

Análisis discriminante de un conjunto de parámetros fonético-acústicos de las pausas llenas para identificar hablantes

Jordi Cicres

ForensicLab – Laboratori de Lingüística Forense, Institut Universitari de Lingüística Aplicada,
Universitat Pompeu Fabra, La Rambla 30-32, 08002 Barcelona, España
e-mail: jordi.cicres@upf.edu

En este artículo se presentan resultados preliminares de un estudio cuya finalidad es la identificación de hablantes en el ámbito de la fonética forense. Se utiliza el análisis discriminante para medir el poder discriminante de un conjunto de variables fonético-acústicas referentes a la calidad vocálica de la voz en los segmentos correspondientes a las pausas llenas (los segmentos dubitativos con contenido fonético, usualmente vocales alargadas). Los resultados muestran un porcentaje bastante elevado (86,3%) de clasificaciones correctas sobre un corpus piloto de 8 hablantes. Si se analizan los casos de hombres y mujeres por separado, los porcentajes de clasificación correcta llegan al 95,8 y al 93,6% respectivamente, por lo que se reafirma la utilidad del análisis de estos segmentos para la identificación de hablantes.

Palabras clave: identificación, hablante, fonética, forense.

This paper presents preliminary results of a forensic phonetics identification study which aims at discriminate speakers in the forensic phonetics field. Discriminant power of a set of phonetic-acoustic variables related to voice quality of the full pauses (or hesitation pauses) in a 8-speakers corpus is investigated. Results show a quite high proportion (86,3%) of correct classifications. If men and women are analysed separately, the percentages of exact classifications are 95.8 and 93.6% respectively. Therefore, it concludes that discriminant analysis of these voice parts is useful for speaker identification.

Key words: identification, speaker, forensic, phonetic.

1. INTRODUCCIÓN

La fonética es la ciencia del habla. Implica el estudio sistemático de los sonidos del lenguaje, incluyendo sus propiedades acústicas, articulatorias y perceptivas. La fonética forense

hace uso de estos conocimientos para aplicarlos en el ámbito judicial, como por ejemplo la identificación de hablantes, la determinación del contenido de una grabación o la construcción de perfiles lingüísticos.

La mayoría de estudios fonéticos se basan en la descripción: describen las características de los sonidos en distintas lenguas o dialectos, estudian sus cambios históricos, los mecanismos de producción y percepción, y la relación con la fonología u otras disciplinas. Así, pues, habitualmente estudian aquellas características de la voz que son comunes a todos los hablantes de un dialecto o lengua, con el fin de realizar una descripción general, obviando las diferencias intrahablantes (en las distintas realizaciones de un mismo hablante) e interhablantes (en las realizaciones de hablantes distintos). La fonética forense, por el contrario, pone el acento en el estudio de estas diferencias, con el fin de individualizar las voces.

Esta tarea de identificación se puede llevar a cabo desde dos perspectivas distintas: auditiva o acústica. El análisis auditivo consiste en una audición atenta del investigador que utiliza su conocimiento lingüístico previo (fonético, fonológico, morfológico, sintáctico, léxico, semántico, discursivo) para comparar muestras de voces y determinar, en base a su experiencia, si dos muestras de voz han sido producidas por una misma persona o por personas diferentes. El análisis acústico, por el contrario, se basa en el uso de aplicaciones informáticas de análisis que permiten detectar aspectos inaudibles para el oído humano y cuantificar y explicitar aquellos rasgos acústicos detectados auditivamente. En ambos casos, las técnicas estadísticas permiten cuantificar las conclusiones. En la actualidad, se utiliza una técnica combinada de análisis auditivo y acústico, a veces complementado con análisis automático.

Cuando se escucha a alguien hablando, el oyente obtiene información sobre el contenido (qué se dice) y sobre la persona que habla (calidad vocal). Según Rose [1], la calidad vocal (*voice quality*) tiene dos componentes: el componente orgánico y el componente (o ajuste) articulatorio. El primero se refiere a los aspectos del sonido determinados por la fisiología del hablante (forma y tamaño de las cavidades bucal y nasal, características de las cuerdas vocales, etc.); el segundo se refiere a la posición y ajustes de los órganos articulatorios (lengua, labios, etc.). El presente estudio concierne sólo al componente orgánico de la calidad vocal.

Los mejores segmentos para analizar la calidad vocal son las llamadas pausas llenas. En [1] se distinguen 5 clases de pausas: las pausas vacías, llenas (dubitativas), de unión, de respiración y de turno. Tanto las pausas vacías como las llenas son recursos comunicativos para dar continuidad al acto del habla. Pueden ser simplemente periodos de silencio (pausas vacías) o bien pueden tener contenido fonético, habitualmente [e:], [↔:], [i:] y [m:]. En estos casos, se pueden considerar como pausas dubitativas. Las pausas de unión son más breves, y se producen entre los límites de unidades de entonación. Las pausas de respiración se producen en el momento de inspirar, y finalmente, las pausas de turno se producen en los cambios de turno en una conversación. Una alternativa a las pausas llenas es el alargamiento de un fonema de una palabra del discurso.

Estudios previos ([2], [3], [4]) han estudiado las pausas llenas desde el punto de vista acústico y sociolingüístico. Así, [2] y [3] analizaron parámetros acústicos tales como la frecuencia de los formantes (F1, F2 y F3) y el tipo de pausa, en alemán e inglés respectivamente, obteniendo como resultado la demostración de que estos sonidos son diferentes respecto de los formantes de las vocales léxicas, y ligeramente más estables. Asimismo, este estudio pone de manifiesto que hay factores sociales (edad, sexo y nivel sociocultural) que influyen en la elección del tipo de sonido dubitativo. El estudio encuentra algunos ejemplos de hablantes cuyas pausas llenas son marcadamente diferentes de las del resto de hablantes participantes, por lo que serían especialmente fáciles de discriminar forensicamente.

Finalmente, en [4] se analiza un grupo de parámetros referentes a la voz y se buscan diferencias estadísticamente significativas entre los hablantes mediante el análisis de la varianza (ANOVA). Los resultados indican que es posible identificar hablantes mediante el estudio de estos

parámetros. Sin embargo, el empleo de estas comparaciones no es suficiente, puesto que hay parejas que no muestran diferencias estadísticamente significativas en ninguno de los parámetros. También concluye que los mejores resultados se obtienen en la discriminación de voces femeninas, y de voces femeninas respecto de las masculinas. Se analizaba al mismo tiempo la variación intrahablante, que no fue considerada estadísticamente significativa en ninguno de los participantes.

En el presente estudio experimental se pretende utilizar información fonética con el objetivo de identificar hablantes. El conjunto de 19 parámetros analizados hacen referencia a la calidad vocal. Se utilizan las técnicas del análisis discriminante para determinar el poder clasificatorio de este grupo de variables.

2. LAS VARIABLES

Las variables analizadas hacen referencia a la frecuencia fundamental (*pitch*), *jitter*, *shimmer*, razón ruido-armónicos e interrupción de la sonoridad. Estas variables se analizan de segmentos vocálicos de larga duración. Dado que el corpus de análisis está formado por grabaciones de habla espontánea, los únicos segmentos fiables para el análisis son las pausas llenas, es decir, los segmentos dubitativos vocálicos utilizados para evitar pausas en el discurso. Típicamente, las pausas llenas se realizan mediante los sonidos vocálicos [e:] y [↔:], seguidas a veces por [m], o bien únicamente [m:], con duraciones medias aproximadas de alrededor de los 600 milisegundos.

En total, se han tenido en cuenta 19 variables (Tabla 2.1.).

Tabla 2.1.
Resumen de las variables de estudio

Tabla 2.1. Resumen de las variables de estudio		
Grupo de parámetros	Parámetro	Explicación (entre paréntesis, unidad de medida)
Referentes a la frecuencia fundamental (F0)	Media	Valor medio de F0 (Hz)
	Mediana	Mediana de F0 (Hz)
	Desviación estándar	Desviación estándar de F0 (Hz)
	Valor mínimo	Valor mínimo de F0 (Hz)
	Valor máximo	Valor máximo de F0 (Hz)
Referentes a la interrupción de la sonoridad y a la aperiodicidad	Porcentaje de periodos sordos	Porcentaje de segmentos sordos (%)
	Grado de segmentos sordos	Porcentaje de marcos, de 1 milisegundo, analizados como sordos (%)
Referentes a la perturbación de frecuencia (jitter)	<i>Jitter</i> (local)	Variabilidad período-a-período, dividida por el valor medio del periodo (%)
	<i>Jitter</i> (local, absoluto)	Variabilidad período-a-período (segundos)
	<i>Jitter</i> (rap)	Variabilidad período-a-período con un suavizado de 3 periodos, dividida por el valor medio del periodo (%)
	<i>Jitter</i> (ppq)	Variabilidad período-a-período con un suavizado de 5 periodos, dividida por el valor medio del periodo (%)
	<i>Jitter</i> (ddp)	Diferencia absoluta media entre diferencias consecutivas entre periodos consecutivos, dividida por el periodo medio (%)

Referentes a la perturbación de amplitud (Zimmer)	<i>Shimmer</i> (local)	Es variabilidad de amplitud de periodos consecutivos dividido por la amplitud media (%)
	<i>Shimmer</i> (local, dB)	Calcula la variabilidad de amplitud en periodos consecutivos en valores absolutos (dB)
	<i>Shimmer</i> (apq3)	Calcula la variabilidad de amplitud pico a pico con un suavizado de 3 periodos, dividido por la amplitud media (%)
	<i>Shimmer</i> (apq5)	Calcula la variabilidad de amplitud pico a pico con un suavizado de 5 periodos, dividido por la amplitud media (%)
	<i>Shimmer</i> (apq11)	Calcula la variabilidad de amplitud pico a pico con un suavizado de 11 periodos, dividido por la amplitud media (%)
	<i>Shimmer</i> (dda)	La diferencia absoluta media entre diferencias consecutivas entre las amplitudes de periodos consecutivos (%)
Referentes al ruido	Razón ruido/armónicos	Mide la cantidad de energía no armónica presente en el sonido respecto a la energía armónica producida por la vibración de las cuerdas vocales. Incluye toda la energía inarmónica presente en la onda sonora, independientemente de su origen (turbulencias, irregularidades en los periodos vibratorios, etc.) (dB)

3. METODOLOGÍA

Se ha trabajado sobre un corpus de habla espontánea constituido por 8 hablantes (la mitad hombres y la mitad mujeres), catalanohablantes, jóvenes, y de un mismo origen social y dialectal.

El análisis acústico se ha realizado con el programa Praat [5], versión 4.5.08. Se han extraído los valores de las variables a partir de la opción *Voice report*. Esta opción es similar al programa MultiDimensional Voice Program [6], si bien presenta algunas diferencias en la metodología de cálculo de algunos parámetros [7]. Aunque la utilidad inicial de esta opción es el análisis exploratorio clínico [6], [8], también puede utilizarse para el análisis de segmentos vocálicos de personas sin ninguna patología con el fin de obtener datos sobre su calidad vocálica.

Previamente se extrajeron los fragmentos dubitativos para el análisis, que se llevó a término en un marco correspondiente al 60% central de la vocal dubitativa, descartando así el 20% inicial y final de la duración de la vocal con el fin de evitar posibles influencias de la coarticulación, por un lado, y de un grado de voz laringalizada excesivo, por el otro, que no representaría la calidad vocálica habitual del hablante, sino únicamente en los finales de frase. De este modo, se analiza la parte más estable de la vocal.

Se han analizado de 10 a 13 pausas llenas por hablante, hasta un total de 95 casos. El análisis estadístico se ha llevado a término mediante el programa SPSS versión 14.0. Se ha utilizado la técnica del análisis discriminante para clasificar los casos. Se trata de una técnica estadística multivariable que permite clasificar los casos en grupos determinados mediante funciones discriminantes (combinaciones de variables que presentan una mínima variación intragrupo y una mayor variación intergrupos). Este análisis también permite predecir los grupos a los que pertenecen nuevos elementos.

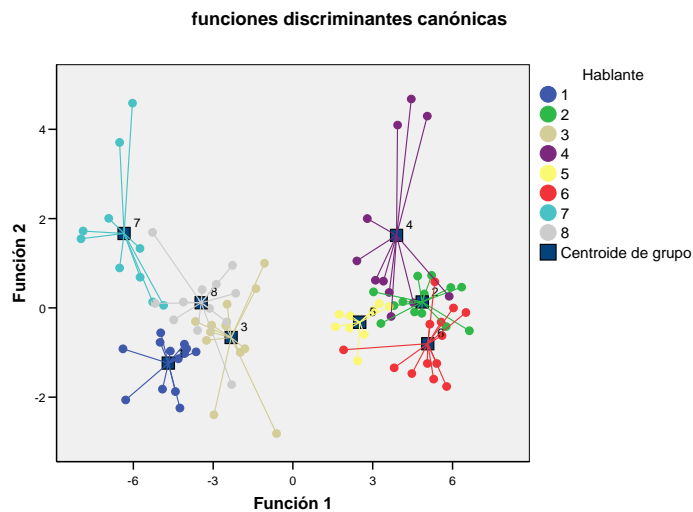
En el análisis forense del habla, esta técnica es útil para comparar los datos multivariados de muestras dubitadas con respecto de las muestras indubitadas estudiando si las funciones discriminantes son compatibles con alguna de las muestras indubitadas.

El gráfico de la Figura 4.2. muestra el diagrama de dispersión de los resultados de las dos primeras funciones discriminantes. Cabe destacar, por un lado, cierto solapamiento de los casos entre algunos hablantes; por el otro, la formación de dos grupos muy bien diferenciados de hablantes, a la izquierda y derecha del gráfico. Los hablantes de cada uno de estos 2 grupos comparten el sexo, por lo que un análisis de hombres y mujeres por separado es pertinente para obtener mejores resultados. Es obvio que el grupo de parámetros referentes a la frecuencia fundamental va a mostrar diferencias muy significativas entre hombres y mujeres, que se ven representadas principalmente en la primera función discriminante.

Tabla 4. 1.
Resultados de la clasificación (todos los hablantes).

Tabla 4. 1. Resultados de la clasificación (todos los hablantes).								
Habla nte	Grupo de pertenencia pronosticado							
	1	2	3	4	5	6	7	8
1	13	0	0	0	0	0	0	0
2	0	13	0	0	0	0	0	0
3	1	0	9	0	0	0	0	2
4	0	4	0	7	0	0	0	0
5	0	0	0	0	10	0	0	0
6	0	2	0	0	1	10	0	0
7	2	0	0	0	0	0	8	0
8	0	0	1	0	0	0	0	12

Figura 4.2.
Diagrama de dispersión de las dos primeras funciones discriminantes con los 8 hablantes.



Al realizar el análisis sólo con hablantes del mismo sexo, los resultados de clasificación mejoran notablemente, llegando al 95,8% en los hombres, y al 93,6% en las mujeres, con sólo 3 funciones discriminantes.

La Figura 4.3. muestra los valores de las funciones en los centroides de los grupos en el análisis discriminante de los hablantes masculinos, y la Tabla 4.2. muestra los resultados de la clasificación. Se puede observar que los dos únicos casos de confusión se producen entre los

hablantes 1 y 7, y 3 y 8. La primera función discriminante (la que más variación explica) es negativa en ambos casos en la primera confusión, mientras que es positiva en la segunda.

Figura 4.3.

Valores acumulados de las funciones en los centroides de los grupos (hablantes masculinos).

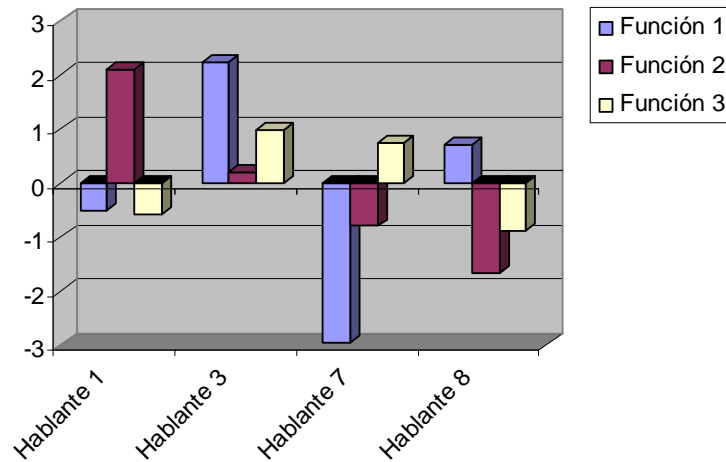
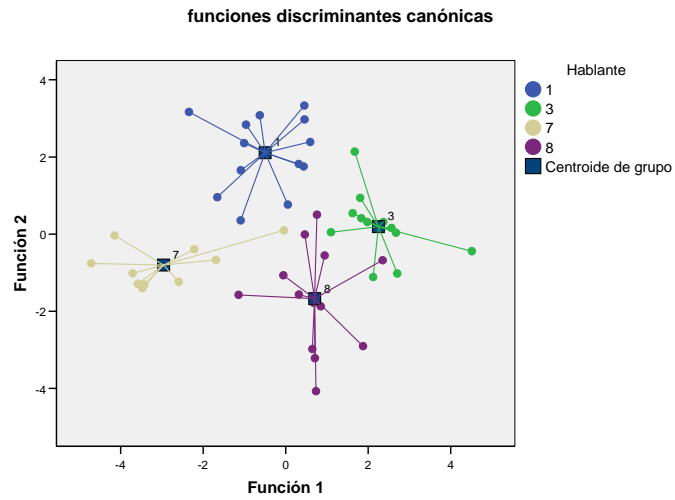


Tabla 4.2. Resultados de la clasificación (hablantes masculinos).

Hablante	Grupo de pertenencia pronosticado			
	1	3	7	8
1	13	0	0	0
3	0	12	0	0
7	1	0	9	0
8	0	1	0	12

La Figura 4.4. muestra el gráfico de dispersión de las dos primeras funciones discriminantes sólo con los hablantes masculinos. Se observa que el grado de solapamiento es mínimo, y que los grupos quedan bien delimitados en su área a excepción de unos pocos casos.

Figura 4.4.
Diagrama de dispersión de las dos primeras funciones discriminantes de los hablantes masculinos



Por su lado, la Figura 4.5. y la Tabla 4.3 muestran los datos referentes al análisis llevado a término únicamente con los hablantes femeninos. En este caso, las tres clasificaciones erróneas se pueden explicar por la tercera función discriminante positiva (hablantes 2 y 4)

Figura 4.5.
Valores acumulados de las funciones en los centroides de los grupos (hablantes femeninos).

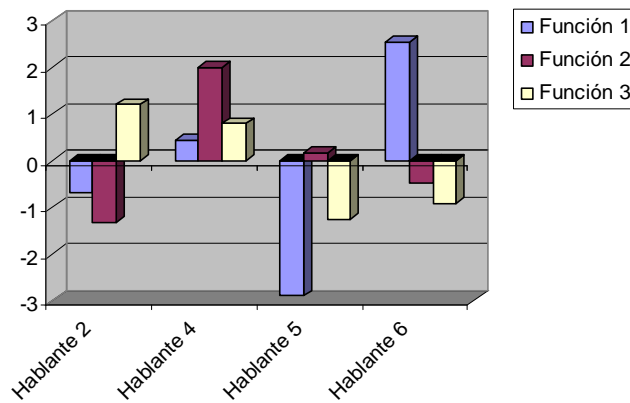


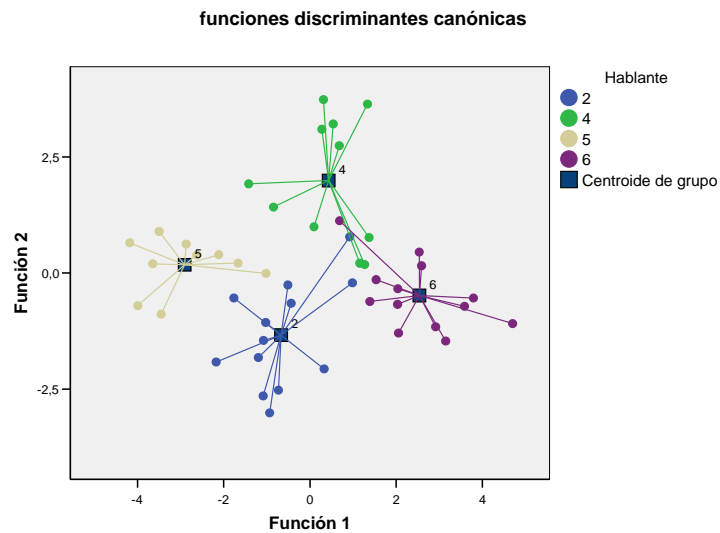
Tabla 4.3. Resultados de la clasificación (hablantes femeninos).

Hablante	Grupo de pertenencia pronosticado			
	2	4	5	6
2	12	1	0	0
4	0	9	1	1
5	0	0	10	0
6	0	0	0	13

La Figura 4.6. muestra el gráfico de dispersión de las dos primeras funciones discriminantes sólo con los hablantes femeninos. En el gráfico se observa la buena delimitación del espacio que ocupan los distintos grupos, con sólo unos pocos casos de solapamiento.

Figura 4.6.

Diagrama de dispersión de las dos primeras funciones discriminantes de los hablantes femeninos.



5. CONCLUSIONES

Los resultados muestran un notable éxito de clasificación a partir de las variables analizadas. Sin embargo, dado el bajo número de hablantes analizados, los resultados tienen que considerarse con cautela. En la práctica forense, hay varios factores que tienen que tenerse en cuenta antes de tomar la decisión de utilizar el análisis multidimensional de la voz para identificar hablantes. En primer lugar, la calidad de la grabación. La precisión de estos parámetros hace que la presencia de ruido de fondo u otras deficiencias en las grabaciones puedan comprometer los resultados. En segundo lugar, las características técnicas de la grabación (tipo de grabadora

utilizada, formato de grabación) también pueden afectar los resultados. En tercer lugar, los aspectos relacionados con el hablante (cambios físicos producidos por el paso del tiempo o por el estado de salud) pueden alterarse notablemente, por lo que el investigador tendrá que cerciorarse de que estas condiciones son similares.

Finalmente, hay que destacar que en ningún caso un dictamen pericial puede basarse únicamente en este análisis, sino que tiene que ser la suma de todo un conjunto de comparaciones acústicas y lingüísticas (en todos los niveles).

REFERENCIAS

- [1] Rose, Ph., *Forensic Speaker Identification*, Taylor & Francis, Londres, 2002.
- [2] Pätzold, M. y Simpson, A., "An acoustic analysis of hesitation particles in German", *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm. Vol. 3. pp. 512-515, 1995.
- [3] Foulkes, P., Carrol, G. y Hughes, S., "Sociolinguistic and Acoustic Variability in Filled Pauses", comunicación oral en el International Association for Forensic Phonetics and Acoustics Conference, Universidad de Helsinki, Finlandia, 2004.
- [4] Cicres, J. y Turell, M.T., "El análisis multidimensional de la voz como herramienta para la identificación del hablante en fonética forense", *Actas del VI Congreso de Lingüística General*, Servicio de Publicaciones de la Universidad de Santiago de Compostela, 2005.
- [5] Boersma, P. y Weenink, D. *Praat: doing phonetics by computer* (Version 4.5.08) [Programa informático]. Disponible en <http://www.praat.org/>.
- [6] Kay Elemetrics Corp, *Multi-Dimensional Voice Program (MDVP) Model 5105* [Manual de usuario]. New Jersey: Kay Elemetrics, 1999.
- [7] Boersma, P. y Weenink, D. Manual on-line del Praat. Epígrafe "Voice 5. Comparison with other programs". http://www.fon.hum.uva.nl/praat/manual/Voice_5__Comparison_with_other_programs.html
- [8] González, J., Cervera, T. y Miralles, J.L., "Análisis acústico de la voz: fiabilidad de un conjunto de parámetros multidimensionales", *Acta Otorrinolaringológica Española* 53. 256-268, 2002.