

Battaner, E., Carbó, C., Gil, J., Llisterri, J., Machuca, M. J., Madrigal, N., . . . Ríos, A. (2007). VILE: Estudio acústico de la variación inter e intralocutor en español. In M. González González, E. Fernández Rei, & B. González Rei (Eds.), *Actas do 3º Congreso Internacional de Fonética Experimental*. (pp. 157-67). Santiago de Compostela: Xunta de Galicia.

[http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE\\_IIICFE05.pdf](http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE_IIICFE05.pdf)

# VILE: ESTUDIO ACÚSTICO DE LA VARIACIÓN INTER E INTRA LOCUTOR EN ESPAÑOL

ELENA BATTANER

*Departamento de Comunicación II, Universidad Rey Juan Carlos*

CARME CARBÓ, JOAQUIM LLISTERRI, MARÍA JESÚS MACHUCA, NATALIA MADRIGAL,

CARME DE LA MOTA, MONTSERRAT RIERA, ANTONIO RÍOS

*Departamento de Filología Española, Universidad Autónoma de Barcelona*

JUANA GIL, VICTORIA MARRERO

*Departamento de Lengua Española y Lingüística General, Universidad Nacional de Educación a Distancia*

## 1. INTRODUCCIÓN

VILE (*Estudio acústico de la variación inter e intralocutor en español*) es un proyecto de investigación básica, financiado por el Ministerio de Ciencia y Tecnología (BFF2001-2551, 2001-2004)<sup>1</sup>, centrado en la caracterización acústica de los elementos segmentales y suprasegmentales que contribuyen a establecer la individualidad de un hablante frente a aquellos que son comunes a un estilo de habla, una variedad geográfica o social, o una lengua. Se pretende con ello obtener el conocimiento fonético necesario para la mejora de los sistemas de identificación o verificación automáticas del locutor, así como dotar a los especialistas en fonética forense de nuevos datos acústicos que permitan comparar, con un mayor grado de certeza, locutores dubitados e indubitados.

En las primeras etapas del proyecto se seleccionaron los parámetros acústicos que se toman en consideración (apartado 2) y el corpus a partir del cual se obtienen los datos (apartado 3). A continuación, se ha llevado a cabo un análisis acústico (apartado 4), del cual se presentan los primeros resultados (apartado 5).

---

1 Para más información puede consultarse <http://liceu.uab.es/~joaquim/VILE.html>

## 2. SELECCIÓN DE LOS PARÁMETROS ACÚSTICOS ANALIZADOS

La revisión de la bibliografía acerca de los parámetros relevantes para el estudio de la individualidad de la voz se centró en tres formas de aproximación a los datos fonéticos: el reconocimiento visual de espectrogramas, el reconocimiento perceptivo y el reconocimiento automático del habla. Debido a la subjetividad implícita en el análisis de los datos que se derivan de la primera aproximación, la búsqueda se limitó a las otras dos aproximaciones que, en cualquier caso, ofrecen un volumen de publicaciones mayor y más en consonancia con las tecnologías actuales. De esta forma, nos centramos en aquella literatura que se refiere a los parámetros denominados *de bajo nivel de información* (Doddington 1985) o los que se corresponden con las dimensiones *fisiológicas* (Kuwabara y Sagisaki 1995). Stevens (1971) destaca dos parámetros generales susceptibles de emplearse en la identificación y discriminación entre hablantes: aquellos referidos a la fuente (frecuencia fundamental, en adelante  $F_0$ , y forma de la onda glotal) y a los resonadores (frecuencias formánticas, anchura de formantes, sonidos turbulentos y consonantes nasales). Hollien (1990 y 1991) añade otros rasgos que parecen decisivos en relación con dicha identificación y discriminación: el espectro a largo plazo, el vector de los formantes vocálicos, el vector temporal y el vector de  $F_0$ . Otros trabajos como Atal (1972), Karlsson (1988), Eskenazi *et al.* (1990), Kuwabara y Takagi (1991) y Kuwabara y Sagisaka (1995) mencionan aproximadamente los mismos parámetros como responsables de la individualidad de la voz.

A la hora de seleccionar los parámetros para el estudio, pueden tenerse en cuenta algunas de las reflexiones que hace Wolf (1972) acerca de cuáles serían los criterios de decisión válidos:

- a) Deberían ser parámetros presentes *natural* y frecuentemente en el habla normal. Los efectos coarticulatorios, en cuanto que son, en cierto grado al menos, 'aprendidos', no se tendrían en cuenta (*cf.* para una opinión contraria, Su *et al.* (1974)).
- b) Han de ser fáciles de medir.
- c) Deben tener la mayor variabilidad posible inter-hablantes y la menor posible intra-hablante. Tanto la forma de la onda glotal como los sonidos turbulentos parecen reunir esas condiciones.
- d) No deberían variar mucho con el tiempo ni verse afectados por las condiciones psico- o fisiológicas del hablante. Las nasales quedarían eliminadas de acuerdo con este último criterio (*cf.* sin embargo Wolf (1972) y Glenn y Kleiner (1968), quienes consideran muy informativo el análisis de las nasales). En el habla espontánea, un rasgo especialmente sensible al estrés experimentado por el hablante es la  $F_0$ , y uno muy resistente es el espectro medio (LTAS, *Long Term Averaged Spectrum*) (Pittam 1987, Hollien 1990).
- e) Han de ser resistentes al posible ruido ambiental y no han de verse afectados por las condiciones de la transmisión. Los valores medios de los formantes vocálicos (del  $F_3$  —que proporciona indicios sobre la longitud del tracto vocal del

hablante-, del F2 y del F1) y su anchura son parámetros muy válidos en este sentido, puesto que son muy resistentes a la distorsión y a las interferencias. Se aconseja estudiar las vocales extremas [a i u]. Por otra parte, se señala que los valores del F1 y F2 de las vocales extremas /i,u,a/ son los más estables y menos sensibles al contexto (Stevens y House 1963).

- f) Finalmente, no pueden ser fácilmente modificables por la mera voluntad del hablante, es decir, deben ser resistentes a los intentos de disimular la voz. La distancia entre los tres primeros formantes no puede ser alterada a voluntad (F1 / F2, F2 / F3).

Existe un importante desacuerdo en lo que se refiere a la importancia relativa de los distintos parámetros y, por ende, a cuál o cuáles deben ser los parámetros prioritarios:

- a) Prioridad del F<sub>0</sub>: Compton (1963), Wolf (1972), Matsumoto *et al.* (1973), Brown (1981), van Dommelen (1987).
- b) Prioridad del LTAS: Bordone-Sacerdote y Sacerdote (1969), Doherty (1976), Hollien y Majewski (1977), Furui (1986), Pittam (1987).
- c) Prioridad de la estructura espectral, bien sea de las frecuencias formánticas absolutas: Shearme y Holmes (1959), Miller (1964), Itoh y Saito (1982), Carrell (1984) Kuwabara y Ohgushi (1987), Kuwabara y Takagi (1991), o bien sea de las trayectorias formánticas: Ingram *et al.* (1996).
- d) Misma prioridad para la F<sub>0</sub> y la estructura formántica: La Riviere (1975).
- e) Variables temporales: Pruzansky (1963), Wolf (1972), Doherty y Hollien (1978), Brown (1981), Johnson *et al.* (1984).
- g) No es factible establecer una prioridad, la importancia de cada parámetro puede diferir de hablante a hablante y depende también de la naturaleza de las muestras: Gobl (1989) y Kuwabara y Sagisaka (1995).
- h) Contorno tonal: Atal (1972), van Dommelen (1987).

Por lo que se refiere a qué formantes o qué distancias entre formantes son las más informativas, tampoco hay coincidencia en las posturas, pero la mayoritaria defiende el estudio de los tres primeros formantes. En este sentido, la falta de resultados coincidentes y concluyentes acerca de cuál es el parámetro más decisivo para el reconocimiento del hablante puede deberse a los enfoques metodológicos empleados, que son muy diferentes. En cualquier caso, parece difícil establecer una jerarquía absoluta entre los parámetros. Varios autores apuntan la interdependencia entre los índices, cuya prioridad relativa dependería, asimismo, del hablante. Por ejemplo, los oyentes pueden tomar como clave primaria para el reconocimiento de un hablante A un F<sub>0</sub> bajo, y sin embargo apoyarse en la estructura formántica para el reconocimiento de un hablante B (van Dommelen 1987). Esto es, todos los rasgos mencionados conllevan un cierto grado de información sobre las características del hablante y son potencialmente válidos para la tarea de reconocimiento.

### 3. SELECCIÓN DEL CORPUS ANALIZADO

El segundo paso fue la revisión de los recursos orales existentes en español, considerando la conveniencia de reutilizar en la medida de lo posible unas bases de datos de gran calidad, obtenidas gracias al trabajo de muchos especialistas, y con frecuencia gracias a financiación pública.

En primer lugar descartamos los que no estaban disponibles para fines investigadores, por ser de propiedad privada<sup>2</sup>. Tampoco se adecuaban a nuestros intereses los que fueron creados con fines específicos, como la variación dialectal o sociolectal<sup>3</sup>, el análisis del discurso o la conversación<sup>4</sup>, etc. Otros (como ROARS, MATE, el corpus oral de la UAM, o el del propio CREA) no nos resultaban accesibles en su totalidad.

Los corpus analizados fueron los siguientes: AHUMADA- GAUDÍ, Albayzín, SpeechDat, EUROM1 y su subconjunto MULTEXT. A continuación resumimos las posibilidades que ofrecen para el estudio de la variabilidad:

#### *Según el tipo de variación*

##### **Variación intralocutor**

El mayor número de locutores repitiéndose a sí mismos los proporciona GAUDÍ: 455 personas leyeron tres veces distintas idéntico texto y diez frases fónicamente equilibrados. En ALBAYZIN encontramos 20 locutores que reproducen dos veces (una de ellas bajo el efecto Lombard) un conjunto de 250 frases equilibradas, mientras que en EUROM1 es una serie de 82 logatomos la que es leída cinco veces por doce emisores.

##### **Variación interlocutor**

La repetición de los mismos estímulos por diferentes sujetos es el fin de todas las bases de datos: SpeechDat, por ejemplo, nos ofrece 999 frases fonéticamente ricas leídas por 1002 personas a través del teléfono. Los 455 locutores mencionados de GAUDÍ también proporcionan una cantidad considerable de estímulos para analizar la variación interlocutor. Por orden decreciente, ALBAYZIN permite comparar las voces de 154 lectores de entre 200 y 500 frases equilibradas. Incluso los 60 locutores de EUROM1 proporcionan un material oral muy abundante.

2 CEUDEX, SPATIS, VESTEL, VOCATEL, VOGATEL, TANGORA, IBM Voice Type, etc.

3 The Sala Project – SpeechDat Across Latin America; Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico. Caracas-77: Estudio sociolingüístico de Caracas, 1977. Caracas-87: Estudio sociolingüístico de Caracas, 1987. CEAP: Corpus de Encuestas en Asunción de Paraguay. CSMV: Corpus Sociolingüístico de Mérida-Venezuela. CSC: Corpus para el estudio del español hablado en Santiago de Compostela. COVJA: Corpus oral de la variedad juvenil universitaria del español hablado en Alicante. ALMECOR: Universidad de Granada. PRE-SEEA Corpus Sociolingüístico del español. Fonética Acústica Experimental del Español de Canarias- U. La Laguna. ILSE – Universidad de Almería. Variedades Urbanas Andaluzas – U. Granada y U. Málaga.

4 Análisis del Discurso Público Actual – Universidad de La Coruña. Análisis del Discurso Oral – Universidad de Granada. ACUAH – Universidad de Alcalá de Henares. Corpus de conversación coloquial- Universidad de Valencia.

### *Según el nivel de análisis*

En el nivel segmental, los datos disponibles permiten analizar todos los sonidos del español en distintas condiciones de naturalidad, desde la lectura de logatomos (con o sin frase portadora), que sería el extremo más artificial, pero también el de mayor control de variables (EUROM1), hasta la lectura de frases equilibradas o fonéticamente ricas (en todas las bases de datos), la lectura de párrafos o textos (EUROM1 y GAUDÍ), o el habla espontánea (sólo en GAUDÍ).

Sin embargo, el estudio del nivel suprasegmental se encontraría limitado a los 20 párrafos con modalidades oracionales interrogativas y exclamativas de EUROM1.

El ritmo de habla, por último, puede ser analizado comparando las lecturas del texto en GAUDÍ, a tres velocidades distintas.

### **3.1 El Corpus Ahumada-Gaudí**

Finalmente, nos decantamos por AHUMADA-GAUDÍ, cuya descripción se encuentra en Ortega *et al.* 2000. Por una parte, porque nos ofrecía abundante material para el estudio de la variación intralocutor, en condiciones de grabación muy controladas, tanto en el aspecto técnico como en el intervalo temporal entre unas sesiones y otras<sup>5</sup>. Por otra, por la existencia de unos minutos de habla espontánea para cada uno de los sujetos. Y por último, porque el material proporcionado también parecía sobradamente abundante para analizar la variación entre hablantes<sup>6</sup>.

Dejando a un lado las tareas con dígitos, a todos los locutores se les pidió:

- a) Lectura de 10 frases fónicamente equilibradas: 409 fonemas
- b) Lectura de un texto de 179 palabras, también equilibrado (712 fonemas), a tres velocidades distintas (normal, rápida y lenta).
- c) Lectura de un texto específico para cada locutor.
- d) Habla espontánea durante más de un minuto, describiendo cuadros y dibujos que tenían preparados, o sus vacaciones, el lugar donde nacieron, etc.

Todas esas tareas se realizaron en seis sesiones de grabación: tres *in situ* con micrófono y tres por teléfono. Las sesiones microfónicas fueron la 1, la 3 y la 5, separadas entre sí por intervalos de entre 20 y 40 días.

## **4. ANÁLISIS ACÚSTICO**

Para el estudio de la variación interlocutor e intralocutor del proyecto VILE se han seleccionado las grabaciones microfónicas del corpus Gaudí-Ahumada correspondientes a

---

5 Limitándonos a las grabaciones con micrófono, las 10 frases leídas tres veces por cada locutor suman 13.650 frases. El texto fue leído nueve veces, tres a velocidad normal, tres a mayor velocidad y tres más lentamente, en total, 4.095 repeticiones.

6 Si se suman las tres repeticiones por sujeto, cada frase ha sido leída 1365 veces, 432 por locutores masculinos y 393 por locutores femeninos. El texto, considerando sólo una velocidad de elocución, ha sido repetido las mismas veces.

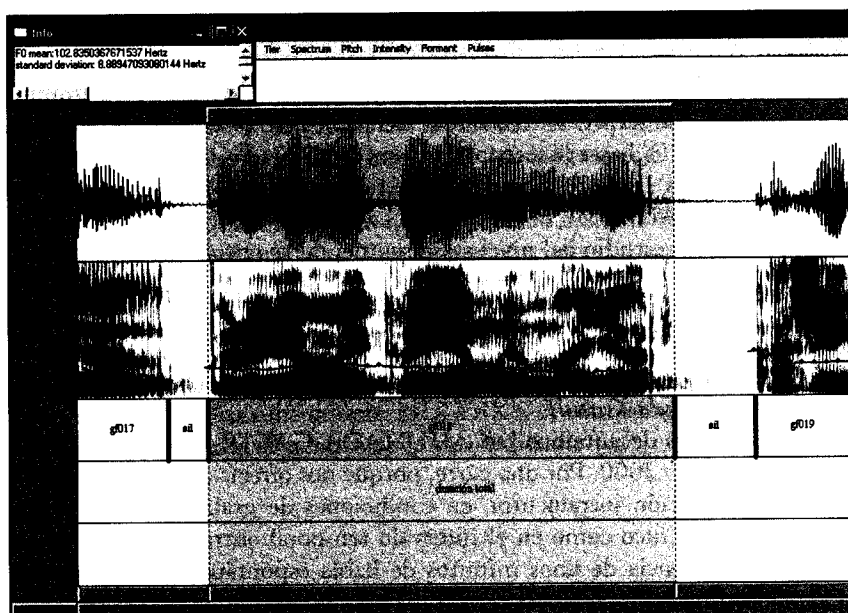


Figura 1. Segmentación y extracción automática de valores de  $F_0$  en el grupo fónico *pero yo me creía muy mayor* procedente del corpus de lectura.

muestras de habla de dos tipos: una, en tarea de lectura y contextos controlados (texto de 179 palabras, a velocidad normal); la otra, a tres sesiones de habla espontánea.

El análisis acústico del corpus, a partir de una selección de las grabaciones de 30 locutores masculinos, se realizó mediante el programa *Praat* (<http://www.praat.org>), desarrollado por Paul Boersma y David Weenink en el Instituto de Fonética de la Universidad de Ámsterdam.

Se eliminaron aquellos locutores cuya grabación de habla espontánea se obtenía mediante la descripción de una lámina, ya que en estos casos el locutor se preparaba con antelación la descripción y, por tanto, su elocución no presenta las características propias del habla espontánea no planificada.

Los parámetros acústicos considerados en el proyecto pueden clasificarse dependiendo del tipo de fuente, de la acción de los resonadores y de los parámetros temporales, tal como se describe a continuación.

#### 4.1 Parámetros referidos a la fuente

El corpus de lectura o de habla espontánea fue segmentado en grupos fónicos (el segmento de habla comprendido entre dos silencios), y de cada uno de ellos se extrajo el valor medio de la  $F_0$  y su fluctuación (la desviación estándar). En la figura 1 puede observarse un ejemplo de segmentación de un grupo fónico y el resultado de la extrac-

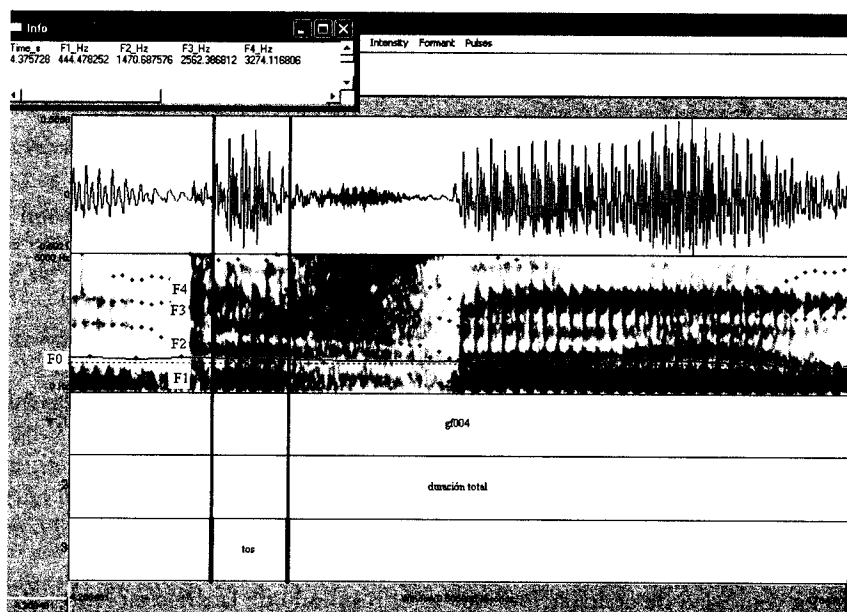


Figura 2. Segmentación y extracción automática de los valores de los cuatro primeros formantes de la vocal [o] del grupo fónico *ese viento suave vuela* procedente del corpus de lectura.

ción automática del valor medio y de la desviación estándar de la  $F_0$  (ventana superior izquierda).

#### 4.2 Parámetros referidos a los resonadores

En el corpus de lectura se segmentaron todas las vocales —tónicas y átonas— precedidas de cualquier oclusiva sorda ( $/p, t, k/$ ) o de fricativa alveolar sorda ( $/s/$ ) y, al mismo tiempo, seguidas de cualquier oclusiva sorda o de fricativa alveolar sorda, que apareciera en una estructura silábica del tipo CV \$ C. Así mismo, se segmentaron todas las fricativas alveolares sordas precedidas y seguidas de las vocales más frecuentes del español, es decir de  $/a/$  y  $/e/$ .

En los segmentos vocálicos se analizaron el valor de la  $F_0$ , las frecuencias centrales de los cuatro primeros formantes y sus anchos de banda (véase la figura 2). A partir de estos valores, se extrajo la distancia entre los valores de frecuencia de  $F_0, F_1, F_2, F_3$  y  $F_4$ . Además, en los segmentos fricativos alveolares sordos se tuvo en cuenta el pico de mayor intensidad situado en la banda de frecuencia entre 0 y 8 kHz.

#### 4.3 Parámetros temporales

En lo que respecta a los parámetros temporales, se consideraron:

- El tiempo total de habla, es decir, la duración total de cada una de las sesiones analizadas.



- b) La proporción de habla, definida como la duración de todos los grupos fónicos de una determinada sesión dividida entre la duración total de dicha sesión.

$$\text{Proporción de habla} = \frac{\text{duración de todos los grupos fónicos de la sesión X}}{\text{duración total de la sesión X}}$$

- c) La proporción de silencios, que se calcula dividiendo la duración total de los silencios de una determinada sesión por la duración total de dicha sesión.

$$\text{Proporción de silencios} = \frac{\text{duración total de silencios de la sesión X}}{\text{duración total de la sesión X}}$$

- d) La velocidad de elocución, entendida como el tiempo total de habla dividido entre el número de sílabas de cada una de las sesiones analizadas.

$$\text{Velocidad de elocución} = \frac{\text{tiempo total de habla}}{\text{número de sílabas de la sesión X}}$$

- e) La velocidad de articulación, que es el resultado de la división entre el tiempo total de habla, exceptuando la duración de los silencios, y el número de sílabas de cada una de las sesiones analizadas.

$$\text{Velocidad de articulación} = \frac{\text{tiempo total de habla} - \text{silencios}}{\text{número de sílabas de la sesión X}}$$

El corpus de lectura se utilizó para el análisis de todos los parámetros descritos; para el estudio de los parámetros temporales y de los parámetros relacionados con la fuente se empleó también el corpus de habla espontánea.

## 5. RESULTADOS

Los resultados obtenidos mostraron variaciones significativas en los tres tipos de parámetros analizados, tal como se expone a continuación

### 5.1 *Parámetros referidos a la fuente*

Los valores medios de la  $F_0$  presentan variaciones muy significativas en función del estilo de habla. Sin embargo, la tendencia no es la misma para los dos estilos observados: en el habla espontánea (excepto para siete locutores, 23,33% del total) se producen variaciones cuadráticas en el valor medio de la  $F_0$  a medida que aumenta el número de sesiones; en lectura, en cambio, no se producen estas variaciones. En 8 locutores (26,66% del total) no se dan diferencias significativas en el valor medio de la frecuencia de  $F_0$  en función del estilo de habla.

La desviación típica de la  $F_0$  no presenta variaciones significativas debidas al estilo de habla. No obstante, la duración del grupo fónico y el número de segmentos del grupo fónico son dos variables que predicen de manera significativa la desviación típica de  $F_0$ .

### 5.2 *Parámetros referidos a los resonadores*

Los valores de frecuencia de los formantes de las vocales analizadas ayudan a identificar a los sujetos. Las resonancias en zonas de alta frecuencia y la  $F_0$  permiten diferenciar en todos los segmentos vocálicos un mayor número de sujetos, mientras que los valores de  $F_1$  y  $F_2$  distinguen un número más reducido.

### 5.3 *Parámetros temporales*

La duración del grupo fónico también presenta diferencias significativas en función del estilo de habla, pero no en relación con la sesión. Sólo 5 locutores (16,66% del total) en el habla espontánea y 2 locutores (6,66% del total) muestran en lectura diferencias significativas en la duración del grupo fónico respecto a la sesión. En la velocidad de elocución y en la velocidad de articulación tampoco se producen diferencias significativas debidas a las distintas sesiones analizadas.

## 6. CONCLUSIONES

Los primeros resultados del análisis acústico hacen pensar que existen elementos que permiten la caracterización de los dos tipos de variación abordados en el proyecto. Por una parte, ponen de manifiesto que el estilo de habla es una variable que debe tomarse en consideración a la hora de comparar muestras de habla —en un contexto judicial, por ejemplo— y, por otra, que ciertos parámetros temporales y frecuenciales parecen mantener una cierta constancia en el mismo locutor a lo largo de las sesiones de grabación.

Además de refinar el tratamiento estadístico de los datos, en la segunda fase de VILE (*Estudio perceptivo de la variación inter e intralocutor en español*, HUM2005-06980/FILO) se plantea la validación perceptiva de los resultados, siguiendo la línea iniciada en el trabajo de Marrero *et al.* (2003), recurriendo tanto a la manipulación del habla natural como a la síntesis.

## BIBLIOGRAFÍA

- ATAL, B.S.: "Automatic Speaker recognition based on pitch contours", *J.A.S.A* 52, 1972, págs. 1687-1697.
- BORDONE-SACERDOTE, C. y G.G. SACERDOTE: "Some spectral properties of individual voices", *Acustica* 21, 1969, págs. 199-210.
- BROWN, R.: "An experimental study of the relative importance of acoustic parameters for auditory speaker recognition", *Language and Speech* 24, 4, 1981, págs. 295-310.

- CARRELL, T.D.: "Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification", *Research on Speech Perception Technical Report* (Indiana University Speech Laboratory), 5, 1984.
- COMPTON, A.J.: "Effects of filtering and vocal duration upon the identification of speakers aurally", *J.A.S.A.* 35, 1963, págs. 1748-1752.
- DODDINGTON, G.R.: "Speaker recognition. Identifying people by their voices", *Proc. IEEE* 73, 1985, págs. 1651-1664.
- DOHERTY, E. T.: "An evaluation of selected acoustic parameters for use in speaker identification", *Journal of Phonetics* 4, 1976, págs. 321-326.
- DOHERTY, E. y H. HOLLIEN: "Multiple factor speaker identification of normal and distorted speech", *Journal of Phonetics* 6, 1978, págs. 1-8.
- DOMMELEN, W. A. van: "The contribution of speech rhythm and pitch to speaker recognition", *Language and Speech* 30, 4, 1997, págs. 325-338.
- ESKENAZI, D., G. CHILDERS y D.M. Hicks: "Acoustic correlates of vocal quality", *Journal of Speech and Hearing Research* 33, 1990, págs. 298-306.
- FURUI, S.: "Research on individuality features in speech waves and automatic speaker recognition techniques", *Speech Communication* 5, 2, 1986, págs. 183-197.
- GOBL, C.: "A preliminary study of acoustic voice quality correlates", *STL-Quarterly Progress Status Report* 4, 1989, págs. 9-22.
- GLENN, J. W. y N. KLEINER: "Speaker identification based on nasal phonation", *J.A.S.A.* 43, 1968, págs. 368-372.
- HOLLIEN, H.: *The Acoustics of Crime. The New Science of Forensic Phonetics*, Nueva York, Plenum, 1990,.
- HOLLIEN, H.: "The profile approach to speaker identification", *Actes du XII<sup>ème</sup> Congrès International des Sciences Phonétiques* (Aix-en-Provence, 1991), Aix, Université de Provence, 1991, págs. 396-401.
- HOLLIEN, H. y MAJEWSKI, W.: "Speaker identification by long-term spectra under normal and distorted speech", *J.A.S.A.* 62, 1977, págs. 975-980.
- INGRAM, J.C.L., PRANDOLINI, R. y ONG, S.: "Formant trajectories as indices of phonetic variation for speaker identification", *Forensic Linguistics*, vol. 3-1, 1996, 129-145.
- ITOH, K. y SAITO, S.: "Effects of acoustical feature parameters of speech on perceptual identification of speaker", *IECE Trans.* Vol. J65-A, 1982, págs. 101-108.
- JOHNSON, C.C., HOLLIEN, H., y J.W. HICKS Jr.: "Speaker identification utilizing selected temporal speech features", *Journal of Phonetics* 12, 1984, págs. 319-327.
- KARLSONN, I.: "Glottal waveform parameters for different speaker types", *Proc. Speech '88, 7<sup>th</sup> FASE Symposium*, vol. 1, 1988, págs. 225-231.
- KUWABARA, H. y OHGUSHI, K.: "Contributions of vocal tract resonants frequencies and bandwidths to the personal perception of speech", *Acustica* 63, 1987, págs. 121-128.

- KUWABARA, H. y SAGISAKA, Y.: "Acoustic characteristics of speaker individuality: Control and conversion", *Speech Communication* 16, 1995, págs. 165-173.
- KUWABARA, H. y TAKAGI, T.: "Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method", *Speech Communication* 10, 1991, págs. 491-495.
- LARIVIERE, C.: "Contribution of fundamental frequency and formant frequencies to speaker identification", *Phonetica* 31, 1975, págs. 185-197.
- MARRERO, V., J. GIL y E. BATTANER, E.: "Inter-Speaker Variation in Spanish. An Experimental and Acoustic Preliminary Approach", *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 2003, págs. 703-706. [http://liceu.uab.es/~joaquim/phonetics/VILE/VILE\\_ICPhS03.pdf](http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_ICPhS03.pdf)
- MATSUMOTO, H. *et al.*: "Multidimensional representation of personal quality of vowels and its acoustical correlates", *IEEE Trans.* Vol. AU, 21, 1973, págs. 428-436.
- MILLER, J.E.: "Decapitation and recapitation: a study of voice quality", *J.A.S.A.* 36, 1964.
- ORTEGA, J., J. GONZÁLEZ y V. MARRERO: "AHUMADA: A large corpus in Spanish for speaker characterization and identification", *Speech Communication* 31 (2-3), 2000, págs. 255-264
- PITTAM, J.: "The long-term spectral measurement of voice quality as a social and personality marker: a review", *Language and Speech* 30, 1987, págs. 1-13.
- PRUZANSKY, S.: "Pattern matching procedure for automatic for automatic talker recognition", *J.A.S.A.* 35, 1963, págs. 354-358.
- SHEARME, J.N. y J.N. HOLMES: "An experiment concerning the recognition of voices", *Language and Speech* 2, 1959, págs. 123-131.
- STEVENS, K.: "Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds", *Proceedings of the 7th International Congress of Phonetic Sciences.*, Montreal, 1971, págs. 206-227.
- STEVENS, K. *et al.*: "Speaker identification and authentication: a comparison of spectrographic and auditory presentation of speech materials", *J.A.S.A.* 44, 1968, págs. 1596-1607.
- SU, L. S, LI, K. P. y FU, K. S.: "Identification of speakers by use of nasal coarticulation", *J.A.S.A.* 56, 1974, págs. 1867-1882.
- WOLF, J. J.: "Efficient acoustic parameters for speaker recognition", *J.A.S.A.* 51, 1972, págs. 2044-2056.