, simulación computarizada, uso del computador en mediciones, uso del análisis matemático constructivo y, más recientemente, aplicación de redes neuronales a diversos aspectos fonéticos.

Por último, hoy los investigadores de esta área también disponen de ediciones electrónicas que aparecen bajo diferentes nombres, como, por ejemplo, IDEAL *(International Digital Electronic Access Library). Computer Speech & Language,* donde se proveen artículos de carácter interdisciplinario relativos a la ciencia fonética, a través de los cuales se dan a conocer experimentos recientes sobre modelos del proceso del habla que pueden ser factibles, aprovechando el uso de la inteligencia artificial, la ciencia computarizada, la ingeniería electrónica, modelos matemáticos, etc.

Cabe señalar además que, debido a los problemas diarios de violencia e inseguridad personal que sufre actualmente parte de la sociedad chilena (y a otros mecanismos más sofisticados que elaborarán los delincuentes en la próxima década), donde la voz juega un papel preponderante en los sistemas de comunicación moderna, los autores de este trabajo iniciaron hace dos años un estudio relacionado con la identificación de hablantes a través del habla, en el marco del Proyecto FONDECYT 1960815 desarrollado en la Universidad de la Frontera, cuyos resultados se exponen en las páginas siguientes, con el fin de contribuir al conocimiento y desarrollo de sistemas relativos a la seguridad personal. Los resultados y las conclusiones mostraron que esto es posible hacerlo en forma interdisciplinaria, aplicando metodologías y técnicas desarrolladas por el equipo de investigación

(Bernales *et al.* 1998) y, de este modo, lograr una estimación predictiva del hablante bajo condiciones ideales de laboratorio. Es decir, lo más importante en este momento es que hoy los valores matemático-estadísticos demuestran que se puede identificar una persona por medio de su voz.

Con el constante aumento de la red mundial de comunicaciones, por diversos medios (telefonía, internet, fax, e-mail, etc.), aumenta también la importancia de tener seguridad en el intercambio de informaciones. La posibilidad de utilizar los parámetros vocales como un dispositivo suplementario de identificación es, por lo tanto, de gran aplicabilidad. Así, sistemas de verificación automática de hablantes podrían ser empleados en transacciones bancarias y comerciales, como filtros para el acceso a áreas restringidas, etc.

1.5 Hipótesis y objetivos

Considerando el problema planteado, se piensa que la hipótesis de trabajo consiste en determinar que, en la medida que se descubra un sistema de reconocimiento automatizado para identificar voces, eventualmente de personas que atenten contra la seguridad personal, ésta sería una alternativa confiable en beneficio de la citada seguridad personal.

En el marco de la hipótesis, el objetivo general de este trabajo es crear entonces una metodología de análisis interdisciplinaria en la cual sean consideradas diferentes áreas de la ciencia (Lingüística-Informática-Fonoaudiología) con el propósito de identificar hablantes a través de grabaciones digitalizadas, con aplicación de redes neuronales artificiales.

Entre los objetivos específicos se propone: (1) Reconocer los padrones acústico-fonéticos de cada hablante e incorporarlos a la base de conocimiento de la red neuronal con el objeto de que ésta "aprenda" dichos padrones y sea capaz de identificarlos cuando corresponda a una segunda señal sonora perteneciente al mismo hablante. Eventualmente, la RNA puede ser perfeccionada constantemente, actualizando la información ya procesada y aumentando así el índice de aciertos; (2) predecir la identificación del hablante en base de la información estadística y probabilística que proporcionan las fases de comparación de los padrones acústicos, considerando criterios matemáticos de estimaciones aceptables; y (3) automatizar el proceso de identificación de hablantes mediante la implementación y construcción de *software* que aplique redes neuronales.

2.0 MATERIALES Y MÉTODOS

2.1 METODOLOGÍA

Las fases metodológicas del Sistema de Identificación de hablantes comprenden las siguientes etapas (ver figura 1).

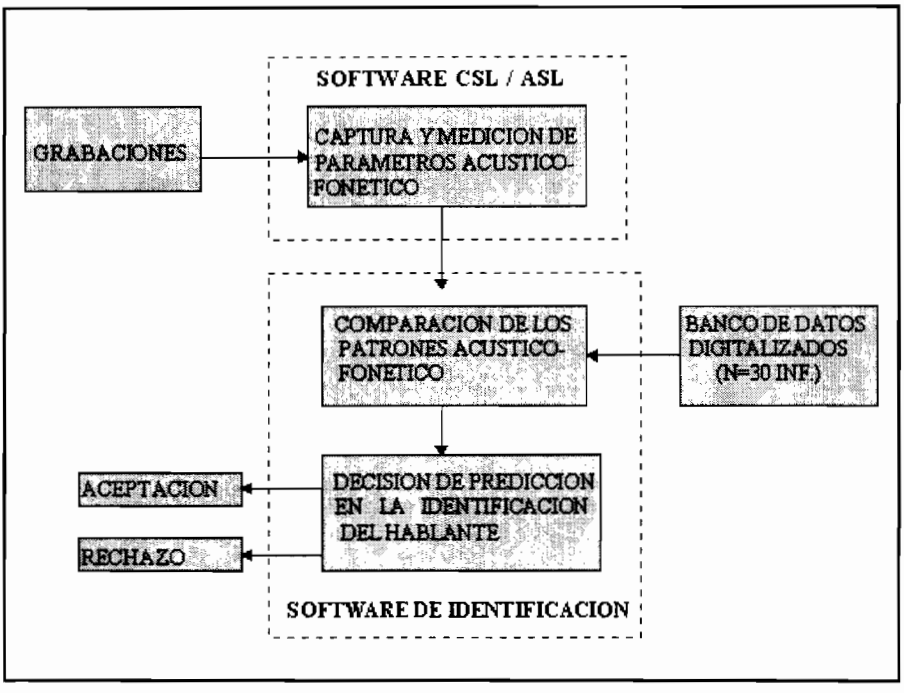


FIGURA 1: Fases metodológicas del sistema de identificación de hablantes.

Las etapas planteadas por este *Sistema* corresponden a:

- *Grabaciones*: En esta fase se registró la voz de un número representativo de hablantes, considerando los sonidos y combinaciones silábicas más relevantes del dialecto del español de Chile (ciudad de Temuco). La elaboración de un cuestionario y la lectura de los textos ad-hoc incluyeron los principales sonidos y sus combinaciones más frecuentes

en el habla local. Los ejemplos seleccionados contemplan la lectura de vocales, de sílabas, nombres de números, frases-claves para cada hablante, lectura veloz y habla espontánea.

- *Captura y medición de parámetros fonético-acústicos*: La captura y análisis de la señal digitalizada de la grabación se procesó por medio del *hardware* y *software* de la Kay Elemetrics, en forma especial, el *software* Computerized Speech Laboratory (CSL 4300B) y el Analysis-Syntesis Laboratory (ASL) que permitió medir los diversos parámetros, utilizando la técnica de predicción lineal (LPC) para obtener los peak, frecuencia fundamental, length, coeficientes de reflexión, formantes y ancho de banda.
- *Banco de datos digitalizados*: Este se construyó con la información fonético-acústica (padrones acústicos) que pertenece al corpus fonético de cada hablante.
- *Comparación de los patrones fonético-acústicos*: Esta fase tiene que ver con el *software* de Identificación de Hablantes, cuya función consiste en comparar la etapa de 'aprendizaje o conocimiento' de la RNA con los datos ingresados desde la fase 'predictiva', a fin de hacer una estimación probabilística sobre la identificación de la voz correspondiente a cada informante.

En relación a la etapa de aprendizaje, ésta tuvo que ver con la 'alimentación' de los patrones acústicos de los hablantes. Se ingresaron los resultados de las variables digitalizadas (por ejemplo, los coeficientes de reflexión, formantes, frecuencia fundamental, peak y otras) generados por el análisis del LPC, los cuales se almacenaron en un archivo con formato especial. Para el entrenamiento de la red neuronal se debió considerar un factor de aprendizaje de la red o penalidad de tipo estadístico, ya que esto permitió optimizar la identificación de los hablantes. En el proceso de identificación se consideraron, además, los resultados de los identificadores que se asignaron a cada uno de los informantes, los que fueron analizados en términos estadísticos y probabilísticos de las variables.

- *Decisión de predicción en la identificación del hablante*: El *software* utilizó los resultados estadístico y probabilístico generados en la fase anterior para aceptar o rechazar la identificación del hablante. Se debió tomar en cuenta la determinación de un umbral de decisión que se fijó como norma general para todos los informantes, de acuerdo a los resultados experimentales, con los cuales se concluyó el proceso de decisión predictiva.

3. ANÁLISIS DE LOS DATOS

3.1 Discusión de los datos

Una vez descritos los pasos fundamentales de esta investigación, presentamos ahora la discusión y análisis de la información fonético-acústica.

Los textos seleccionados, leídos y grabados en dos oportunidades por los 30 informantes, con un intervalo de seis meses, conformaron el corpus fonético. Estos textos correspondieron a secuencias de números de carácter aleatorio, distribuidos en tablas de cuatro dígitos por fila y que en total alcanzaron a 120 dígitos. Desde el punto de vista lingüístico, la combinación de los elementos que conforman los nombres de números (del 'cero' al 'nueve') corresponden a la estructura fonológica de la lengua española. En cuanto a las sílabas, hubo especial cuidado que éstas estuvieran integradas por vocales tónicas y átonas, sílabas abiertas de estructura CV, y en diptongos crecientes de estructura CD y CDC.

La información relativa a las características acústico-matemáticas de los sonidos que integran las voces (*cero, uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve*) se obtuvo a través de los programas de análisis que proporcionan el CSL y el ASL, de manera especial y por ser altamente representativa para la metodología propuesta, se consideraron solo la frecuencia fundamental *(fo)*, la intensidad *(pk)* y los coeficientes de reflexión *(k)*.

Para extraer los valores de las mediciones acústicas, el LPC se configuró de la siguiente manera:

Método de análisis	: Autocorrelación
Método de *frame*	: Sincrónico
Tipo de ventana	: Hamming
Tamaño del *frame*	: 10mseg.
Núm. de coeficientes	: 12
Ancho de banda	: 500Hz

En seguida, con la información numérica obtenida se procedió a elaborar planillas electrónicas debidamente individualizadas y construidas solo con señales periódicas, cuidando de que la frecuencia fundamental de cada *frame* no fuera igual a cero, pero cuando esta condición no se cumplió, la *fo* fue eliminada. Además, para conocer el grado de confiabilidad de los datos estadísticos de cada informante se incorporó en el *software* un indicador porcentual que permitió estimar el error asociado a ellos.

Los parámetros fonético-acústicos (*fo, pk, k*) citados más arriba, relativos a la primera grabación se utilizaron para construir la base de conocimiento de la red neuronal artificial (RNA). En cambio, para la denominada fase de predicción, también se utilizaron los resultados entregados por el LPC, pero ahora correspondientes a los nombres de los números pronunciados en la segunda grabación por cada informante. De este modo, se pudo llevar a cabo la identificación de las voces, o sea, de los nombres de cada número.

Número de Grupo	Identificación del Hablante	Aprobación del Modelo	Mediana [%]	Estimación de Error [%]
1	CH	SI	72	60
	HR	SI	97	31
	VV	SI	78	48
2	CS	SI	100	28
	LJ	SI	26	76
	RR	SI	100	13
3	IV	SI	38	73
	JV	SI	58	59
	LQ	SI	99	41
4	EA	SI	100	25
	RT	SI	89	49
	WL	SI	16	79
5	CSC	SI	67	52
	FF	SI	89	46
	LM	SI	59	64
6	EM	SI	79	47
	MB	SI	85	48
	MO	SI	98	36
7	IG	SI	7	86
	JN	SI	75	54
	JU	SI	79	51
8	EG	SI	57	65
	FP	SI	100	34
	HS	SI	75	56
9	DA	SI	68	57
	JVA	SI	36	76
	OR	SI	91	53
10	HM	SI	86	51
	NC	SI	92	42
	SC	SI	95	36

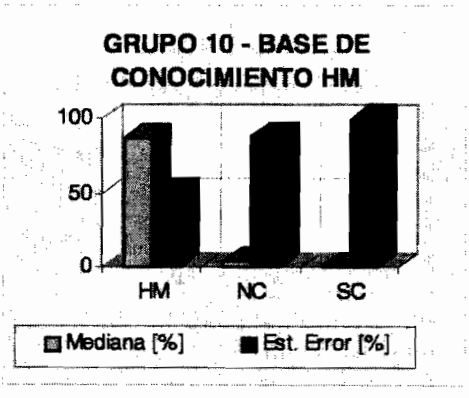
FIGURA 2: Tabla de valores de predicción.

Teniendo en cuenta la información matemática disponible se procedió a formar al azar 10 grupos de análisis integrados por tres informantes cada uno, debido a que las características del diseño del *software* facilitaba solo la comparación de grupos pequeños.

En la tabla de valores (ver figura 2) se observan los 10 grupos citados, la individualización de cada informante con sus respectivas iniciales y el resultado positivo, señalado con 'SI' de la aplicación del modelo usado en esta investigación. En las columnas finales se especifican los valores de predicción porcentual, estimándose en este caso más relevantes la mediana y la estimación de error.

De acuerdo con los resultados presentados en esta tabla, se deduce que en los 10 grupos es posible identificar a sus integrantes. También se observa que a los valores de la mediana aparece asociada la estimación de error. Por ejemplo, en los grupos 1, 6 y 10, la mediana es alta y muy significativa, está sobre un 70%; en cambio la estimación de error alcanza valores inferiores al 60%, de donde se deduce que el modelo elegido es eficaz en la estimación predictiva de los informantes, cuando los resultados de la mediana se aproximan a un 100% y la estimación de error disminuye progresivamente. En los otros grupos, el modelo también funciona a pesar del comportamiento diferente de los resultados estadísticos.

Un buen ejemplo de la aplicación del modelo comentado lo constituye el grupo 10, el cual está integrado por HM, NC, SC, cuyos valores de la mediana indican un 86%, 92%, 95%, respectivamente, y se encuentran asociados a una estimación de error equivalente a 51%, 42% y 36%. Ahora, para poder establecer la identificación de cada integrante de este grupo, se utilizó la base de conocimiento respectiva, según puede apreciarse en los gráficos siguientes. En los histogramas se muestran los valores indicados, correspondientes a la identificación de HM, NC y SC (Ver figura 3)



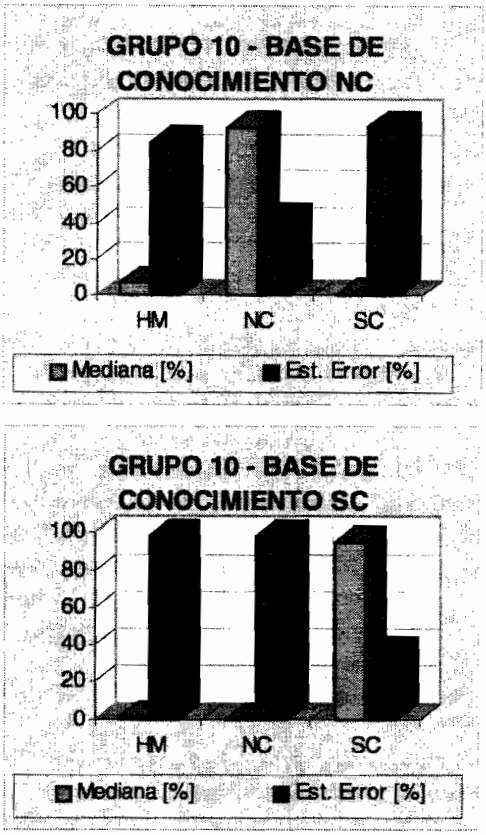


FIGURA 3: Gráficos correspondientes a cada uno de los integrantes del grupo 10

Finalmente, en la figura 4 se observan los 10 grupos seleccionados, donde cada integrante de grupo se identifica con las iniciales correspondientes a cada hablante identificado. En este histograma se aprecia también el valor porcentual de la estimación de dos mediciones estadísticas: la mediana y la estimación de error.

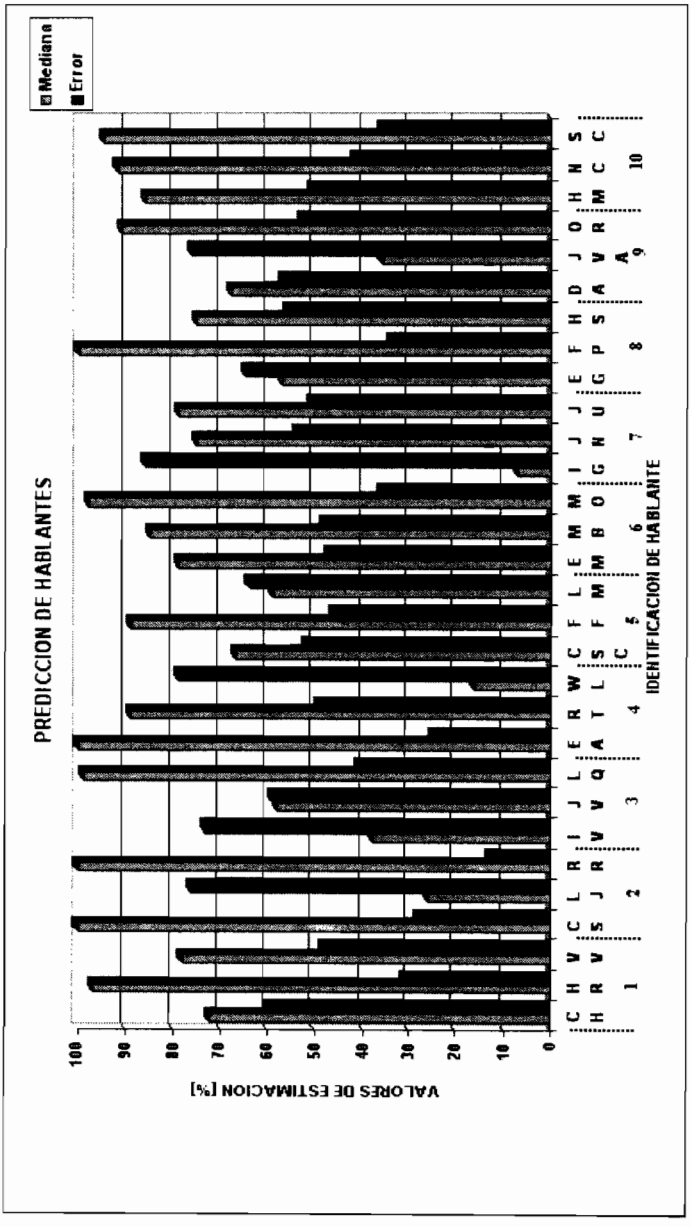


FIGURA 4: Gráfica de predicción de valores estimativos

4.0 CONCLUSIONES

4.1 La metodología y las técnicas usadas en esta investigación demostraron que el modelo propuesto fue capaz de identificar y reconocer a los 30 informantes en base a las mediciones estadísticas.

4.2 El *software* de automatización desarrollado por el equipo permitió procesar la información generada por el CSL y asistir el análisis matemático-estadístico con la red neuronal artificial (RNA) escogida para los objetivos de es trabajo.

REFERENCIAS BIBLIOGRÁFICAS

ANDERSON, J. A., 1983 "Cognitive and psychological computation with neural models". *IEEE Transactions on system, Man and Cybernetics*, N° 13, 799-815.

ATAL, B.S.,1976, "Automatic recognition of speakers from their voices", *Proc. IEEE*, 64, 4, 460-475.

BEHLAU, 1984, *Uma Análise das Vogais do Português Brasileiro falado em São Paulo: Perceptual, Espectrográfica de Formantes e Computarizada de Freqüência Fundamental*. Diss. Mestrado, Escola Paulista de Medicina.

BERNALES, *et al.*, 1998, "Desarrollo de un sistema automatizado para la verificación de hablantes". En *Por los caminos del lenguaje*, Temuco, EDUFRO, 185-194.

BRAUN, A., 1995, "Procedures and perspectives in forensic phonetics", *Proc. ICPhS* 95, Stockholm, 146-153.

BRICKER, P.D. Y PRUZANSKY, S., 1976, "Speaker recognition", en *N.J. Lass* (ed.), 295-326.

BRICKER, P.D. Y PRUZANSKY, S., 1966, "Effects of stimulus content and duration on talker identification", *JASA*, 40, 1441-1449.

BROEDERS, A.P.A., 1995, "The role of automatic speaker recognition techniques in forensic investigations". *Proc. ICPhS* 95, Stockholm, 154-161.

DODDINGTON, G.R., 1985, "Speaker recognition-Identifying people by their voices", *Proc. IEEE* 73, 1651-1664.

DOHERTY, E.T., 1975, "Evaluation of selected acoustic parameters for use in speaker identification", *JASA* 58, 107.

ENDRES, W. *et al.*, 1971, "Voice spectrograms as a function of age, voice disguise and voice imitation", *JASA* 49, 1842-1848.

FUKUSHIMA, K., 1980, "Neocognitron: A seft-organizing neural network model for a mechanism of pattern recognittion unaffected by shift in position". *Biological Cybernetics*, N° 36, 193-202.

FOULKES, 1961, "Computer identification of vowel types". *JASA* 33, 7-11

GROSSBERG, S., 1980, "How does the brain build a cognitive code?", *Psychological Review*, N° 87, 1-51.

HALL Y TOSI, 1975, "Spectrographic and aural examination of professionally mimicked voices", *JASA* 58, 107.

HECKER, M. H. L., 1971, "Speaker recognition: basic considerations and methodology", JASA 49, 138 (A).

HERTZ, JOHN, KROGH, ANDERS, PALMER, RICHARD G., 1991, "Introduction to the theory of neural computation", California, *Addison-Wesley Publishing Company*, Redwood City.

HOLLIEN, H. Y SHIPP, T., 1972, "Speaking fundamental frequency and chronologic age in males", *JSHR* 15, 155-159.

HOLLIEN, H. y MAJEWSKI, W., 1977, "Speaker identification by long-term spectra under normal and distorted speech conditions", *JASA* 62, 975-980.

HOLLIEN, H., 1995, "The future of speaker identification: a model". *Proc. ICPhS* 95, Stockholm, 138-145.

HOLLIEN, H., 1990, *The Acoustics of Crime: The New Science of Forensic Phonetics*, NY-London, Plenum Press,

HOPFIELD, John, 1984, "Neurons with graded response have collective computational properties like those of two-state neurons". Proc. Natl. Acad. Sci. USA, *Biophysics* N° 81, 3088-3092.

HOPFIELD, John, 1982, "Neural networks and physical systems with emergent collective computational abilities". Proc. Natl. Acad. Sci. USA, *Biophysics* N° 79, 2554-1558.

IOST, HANS Y RIVERA, RICARDO, 1993, *Diseño, Implementación y Prueba de una Herramienta Computacional para el manejo de Redes Neuronales, Basada en el Mecanismo de Back-Propagation y sus variaciones*. Trabajo de Título para optar al Título de Ingeniero Civil Industrial mención Informática, Universidad de la Frontera. Fac. de Ing. y Adm. Dpto. de Ing. Eléctrica, Temuco.

JOHNSON *et al.*, 1984, "Speaker identification utilizing selected temporal speech features". *J. Phon.* 12, 319-326.

KOHONEN, T., 1982, "Clustering, taxonomy, and topological maps of patterns". In M. Lang (Ed.), *Proceeding of the Sixth International Conference on Pattern Recognition, Silver Spring,* MD: IEEE Computer Society Press, 1982, pp. 114-125.

KUWABARA Y TAKAGI, 1991, "Acoustic parameters of voice individuality and voice quality control by analysis-synthesis method". *Speech Communication* 10, 491-495

KÜNZEL, H. J., 1990, *Phonetische Untersuchungen zur Sprecher-Erkennung durch Linguistisch Naive Personen*, Stuttgart, Franz Steiner Verlag.

LASS, N. J. Y DAVIS, M., 1976, "An investigation of speaker height and weight identification", *JASA* 59, 700-703.

LASS, N. J. *et al.*, 1980, "A comparative study of speaker height and weight and weight identification from voiced and whispered speech", *J. Phonetics* 8, 195-204.

LAVER, J. Y TRUDGILL, 1979, "Phonetic and lingüistic markers in speech". En Scherer, K.R. y Giles, H. (eds), 1-32.

LINDBLOM, 1963, "Spectrographic study of vowel reduction". *JASA* 35, 1773-1781.

LLISTERRI, J., 1996, "Los sonidos del habla", en Carlos Martín Vide (ed.), *Elementos de Lingüística*, 65-128.

LLISTERRI, J., 1991, *Introducción a la fonética: El método experimental,* Barcelona, Anthropos.

MC CULLOCH, W. S. & PITTS, W., 1943, "A logical calculus of ideas immanent in nervous activity". *Bulley of Mathematical Biophysics*, N° 5, 115-133.

MILLER, B. & REINHARDT, J., 1990, *Neural Network, An Introduction*, Springer-Verlag, Berlin Heidelberg, Germany.

MOLINA, R., 1994, *Identificação de falantes. Aspectos Teóricos e Metodológicos*. Tesis de Doutorado. Univ. de Campinas.

NOLAN, 1983, *The Phonetic Bases of Speaker Recognition*. Cambridge

POLLACK, Y. *et al.*, 1954, "On the identification of speakers by voice", *JASA* 26, 403-406.

PETERSON Y BARNEY, 1952, "Control methods in a study of the vowels". *JASA* 24, 175-184.

ROSENBERG, A. E., 1976, "Automatic speaker verification: a review", *Proc. IEEE* 64, 4, 475-487.

WU, K. Y CHILDERS, D. G., 1991, "Gender recognition from speech. Part. I: coarse analysis", *JASA* 90, 1828-1840.