VARIABILIDAD INTRA- E INTER-HABLANTE DE LA FRICATIVA SIBILANTE /s/ EN EL ESPAÑOL DE ARGENTINA

INTRA- AND INTER-SPEAKER VARIABILITY OF SIBILANT FRICATIVE /s/ IN ARGENTINE SPANISH

PEDRO UNIVASO
LIS, INIGEM, CONICET-UBA
punivaso@yahoo.com.ar

MIGUEL MARTÍNEZ SOLER
LIS, INIGEM, CONICET-UBA y Universidad Austral
miguelmsoler@gmail.com

JORGE A. GURLEKIAN
LIS, INIGEM, CONICET-UBA
jag@fmed.uba.ar

Artículo recibido el día: 23/10/2013
Artículo aceptado definitivamente el día: 29/04/2014
Estudios de Fonética Experimental, ISSN 1575-5533, XXIII, 2014, pp. 95-124

RESUMEN

En este trabajo se propone estudiar y valorizar el poder discriminativo de la fricativa sibilante /s/, de manera de incorporar este conocimiento en futuros sistemas automáticos de reconocimiento de hablantes. Se seleccionó esta fricativa por ser la principal consonante de acuerdo a la frecuencia de aparición en el corpus. Se determinó un ranking de los parámetros acústicos de dicho fonema que mejor discriminan a un hablante, teniendo en cuenta la menor variabilidad intra-hablante y la máxima variabilidad inter-hablante. El material de evaluación fue extraído de la base de datos SpeechDat, con muestras de habla de telefonía fija en Español de Argentina. Los parámetros con mejor puntaje fueron: la Intensidad, el tercer formante (F3), el primer formante (F1) y el primer momento espectral o Centro de Gravedad (CG). El poder discriminante de la fricativa sibilante /s/, con respecto al resto de los fonemas, ha quedado corroborado por su importante aporte a la tasa de reconocimiento de hablantes obtenida, siendo el sexto fonema en importancia después de las vocales /e/, /a/, /o/ y /i/, y la nasal /n/. La tasa de igual error, empleando solamente este fonema, resultó un 35% menor que la media del total de los 30 fonemas involucrados.

Palabras clave: *fricativas sibilantes, reconocimiento de hablantes, parámetros acústicos, variabilidad intra-hablante, variabilidad inter-hablante.*

ABSTRACT

This paper focuses on the analysis of the discriminative power of the sibilant fricative /s/, in order to incorporate this knowledge in future automatic speaker recognition systems. The selected fricative is the most frequent consonant in the corpus. An acoustical parameter ranking of /s/ was performed based on minor intra-speaker variability and maximun inter-speaker variability. Evaluation is performed on Argentine-Spanish voice samples from the SpeechDat database recorded on a fixed phone environment. The intensity, the third formant (F3), the first formant (F1) and the first spectral moment or Center of Gravity (CG) were the best ranked parameters. The sibilant fricative /s/, considered in isolation, has a speaker recognition equal error rate (EER) of 35% lower than the average of the total of 30 phonemes involved, confirming the importance of this phoneme for the discrimination of speakers as the sixth phoneme in importance, preceded by the vowels /e/, /a/, /o/ and /i/, and the nasal /n/.

Keywords: *sibilant fricatives, speaker recognition, acoustic parameters, intra-speaker variability, inter-speaker variability.*

1. INTRODUCCIÓN

Resulta de interés la incorporación de conocimiento fonético en los actuales sistemas de reconocimiento automático de hablantes. Éstos son desarrollados a partir de una combinación de algoritmos empíricos que son evaluados fundamentalmente por los resultados, expresados como tasas de reconocimiento estadísticas. En su desarrollo, los diferentes investigadores emplean ciencias tan diversas como la biomedicina, la matemática, la estadística y la computación, con la finalidad de desarrollar sistemas que logren tasas de reconocimiento de hablantes superadoras. En este camino hacia la eficiencia, es poco frecuente encontrar sistemas de reconocimiento automático de hablantes que basen sus principios de funcionamiento en conceptos extraídos de la fonología o la fonética.

El área del «reconocimiento del hablante» tiene como objetivo la determinación de una persona a partir de su voz, pudiéndosela dividir en dos sub-áreas: la «identificación del hablante» y la «verificación del hablante». En la aproximación tradicional, tanto la verificación como la identificación requieren que el hablante emita frases de prueba, luego, que se hagan mediciones características de esas frases de prueba y se computen una o varias funciones de distancia entre el vector de mediciones y el vector de referencia almacenado. En términos de procesamiento de la señal, los métodos son similares. La principal diferencia radica en la lógica de decisión y los parámetros utilizados para medir las distancias.

En la verificación se pretende determinar si un hablante es quien dice ser mediante su voz, o detectarlo en una conversación, estableciendo si un segmento de habla fue emitido por él o no. En la verificación, la respuesta del sistema es binaria: acepta o rechaza la identidad del hablante, haciendo una sola comparación y utilizando un umbral que pesa el costo de aceptar un impostor o rechazar un hablante verdadero.

La identificación realiza la comparación de los rasgos de un hablante incógnita con un número N de hablantes. Se elige el hablante con la mínima probabilidad de error. La probabilidad de error tiende a uno en la medida que el número de hablantes, con que debe compararse, aumenta. Al aumentar el número de hablantes crecen las probabilidades de que dos o más hablantes tengan distribuciones muy cercanas unas a otras. En esas circunstancias, la identificación es una tarea difícil de resolver.

En años recientes, los sistemas de reconocimiento del hablante basados en Modelos de Mezclas de Gaussianas (GMM) han generado los mejores resultados. Cada hablante es representado por un modelo probabilístico genérico de densidades multivariadas, capaz de representar densidades arbitrarias, lo cual lo hace factible de ser empleado en aplicaciones independientes del texto. Asimismo, son modelos estadísticos conocidos, computacionalmente económicos e insensibles a las características temporales, donde se modelan las características subyacentes dependientes del hablante a partir de datos acústicos segmentales. Este tipo de modelado no permite discriminar los diferentes componentes del habla a reconocer, con lo cual no son susceptibles de ser empleados directamente para la incorporación de parámetros basados en el conocimiento fonético. Y por lo tanto, no cumple con los requisitos planteados en el presente trabajo.

En cambio, los sistemas de reconocimiento de hablantes basados en modelos ocultos de Markov (HMM), que emplean fonemas como unidades básicas del modelado acústico, tienen la posibilidad de incorporar factores proporcionales para cada modelo de acuerdo al poder discriminante que posea cada fonema. Debido al empleo de modelos de fonemas, estos sistemas son dependientes del texto, lo cual restringe su empleo.

En el caso de los sistemas de identificación de hablantes empleados en el ámbito forense, dicha restricción no es insalvable. La identificación generalmente parte de la grabación de una voz relacionada con un hecho delictivo (grabación dubitada, prueba o evidencia), la cual es comparada con otros registros atribuidos a una persona, normalmente conocida (grabación indubitada o del sospechoso), con lo cual se conocen, previamente a la identificación, las palabras presentes en la prueba y las emitidas por el sospechoso. Generalmente dichas grabaciones provienen de comunicaciones telefónicas de redes fijas y/o móviles.

También pueden emplearse en forma conjunta con sistemas de reconocimiento automático de habla (ASR), que permiten recomponer la secuencia de fonemas emitidos por el hablante en forma previa a la identificación del hablante. En estos casos la tasa de reconocimiento de las palabras emitidas introduce un factor adicional en el error de identificación de hablantes.

Podemos enmarcar la identificación de hablantes en el ámbito forense dentro de la fonética forense (Rose, 2002:19), y más ampliamente, dentro de la lingüística forense (Gibbons *et al*, 2008). Las áreas relacionadas con la fonética forense, según Rose (2002:19), son las siguientes: la identificación de hablantes –que es la abarcaremos en este trabajo-; la determinación de perfiles de hablantes (ante la

falta de un sospechoso, dar información sobre su acento socioeconómico); la construcción de ruedas de voces (para el reconocimiento de voces por parte de testigos o víctimas); la identificación de contenido (determinando lo que fue dicho cuando la grabación es de mala calidad, o cuando la voz es patológica o tiene un acento extranjero); y la autenticación de los registros de audio (determinando si una grabación ha sido manipulada).

1.1. Fundamentos de la selección de /s/ como variable de estudio

El presente trabajo se restringirá a un único fonema, la consonante fricativa sibilante /s/ en la secuencia VCV.

Dicha selección tuvo en cuenta que /s/ es la consonante más frecuente (figura 1) en el corpus empleado, con un 8,9% de ocurrencia, y la secuencia VCV la más frecuente, entre las que involucran a dicho fonema (figura 2), con un porcentaje del 37,6%.

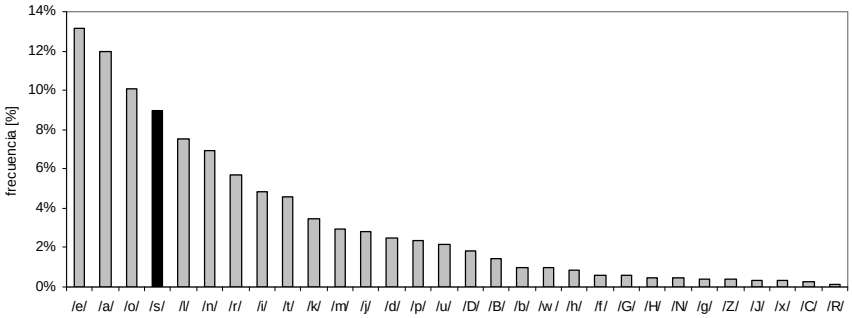

Figura 1. *Frecuencia de los fonemas en el corpus empleado. La nomenclatura empleada utiliza el alfabeto SAMPA adaptado a la Argentina (Gurlekian, Colantoni y Torres, 2001).*

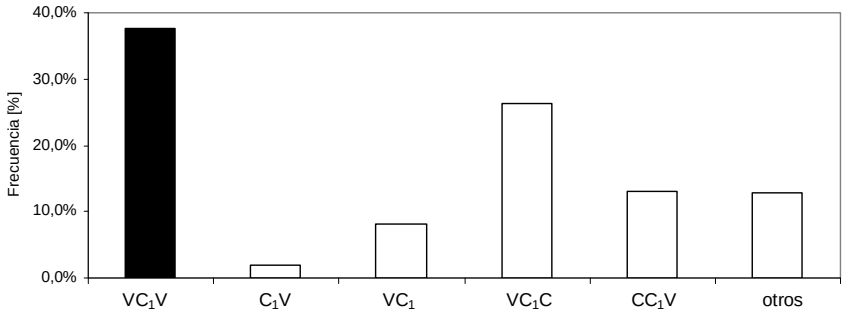

Figura 2. *Frecuencia de aparición en el corpus de las secuencias que incluyen al fonema /s/, representado en el gráfico por C_1. Con fondo negro, la secuencia VC_1V empleada en el presente trabajo.*

1.2. Objetivo: valorización del poder discriminativo de la /s/

El objetivo del presente trabajo es la valorización del poder discriminativo de la /s/, para incorporarlo a un sistema HMM que pueda aprovechar las características fonéticas distintivas de cada modelo. Dicha valorización se efectivizará por medio de la determinación de un listado de los parámetros acústicos de la /s/ que mejor discriminan a un hablante. De esta manera, los parámetros de la lista con mayor puntaje podrán ser incorporados posteriormente al sistema automático de reconocimiento de hablantes para mejorar su performance. Los actuales sistemas generalmente no discriminan entre fonemas y aquellos que lo hacen emplean los mismos parámetros para todos los fonemas.

1.3. Hipótesis: rasgos discriminantes y rasgos distintivos

La hipótesis general del presente trabajo es que los parámetros que mejor discriminan a un hablante del resto son aquellos que dependen intrínsecamente de las características físicas particulares de cada hablante. La característica discriminante de estos parámetros los hace poseedores de lo que denominaremos «rasgos discriminantes». Dentro de la hipótesis general, proponemos que los mismos posean la máxima variabilidad inter-hablante y la menor variabilidad intra-hablante. Para ello se creará el «Coeficiente discriminante [Cd]», determinado como el cociente entre los coeficientes de variación de Pearson inter-

e intra-hablante. Como primera hipótesis particular se propone el tercer formante (F3) como el parámetro acústico más relevante para la discriminación de hablantes dada su mejor asociación con la longitud del tracto vocal, la cual es la más común de las diferencias físicas entre hablantes (Stevens, 1972; Nordström y Lindblom, 1975; Weirich, 2010). Y como segunda hipótesis particular se considerará a la intensidad sin normalizar, la cual debería proporcionar una variación importante entre hablantes, de acuerdo a lo estudiado por Behrens y Blumstein (1988).

Los parámetros iniciales empleados serán los que representan los rasgos distintivos de las fricativas, de los cuales surgirán aquellos que posean rasgos discriminantes. Como señala Cuadrado (1995:42), los fonemas se hallan constituidos por un conjunto de rasgos distintivos, es decir, de señales fónicas complejas capaces *de cambiar de un fonema en otro por sustitución y, como consecuencia, de originar transformaciones significativas* (Delattre, 1967:178-179). Este procedimiento supone que los rasgos discriminantes están incluidos dentro de los rasgos distintivos o, lo que es lo mismo, que algunos parámetros que nos permiten diferenciar un fonema de otro son los mismos que también nos ayudan a discriminar hablantes.

El resto del presente trabajo está organizado de la siguiente manera: en la Sección 2 se realizará una revisión de la literatura referida a la variabilidad y a la fricativa /s/; en la Sección 3, se presenta la metodología empleada; en la Sección 4, los resultados obtenidos; en la Sección 5, la discusión de dichos resultados y en la Sección 6, las conclusiones del presente trabajo.

2. VARIABILIDAD DE LA FRICATIVA SIBILANTE /s/

Los primeros estudios de las fricativas se remontan a 1956 cuando Hughes y Halle (1956) señalaron, por primera vez, algunas de las características espectrales de dichas consonantes. Posteriormente, se han publicado diversos trabajos experimentales sobre las distintas lenguas, entre los que sobresalen Strevens (1960), Heinz (1961), Heinz y Stevens (1961), Jassem (1965 y 1968), Behrens y Blumstein (1988). En dichos trabajos se analizan diferentes parámetros acústicos que diferencian a las fricativas o grupos de fricativas entre sí, pero ninguno analiza conjuntamente la variabilidad intra- e inter-hablante de las mismas.

La investigación para el español no es muy numerosa y, en general, se limita a ofrecer descripciones globales de lo que se percibe en el espectro. No obstante, en

algunos casos, se señalan las frecuencias de los picos espectrales, aunque no se detalla su relación en términos de amplitud. Los sonidos fricativos que se encuentran en el Español de Argentina son las sibilantes [s], [ʃ], [z], [ʒ] y las no sibilantes [f], [θ], [x], [h]. La fricativa [s] es del tipo sibilante sorda, aunque en algunos casos particulares puede sonorizarse, siendo su punto de articulación el alveolar apical de acuerdo al alfabeto fonético internacional (IPA). Los trabajos de Borzone de Manrique (1980) y Borzone de Manrique y Massone (1979 y 1981) examinan las propiedades acústicas de las fricativas del Español de Argentina y los picos espectrales, que permiten distinguir estas consonantes entre sí, encontrándose, que para la [s], dichos picos se encuentran entre los 5.000 y los 8.000 Hz. La incidencia de los parámetros acústicos en la percepción de /s/ y /f/ fue estudiada por Gurlekian (1981) mediante la síntesis de sus rasgos distintivos.

Los principales parámetros acústicos descriptos en la literatura, para clasificar las fricativas, son: la duración (Baum y Blumstein, 1987), la amplitud (Strevens, 1960; Behrens y Blumstein, 1988), y las características espectrales (Strevens, 1960; Forrest *et al*, 1988: Borzone de Manrique y Massone, 1981; Gurlekian, 1981). Las principales características espectrales estudiadas fueron la ubicación de los picos espectrales (Strevens, 1960) y los momentos espectrales de primero a cuarto orden (Forrest *et al*, 1988; Flipsen *et al*, 1999; Jongman *et al*, 2000; Tabain, 2001): centro de gravedad, desviación estándar, asimetría y curtosis. Para el caso de las fricativas sibilantes también se emplearon los formantes como parámetros (Toda *et al*, 2009).

La mayoría de la literatura sobre reconocimiento automático de hablantes se centra en el aporte de ciencias como la biomedicina, la matemática, la estadística y la computación, como lo muestra, por ejemplo un reconocido tutorial sobre el tema (Campbell, 1997). Aunque la fonética se encuentra mayormente marginada, los primeros trabajos emplearon parámetros basados en los espectros de predicción lineal de las fricativas (Sambur, 1975). En un trabajo actual para el francés (Kahn *et al*, 2010) se complementan los resultados obtenidos por otras técnicas con información fonéticamente relevante, a partir de parámetros LFCC, Delta y Aceleración, extractados de cada fonema. En dicho trabajo, se propone continuar trabajando para determinar cuál tipo de información acústica es la que induce los mejores resultados en la verificación de hablantes, y qué tipo de información es la que la degrada. Otra investigación para el español, en esta línea de trabajo, analiza la variación intra-hablante (Marrero *et al*, 2003), por medio de sendos análisis acústicos y perceptuales de diferentes fonemas. Dichos autores seleccionaron la [s] para analizar las fricativas, habiendo empleado como parámetros: la máxima intensidad y la frecuencia del espectro al comienzo de la turbulencia. A manera de

conclusión y para el caso de la sibilante en cuestión se hace notar que la frecuencia de la turbulencia acompaña el F2 de la vocal contextual correspondiente. Como señalan Carney y Moll (1971), en el segmento fricativo encontraremos información del hablante emisor, transmitida por la vocal contextual, además de la inherente a la fricación en sí misma. Un trabajo reciente evalúa la capacidad de los sonidos fricativos sordos del español para discriminar hablantes en fonética forense, empleando los momentos espectrales, el pico espectral de mayor intensidad y los valores estandarizados de las bandas LTAS (Cicres, 2011:37-41).

Con respecto a la importancia de la fricativa sibilante /s/ en el estudio de la discriminación de hablantes, Magrin-Chagnolleau *et al*. (1995) habían mostrado que algunos segmentos fonéticos eran más eficientes que otros para modelar a un hablante. Posteriormente Kahn *et al*. (2010) demostraron que, para el francés, el principal fonema discriminador era la fricativa sibilante /s/, en experimentos realizados con hablantes de ambos géneros.

3. METODOLOGÍA

El corpus empleado, forma parte del proyecto SALA (*SpeechDat Across Latin America*) (Moreno *et al*, 2000), y sigue las definiciones establecidas para la creación de base de datos de habla conversacional encontrada en las comunicaciones de redes de telefonía fija (Winsky, 1997:22). El subconjunto correspondiente al español de Argentina (Gurlekian *et al*, 2001) está constituido por cinco regiones distribuidas en todo el país. El estilo de habla corresponde a párrafos leídos, extraídos de diarios y libros de la Argentina o elaborados por lingüistas. Las grabaciones fueron realizadas por sujetos de ambos sexos y distintos grupos de edades, y efectuadas a través de la red de telefonía fija hacia una computadora equipada con una plaqueta AVM-ISDN-A1 y una interfaz de acceso básico a ISDN (BRI).

El empleo del canal telefónico para la realización de las grabaciones incorporó un factor generalmente presente en las tareas forenses. En cambio, el estilo de habla leída, aunque no comparable con el habla espontánea (propia de dichas tareas), permitió desarrollar un corpus de habla controlada que posibilitó la realización del presente trabajo.

La frecuencia de muestreo empleada fue de 8 KHz / 16 bits quedando un ancho de banda útil de 4.000 Hz, en concordancia con los límites de frecuencia establecidos por el canal telefónico. Esta frecuencia es inferior a la de los picos espectrales de

[s], con lo cual no se podrá hacer uso de los mismos como parámetros discriminantes; así como tampoco los emplea el sistema perceptual en las comunicaciones telefónicas habituales. En estos casos, la información remanente es suficiente para identificar a un hablante. Cabe hipotetizar que la concentración de energía en el rango de 5.000 a 8.000 Hz es un rasgo distintivo de la fricativa y la distribución de energía en el resto de las bandas posee características particulares de los hablantes.

Para este trabajo se seleccionaron las emisiones de los hablantes masculinos de la región SUR. Esta región comprende las provincias de Buenos Aires, Santa Fe, Entre Ríos, La Pampa, Neuquén, Río Negro, Chubut, Santa Cruz y Tierra del Fuego. La región SUR es la más populosa de Argentina con un número aproximado de 21 millones de habitantes, la cual corresponde al 65% del total del país y forma parte de una de las divisiones dialectales propuestas por Vidal de Battini (1964).

Durante la selección de las emisiones acústicas, se eliminaron aquellas que presentaban alteraciones groseras (como baja relación señal a ruido y errores de emisión). El corpus quedó conformado por 239 emisiones de 47 hombres entre 16 y 45 años conteniendo la secuencia VCV, dónde C es el fonema /s/, incluidas las variantes de acento léxico en las vocales contextuales, p ej. ***peso****s* ['pe sos], *as****eso****r* [a se 'sor], ***eso****térico* [e so 'te ri ko]. La base de datos intra-hablante quedó compuesta por 48 emisiones de 6 hablantes, es decir, un promedio de 8 emisiones por hablante.

Se empleó un sistema de reconocimiento automático de habla, basado en Modelos Ocultos de Markov (HMM), para etiquetar los segmentos fricativos dentro de las emisiones de cada hablante. Para mejorar la performance del etiquetado, se empleó la metodología de alineamiento forzado, y posteriormente se verificó manualmente la precisión del etiquetado para algunos casos (figura 3). El cálculo de los parámetros distintivos se realizó por medio del software de análisis y síntesis de señales de habla PRATT (Boersma *et al*, 2005), empleándose para todas las mediciones la configuración estándar. Para todas las mediciones, excepto para la duración, se empleó una ventana de 25 milisegundos de ancho, centrada en el segmento fricativo seleccionado.

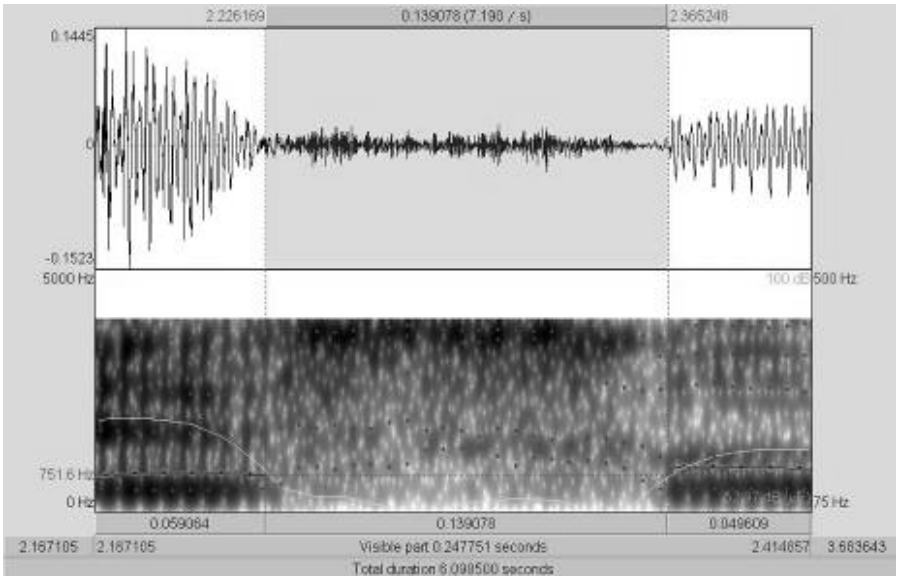

Figura 3. *Selección del segmento fricativo [s] de la secuencia VCV. A la izquierda y derecha pueden visualizarse los segmentos sonoros de las vocales contextuales delimitantes.*

En base a los diferentes parámetros empleados en la literatura para representar a las fricativas sibilantes, se seleccionaron algunos que contienen información relevante de los formantes, el espectro, la energía y la duración del segmento fricativo. Los parámetros que se emplearon fueron los siguientes: Intensidad; Duración; los momentos espectrales Centro de Gravedad (CG), Desviación Estándar, Asimetría y Curtosis; y los formantes F1, F2, F3, F4 junto a los parámetros relacionados F2-F1 y F4-F3. A modo de ejemplo, en la tabla 1 se muestran los resultados de las mediciones de dichos parámetros para la emisión de un hablante.

Emisión	s1sh1-002-7
F1 [Hz]	1296
F2 [Hz]	1687
F3 [Hz]	2728
F4 [Hz]	3536
F2-F1 [Hz]	391
F4-F3 [Hz]	808
Intensidad [dB]	51
Duración [mseg]	130
CG [Hz]	3229
Desv.Est. [Hz]	839
Asimetría [-]	-1.8
Curtosis [-]	2.4

Tabla 1. *Ejemplo de parámetros calculados dentro del segmento fricativo [s] para la emisión número 7 del hablante s1sh1-002.*

En la figura 4 se puede ver el espectro FFT resultante de la aplicación de la transformada rápida de Fourier a la señal temporal, y en la figura 5, los formantes calculados a partir del cómputo de los coeficientes LPC. En la figura 6 se visualiza la variabilidad del espectro LPC inter- e intra-hablante para dos hablantes del corpus.

En la figura 7 se muestran las distribuciones de las emisiones inter-hablantes y de tres emisiones intra- hablantes en un plano de coordenadas F1 y F3. Se puede ver cómo las variaciones intra-hablantes en ambas coordenadas son menores a las correspondientes a las inter-hablantes, lo cual corroboraría la hipótesis de que dichos parámetros son un aporte para la discriminación de hablantes.

En la tabla 2 se muestran los valores que toman todos los parámetros seleccionados en el presente trabajo para las bases de datos intra- inter-hablante.

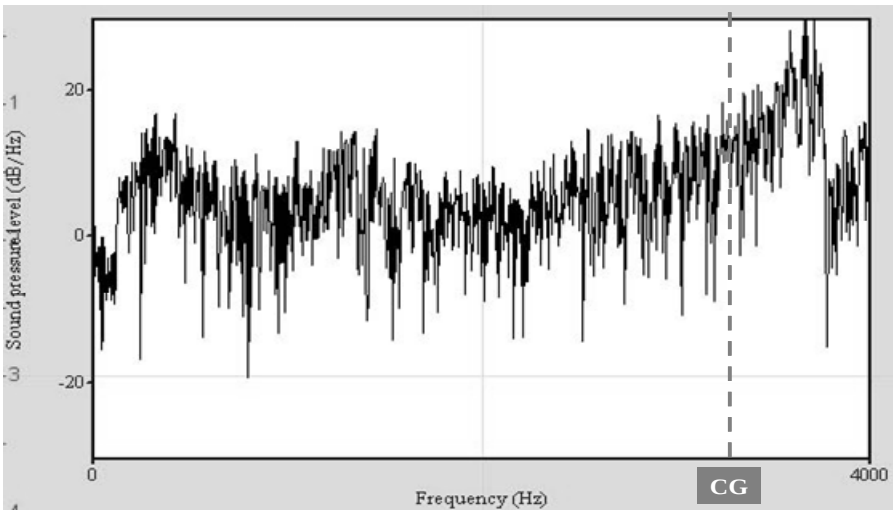

Figura 4. *Espectro FFT correspondiente a la emisión s1sh1-002-7, a partir del cual se calculan los momentos espectrales.*

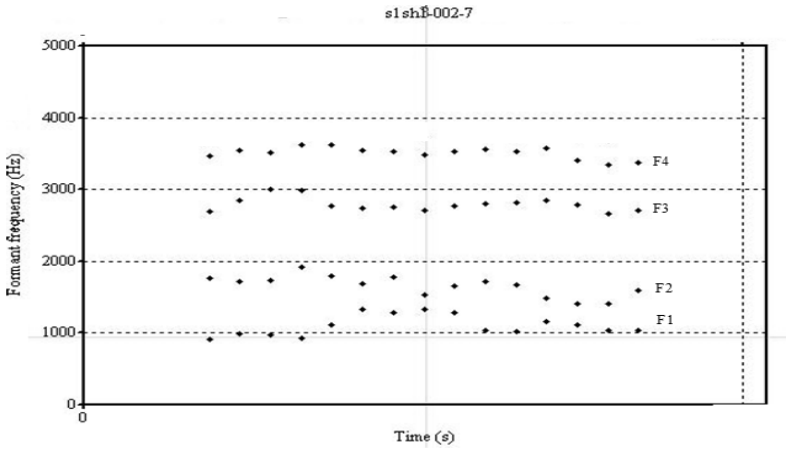

Figura 5. *Variación temporal de la frecuencia de los formantes F1 a F4 correspondientes a la emisión s1sh1-002-7. En el presente trabajo solo se emplean los formantes señalados en el centro de la emisión.*

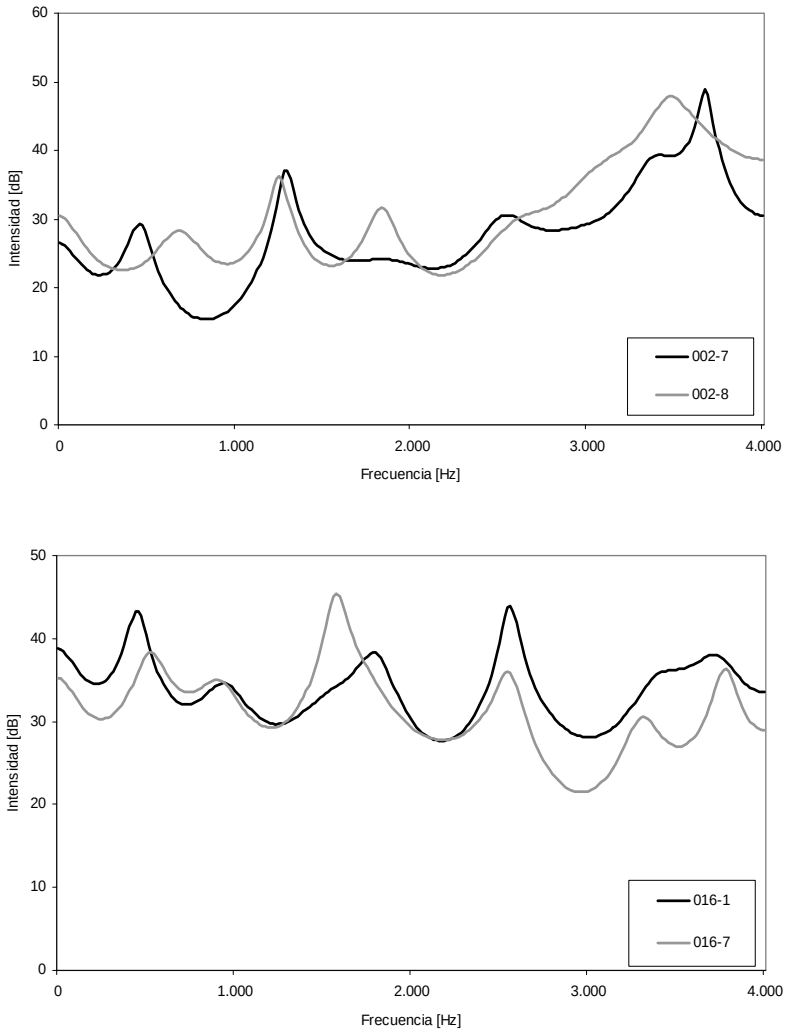

Figura 6. *Espectros LPC de cuatro emisiones, correspondientes a dos hablantes empleados en el presente trabajo. Los códigos iniciales de cada emisión corresponden al número de hablante y los segundos al número de emisión correspondiente.*

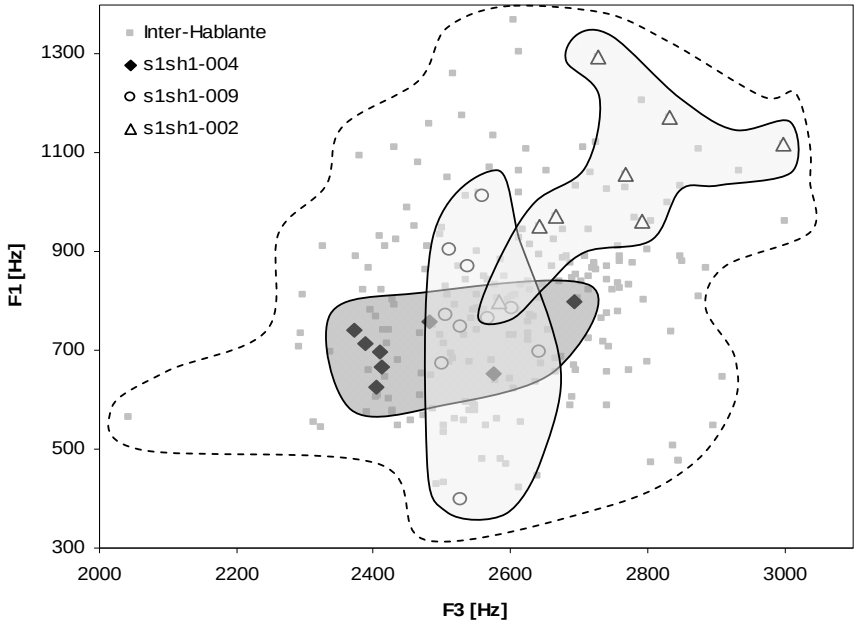

Figura 7. *Distribución en el plano F1-F3 de emisiones inter-hablantes y emisiones intra-hablantes correspondientes a los hablantes s1sh1-004, 9 y 2.*

Parámetro	Intra-hablante						Inter-hablante
	s1sh1-001	s1sh1-002	s1sh1-003	s1sh1-004	s1sh1-009	s1sh2-006	
F1 [Hz]	898 ± 47	1041 ± 155	783 ± 197	707 ± 58	731 ± 76	698 ± 125	796 ± 181
F2 [Hz]	1861 ± 89	1804 ± 138	1854 ± 71	1606 ± 125	1727 ± 63	1728 ± 59	1791 ± 142
F3 [Hz]	2703 ± 72	2751 ± 130	2847 ± 57	2467 ± 112	2495 ± 89	2533 ± 66	2596 ± 147
F4 [Hz]	3249 ± 136	3415 ± 93	3313 ± 127	3179 ± 153	3219 ± 128	3247 ± 128	3248 ± 155
F2-F1 [Hz]	963 ± 110	763 ± 174	1071 ± 204	899 ± 161	996 ± 89	1029 ± 135	995 ± 188
F4-F3 [Hz]	547 ± 143	663 ± 106	466 ± 117	712 ± 199	725 ± 115	714 ± 120	651 ± 156
Int. [dB]	50 ± 4	52 ± 3	62 ± 4	50 ± 3	39 ± 3	43 ± 5	52 ± 9
Dur. [mseg]	0,11 ± 0,02	0,12 ± 0,02	0,1 ± 0,03	0,1 ± 0,01	0,1 ± 0,02	0,09 ± 0,02	0,11 ± 0,03
CG [Hz]	2418 ± 380	2916 ± 596	2843 ± 496	1598 ± 421	1768 ± 417	1706 ± 453	2188 ± 718
D.E. [Hz]	785 ± 236	773 ± 255	774 ± 328	974 ± 194	999 ± 245	1062 ± 203	892 ± 262
Asim. [-]	-0,82 ± 1,34	-1,99 ± 1,62	-2,17 ± 1,28	0,45 ± 0,74	0,03 ± 0,55	0,14 ± 0,74	-0,58 ± 1,4
Curtosis [-]	4,7 ± 13,4	6,3 ± 8,6	7,4 ± 8,5	-0,2 ± 1,2	-0,6 ± 0,7	-0,8 ± 0,9	2,1 ± 5,6

Tabla 2. *Valores promedio y desviación estándar de los parámetros empleados en le presente trabajo, considerando las emisiones inter-hablante y las emisiones de cada uno de los hablantes de la base de datos intra-hablante.*

4. RESULTADOS

4.1. Coeficiente discriminante

De manera de poder evaluar la discriminación entre hablantes por medio de los diferentes parámetros estudiados, se definió el Coeficiente discriminante (Cd), teniendo en cuenta que un parámetro con alto poder de discriminación debe presentar una baja variabilidad intra-hablante y una alta variabilidad inter-hablante. La definición clásica utilizada en estadística para el Coeficiente de Variación de Pearson (CV) es la siguiente:

$$(1)\ CV = \text{Coef. de Variación de Pearson} = \frac{\text{Desviación Estándar}}{\left|\text{Valor medio}\right|}$$

Y si se considera que, el Coeficiente de Variación de Pearson corresponde a una distribución de baja dispersión cuando se cumple que:

$$(2)\ CV < 1$$

Entonces, para esos casos, definiremos el Coeficiente discriminante (Cd) como:

$$(3)\ Cd\,[\%] = \frac{CV\,\text{inter} - CV\,\text{intra}}{CV\,\text{inter}} \times 100$$

Donde:

$$(4)\ CV\ \text{inter} = \frac{\text{Desviación Estándar interhablante}}{\text{Valor medio interhablante}}$$

$$(5)\ CV\ \text{intra} = \frac{\text{Desviación Estándar intrahablante}}{\text{Valor medio intrahablante}}$$

En la tabla 3 se muestran los resultados de las mediciones intra-hablante e inter-hablante correspondientes a los diferentes parámetros seleccionados. Los parámetros Asimetría y Curtosis no pueden brindar información fehaciente de su potencial

discriminador, en base al cálculo del Coeficiente discriminador, dado que sus Coeficientes de Variación de Pearson son mayores que la unidad.

Parámetros	Coef. de Variación de Pearson		Cd [%]
	CV inter	CV intra	
Intensidad	0.17	0.06	65%
F3	0.06	0.04	38%
CG	0.33	0.21	37%
F1	0.23	0.15	32%
F2	0.08	0.06	21%
F4	0.05	0.04	19%
F2-F1	0.19	0.18	3%
Desv.Est.	0.30	0.30	-1%
Duración	0.24	0.25	-3%
F4-F3	0.24	0.25	-5%
Asimetría	2.39	--	--
Curtosis	2.69	--	--

Tabla 3. *Coeficientes de Variación de Pearson inter-hablante CV inter e intra-hablante CV intra y Coeficientes discriminantes Cd [%].*

Para que un parámetro pueda considerarse efectivo para discriminar hablantes debe cumplir que su Coeficiente discriminante (Cd) sea elevado, lo que significa que debe poseer no solo una baja variabilidad intra-hablante, sino que debe estar acompañada de una alta variabilidad inter-hablante. En la figura 8 se muestran los Cd correspondientes a cada parámetro, pudiéndose ver que los parámetros Intensidad, F3 y CG son los que, según esta metodología, mejor discriminan hablantes. En la tabla 4 se presentan los mismos resultados en forma de ranking. Por otra parte, los parámetros Desviación Estándar, Duración y F4-F3, que poseen valores negativos de Cd, no estarían aportando a la discriminación de hablantes.

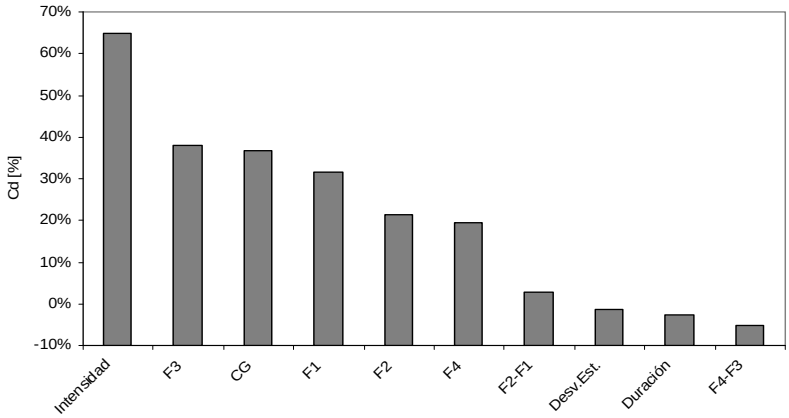

Figura 8. *Coeficientes discriminantes (Cd) para cada uno de los parámetros calculados.*

Parámetros	Cd [%]	Ranking 1
Intensidad	65%	1
F3	38%	2
CG	37%	3
F1	32%	4
F2	21%	5
F4	19%	6
F2-F1	3%	7
Desv.Est.	-1%	8
Duración	-3%	9
F4-F3	-5%	10
Asimetría	76%	--
Curtosis	60%	--

Tabla 4. *Ranking de efectividad en la discriminación de hablantes en base a los Coeficientes discriminantes (Cd) de cada parámetro.*

4.2. Algoritmo de identificación de hablantes cíclico

De manera de contrastar los resultados obtenidos a través del cálculo de los Coeficientes discriminantes, se realizó un experimento adicional, empleando un algoritmo de identificación de hablantes cíclico, que mide la efectividad de cada parámetro de acuerdo al resultado de la tasa de reconocimiento obtenida. Para ello se calculó la distancia euclídea n-dimensional (con *n* igual a la cantidad de parámetros empleados) entre las emisiones de todos los hablantes de la base inter-hablante con respecto al promedio de las emisiones de cada uno de los hablantes de la base intra-hablante. En la tabla 5 se pueden ver los resultados individuales, medidos por medio de la tasa de igual error (EER), que es la tasa de error en la que el porcentaje de falsas alarmas es igual al porcentaje de casos perdidos. En la misma tabla se ha incorporado el ranking de parámetros, de acuerdo a estos resultados.

Parámetros	EER[%]	Ranking 2
Intensidad	27.0%	1
F3	30.4%	2
F1	31.0%	3
CG	35.5%	4
F2-F1	37.0%	5
Asimetría	37.7%	6
Duración	38.5%	7
F4-F3	41.6%	8
F4	41.8%	9
F2	42.5%	10
Curtosis	45.4%	11
Desv.Est.	45.4%	12

Tabla 5. *Ranking de efectividad en la discriminación de hablantes en base a la tasa de igual error (EER) de cada parámetro en forma individual.*

El algoritmo clasificador propuesto comienza empleando el parámetro de mejor rendimiento individual (menor valor de EER), para luego ir adicionando el siguiente parámetro, en orden decreciente de rendimiento. El procedimiento anterior continúa cíclicamente hasta que no queda ningún parámetro sin incluir en el clasificador. En la tabla 6 se pueden ver los resultados obtenidos para cada ciclo. La mejor combinación de parámetros, empleando esta metodología, se observa en el tercer ciclo cuando el clasificador utiliza los parámetros de Intensidad, F3 y F1.

Ciclo	EER[%]	Parámetros empleados
1	27.0%	Intensidad
2	30.5%	Intensidad-F3
3	25.0%	Intensidad-F3-F1
4	33.2%	Intensidad-F3-F1-CG
5	33.4%	Intensidad-F3-F1-CG-(F2-F1)
6	33.4%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría
7	33.4%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría-Duración
8	33.0%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría-Duración-(F4-F3)
9	31.3%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría-Duración-(F4-F3)-F4
10	30.7%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría-Duración-(F4-F3)-F4-F2
11	30.7%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría-Duración-(F4-F3)-F4-F2-Curtosis
12	31.3%	Intensidad-F3-F1-CG-(F2-F1)-Asimetría-Duración-(F4-F3)-F4-F2-Curtosis-Desv.Est.

Tabla 6. *Tasa de igual error (EER) empleando diferente cantidad de parámetros con un algoritmo de identificación de hablantes cíclico.*

4.3. Árbol de decisión

Para analizar los resultados a partir de un clasificador estándar, se utilizó el algoritmo C.45, que genera árboles de decisión. Para su implementación se empleó el paquete de algoritmos de minería de datos WEKA desarrollado por la universidad de Waikato (Nueva Zelanda) (Hall *et al.* 2009). El análisis, que empleó exclusivamente la base intra-hablante, utilizó la metodología de validación cruzada de dejar uno afuera, que consiste en eliminar casos uno a uno y clasificarlos según el modelo resultante del resto de casos. La tasa de igual error promedio fue del 12,5%, lográndose un 70,8% de casos clasificados correctamente.

4.4. Sistema de reconocimiento automático de hablantes

De manera de poder evaluar el poder discriminatorio del fonema /s/, con respecto al resto de los fonemas, se realizó un último experimento considerando el corpus completo y empleando un sistema de reconocimiento de hablantes basado en modelos ocultos de Markov (HMM), que emplea fonemas como unidades básicas del modelado acústico (Univaso *et al*. 2012).

Partiendo de la transcripción manual de las frases emitidas por cada hablante y por medio de un transcriptor fonético, desarrollado especialmente para el Español de Argentina de la región SUR, se extrajo la información acústica de cada fonema por medio de la alineación forzada realizada con un sistema de reconocimiento de habla (ASR), basado en modelos ocultos de Markov (HMM). Posteriormente se empleó dicha información para identificar el hablante, por medio de un clasificador basado en la prueba del cociente de verosimilitudes, con respecto a un modelo universal (UBM).

Se empleó la metodología de validación cruzada de 10 particiones (10-fold cross validation), la cual permite realizar particiones múltiples de los datos para luego estimar el error en base al promedio de las mismas, habiéndose compuesto cada partición con 124 hablantes para la conformación del UBM y 12 hablantes para el testeo de identidades.

Es decir, se realizó la identificación de hablantes empleando un modelo (HMM) a la vez, resultando las tasas de igual error correspondientes a cada fonema como pueden verse en la figura 9. La tasa de igual error (EER) de la fricativa sibilante /s/ fue del 24%, siendo el promedio de los 30 fonemas del 37%. La /s/ se ubicó como el sexto fonema en importancia después de las vocales /e/, /a/, /o/ y /i/, y la nasal /n/. Las vocales que la preceden y el orden relativo de las mismas son similares a las obtenidas por Kahn *et al*. (2011) en su análisis de la variabilidad de vocales para el francés.

Finalmente se realizó un último experimento con el mismo reconocedor de hablantes basado en HMM, pero empleando, en este caso, todos los fonemas en forma simultánea, resultando una tasa de igual error (EER) de 4,5%.

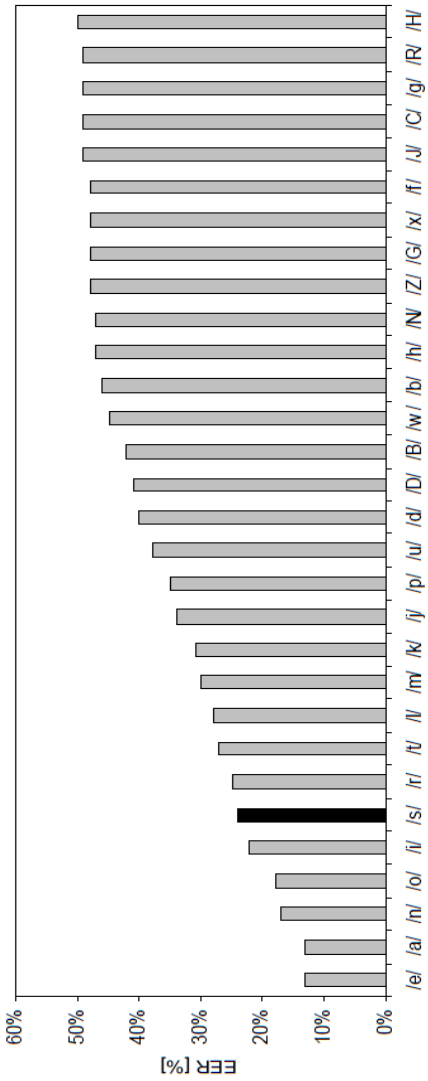

Figura 9. *Tasa de igual error (EER) del corpus completo considerando el empleo de cada fonema en forma independiente con un sistema de reconocimiento de hablantes basado en modelos ocultos de Markov (HMM).*

5. DISCUSIÓN

Comparando los rankings de las tablas 4 y 5 puede observarse que los cuatro primeros parámetros son coincidentes (Intensidad, F3, F1 y CG), aunque los dos últimos se encuentran en diferente orden. Un moderado factor de correlación (fc: 0,80) entre ambas metodologías apoya la hipótesis inicial de la existencia de rasgos discriminantes relacionados con las variaciones intra- e inter-hablante. Los rasgos discriminantes F3 e Intensidad, propuestos en la hipótesis inicial, formaron parte de los principales parámetros en los resultados obtenidos con los métodos del coeficiente discriminante y del algoritmo de identificación de hablantes cíclico.

Ante la ausencia del principal rasgo distintivo de este sonido, como es la concentración de energía en los 5.000-8.000 Hz, cabe hipotetizar que la distribución de energía en el resto de las bandas corresponde a un rasgo discriminante del hablante, como muestran los resultados sobre la estructura de formantes.

La importancia demostrada por el CG, como parámetro discriminador, estaría relacionada con la estructura de los formantes, que en el caso de las consonantes sibilantes está influenciada por los detalles anatómicos alrededor y delante del reborde alveolar.

Era predecible que el F3 fuera uno de los parámetros de mejor discriminación, dada su relación con la longitud del tracto vocal, una de las diferencias físicas más comunes entre hablantes.

La menor tasa de igual error, empleando el algoritmo de identificación de hablantes cíclico, se obtuvo combinando los parámetros de Intensidad, F3 y F1 (tabla 6), lográndose un valor del 25,0%. Otro resultado interesante observado es que el uso de un solo parámetro (Intensidad) mejora el reconocimiento obtenido empleando la totalidad de los mismos.

El elevado valor obtenido en la tasa de reconocimiento de hablantes del 70,8%, empleando el clasificador C.45, corrobora la hipótesis inicial de que los principales rasgos discriminantes podrían ser extraídos a partir de los propios rasgos distintivos de la fricativa sibilante /s/. Una de las limitaciones del presente trabajo fue el empleo de un canal telefónico, que restringe el ancho de banda a frecuencias inferiores a 3.500 Hz. A pesar de esta limitación, la tasa de reconocimiento es similar a la obtenida por Cicres (2011:46), quién obtuvo un 79,4% utilizando un ancho de banda de entre 500 y 16.000 Hz.

El poder discriminante de la fricativa sibilante /s/, con respecto al resto de los fonemas, ha quedado corroborado por su importante aporte a la tasa de reconocimiento de hablantes obtenida, siendo el sexto fonema en importancia en la discriminación de hablantes. La tasa de igual error empleando solamente este fonema resultó un 35% menor que la media de todos los fonemas.

6. CONCLUSIONES

La principal contribución del presente trabajo es la caracterización de los parámetros acústicos (Intensidad, F3, F1 y CG) como rasgos discriminantes de hablantes, correspondientes a los segmentos fricativos sibilantes /s/ en la secuencia VCV. De esta manera se revalorizó el aporte que puede realizar la fonética al campo del reconocimiento de hablantes, ciencia que ha estado poco presente en los últimos años del desarrollo de sistemas automáticos de identificación y verificación de hablantes. También se ha determinado la importancia discriminante del fonema /s/ en el total del grupo fonémico, lo cual deberá también tenerse en cuenta en el desarrollo de esos sistemas.

Hemos podido demostrar que, para el caso de la fricativa sibilante /s/, los rasgos discriminantes del hablante se encuentran dentro de los propios rasgos distintivos de su producción fonética, de manera que los parámetros acústicos que permiten diferenciar a este fonema del resto de los fonemas también llevan información del hablante emisor. Las teorías que estudian tanto la percepción como la producción de habla no introducen este tipo de concepto idiolectal, apoyando las críticas de De Saussure (1987:41-44) y de algunos lingüistas que niegan que el estudio del idiolecto se base en los métodos habituales de la lingüística, mientras que otros consideran marginal el estudio de estos fenómenos para la comprensión de la lengua (Ducrot y Todorov 2005:74). Considerar el idiolecto como un fenónemo comunicacional podría permitir la ampliación de dichas teorías, de manera de incorporar en las características del emisor las propias del receptor en cuestión, con lo cual la discriminación del hablante permitiría adicionar otra dimensión suprasegmental al lenguaje. El reconocimiento del interlocutor puede generar cambios en la propia producción del habla, como es el caso del habla espontánea coloquial.

En el futuro, la metodología empleada podrá extenderse al resto de los fonemas, determinando rasgos discriminantes de cada fonema por medio del Coeficiente discriminante (Cd). Esto permitirá realizar un mapa fonémico de parámetros discriminantes.

7. REFERENCIAS BIBLIOGRÁFICAS

BAUM, S. y S. E. BLUMSTEIN (1987): «Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English», *Journal of the Acoustical Society of America*, 82, pp. 1073-1077.

BEHRENS, S. J. y S. E. BLUMSTEIN (1988): «On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants», *Journal of the Acoustical Society of America*, 84, pp. 861–867.

BOERSMA, P. y D. WEENINK (2005): *Praat software (version 5.2.01)*, Amsterdam, Universidad de Amsterdam.
http://www.fon.hum.uva.nl/praat. [11/11/2012]

BORZONE DE MANRIQUE, A. M. (1980): *Manual de fonética acústica*, Buenos Aires, Hachette.

BORZONE DE MANRIQUE, A. M. y M. I. MASSONE (1979): «On the identification of Argentine Spanish Fricatives», en E. Fischer-Jörgensen, J. Rischel y N. Thorsen (eds): *Proceedings 9th International Congress of Phonetic Sciences*, Copenhague, Universidad de Copenhague, vol. I, p. 237.

BORZONE DE MANRIQUE, A. M. y M. I. MASSONE (1981): «Acoustic analysis and perception of Spanish fricative consonants», *Journal of the Acoustical Society of America*, 69, pp. 1145-53.

CAMPBELL, J. P. JR. (1997): «Speaker recognition: A tutorial» en *Proceedings of the Institut of Electrical and Electronics Engineers*, Nueva York, IEEE, vol. 85, 9, pp. 1437-1462.

CARNEY, P. J. y K. L. MOLL (1971): «A cinefluorographic investigation of fricative consonant-to-vowel coarticulation», *Phonetica*, 23, pp. 193-202.

CICRES, J. (2011): «Los sonidos fricativos sordos y sus implicaciones forenses», *Estudios filológicos*, 48, pp. 33-48.

CUADRADO, L. A. H. (1995): *Introducción a la teoría y estructura del lenguaje*, Madrid, Verbum Editorial.

DE SAUSSURE, F. (1916): *Curso de lingüística general* (C. Bally, y A. Sechehaye, eds.), Madrid, Alianza, 1987.

DELATTRE, P. (1967): «Acoustic or articulatory invariance», *The General Phonetic Characteristics of Languages*, Santa Bárbara, Universidad de California.

DUCROT, O. y T. TODOROV (2005): *Diccionario enciclopédico de las ciencias del lenguaje*, Buenos Aires, Siglo XXI Editores Argentina.

FLIPSEN, P. JR.; L. SHRIBERG; G. WEISMER; H. KARLSSON y J. MCSWEENY (1999): «Acoustic characteristics of /s/ in Adolescents», *Journal of Speech, Language and Hearing Research*, 42, 3, pp. 663-677.

FORREST, K.; G. WEISMER; P. MILENKOVIC y R. DOUGALL (1988): «Statistical analysis of word- initial obstruents: Preliminary data», *Journal of the Acoustical Society of America*, 84, pp. 115-123.

GIBBONS, J. y M. T. TURELL (eds.) (2008): *Dimensions of forensic linguistics*, Amsterdam/Filadelfia, John Benjamins Publishing.

GURLEKIAN, J. A. (1981): «Recognition of Spanish Fricatives /s/ and /f/», *Journal of the Acoustical Society of America*, 70, 6, pp. 1624-1627.

GURLEKIAN, J. A.; L. COLANTONI; H. TORRES; A. RINCÓN; A. MORENO y J. MARIÑO (2001a): «Database for an Automatic Speech Recognition System for Argentine Spanish», en S. Bird, P. Buneman y M. Liberman (eds.): *Proceedings of the IRCS Workshop on Linguistic Databases*, Filadefia, Editorial LDC-Upenn, Research in Cognitive Sciences and the NSF Project on International Standards in Language Engineering, 1, pp. 219-227.

GURLEKIAN, J. A.; COLANTONI, L. y TORRES, H. (2001): «El alfabeto fonético SAMPA y el diseño de córpora fonéticamente balanceados», *Fonoaudiológica*, Editorial ASALFA, 47, 3, pp. 58-69.

HALL, M.; E. FRANK; G. HOLMES; B. PFAHRINGER; P. REUTEMANN e I. WITTE (2009): «The WEKA Data Mining Software: An Update», *SIGKDD Explorations*, vol. 11, 1, pp. 10-18.

HEINZ, J. M. (1961): «Analysis of fricative consonants», *MIT Research Lab of Electronics Quartely Progress Report*, 60, pp. 181-84.

HEINZ, J. M. y K. N. STEVENS (1961): «On the Properties of Voiceless Fricatives Consonants», *Journal of the Acoustical Society of America*, 33, pp. 589-96.

HUGHES, G. W. y M. HALLE (1956): «*Spectral Properties of Fricative Consonants*», *Journal of the Acoustical Society of America*, 28, 2, pp. 303-310.

JASSEM, W. (1965): «The formants of fricative consonants», *Language and Speech*, 8, 1, pp. 1-16.

JASSEM, W. (1968): «Acoustical description of voiceless fricatives in terms of spectral parameters», *Speech analysis and synthesis*, 1, pp. 189-206.

JONGMAN, A.; R. WAYLAND y S. WONG (2000): «Acoustic characteristics of English fricatives», *Journal of the Acoustical Society of America*, 108, p. 1252.

KAHN, J.; N. AUDIBERT; S. ROSSATO y J. F. BONASTRE (2010): «Intra-speaker variability effects on speaker verification performance», *Odyssey 2010*, Brno (República Checa), pp. 109-116.

KAHN, J.; N. AUDIBERT; J. F. BONASTRE y S. ROSSATO (2011): «Inter and intra-speaker variability in French: an analysis of oral vowels and its implication for automatic speaker verification», en W.-S Lee y E. Zee (eds): *International Congress of Phonetic Science*, Hong Kong, Universidad de Hong Kong, pp. 1002-1005.

MAGRIN-CHAGNOLLEAU, I.; J. F. BONASTRE y F. BIMBOT (1995): «Effect of utterance duration and phonetic content on speaker identification using second-order statistical methods», en J. M. Pardo (ed): *Eurospeech '95*, Madrid, ISCA, pp. 337-340.

MARRERO, V.; J. GIL y E. BATTANER (2003): «Inter-speaker variation in Spanish. an experimental and acoustic preliminary approach», en Ma. J. Solé y J. Romero (eds): *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, UAB, pp. 703-706.

MORENO, A.; R. COMEYNE; K. HASLAM; H. VAN DEN HEUVEL; H. HÖGE; S. HORBACH y G. MICCA (2000): «SALA: SpeechDat Across Latin America. Results of the first phase», en M. Gavrilidou, G. Carayannis, S. Markantonatou, S. Piperidis y G. Steinhaouer (eds): *Proceedings of the Second International Conference on Language Resources and Evaluation*, Universidad Técnica Nacional de Atenas, II, pp. 877–882.

NORDSTRÖM, P. E. y B. LINDBLOM (1975): «A normalization procedure for vowel formant data», *International Congress on Phonetic Sciences*, Leeds, Universidad de Leeds, paper 212.

ROSE, P. (2002): *Forensic Speaker Identification*, London, Taylor & Francis.

SAMBUR, M. (1975): «Selection of Acoustic Features for Speaker Identification», *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 23, 2, pp. 176-182.

STEVENS, K. N. (1972): «Sources of inter- and intra- speaker variability in the acoustic properties of speech sounds», en A. Rigault y R. Charbonneau (eds.): *Proceedings of the 7th International Congress of Phonetic Sciences*, La Haya, Mouton, pp. 206-232.

STREVENS, P. (1960): «Spectra of fricative noise in human speech», Language and Speech, 3, 1, pp. 32-49.

TABAIN, M. (2001): «Variability in Fricative Production and Spectra. Implications for the Hyper-and Hypo-and Quantal Theories of Speech Production», *Language and Speech*, 44, 1, pp. 57-93.

TODA, M.; S. MAEDA y K. HONDA (2010): «Formant-cavity affiliation in sibilant fricatives», *Turbulent Sounds: An Interdisciplinary Guide*, 21, pp. 343-374.

UNIVASO, P.; M. MARTÍNEZ SOLER; D. EVIN y J. A. GURLEKIAN (2012): «A preliminary approach to forensic speaker recognition using phonemes», en D. Torre, A. Ortega, A. Teixeira, J. González, L. Hernández, R. San Segundo y D. Ramos Castro (eds.): *IberSPEECH 2012, VII Jornadas en Tecnología del Habla and III Iberian SLTech Workshop*, Madrid, UAM.
http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf [01/12/2012]

VIDAL DE BATTINI, B. (1964): *El Español de Argentina*, Buenos Aires, Consejo Nacional de Educación, 1983.

WEIRICH, M. (2010): «Articulatory and Acoustic Inter-Speaker Variability in the Production of German Vowels», *ZAS Papers in Linguistics*, 52, pp. 19-42.

WINSKY, R. (1997): «Definition of Corpus, Scripts and Standards for Fixed Networks», *SpeechDat project, doc ref LE2-4001-SD1*, vol. 1, pp. 22.