

Battaner, E., Carbó, C., Gil, J., Llisterri, J., Machuca, M. J., Madrigal, N., . . . Ríos, A. (2004). VILE: Estudio acústico de la variación inter e intralocutor en español. In *VI Congreso de Lingüística General*. Universidad de Santiago de Compostela, Santiago. 3-7 de mayo de 2004. (pp. 512-4). Santiago de Compostela: Área de Lingüística Xeral, Facultade de Filoloxía, Universidade de Santiago de Compostela.

http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE_CLG04.pdf

-un foro de debate sobre todo lo concerniente a esta disciplina y sobre todo

-un instrumento informático que facilita un aparato estadístico accesible, eficaz, rápido y actualizado.

Nace con el deseo de convertirse en una página de y para todos los investigadores en este campo con el compromiso de que los contenidos serán actualizados periódicamente a partir de las aportaciones recibidas.

VILE: Estudio acústico de la variación inter e intra locutor en español

ELENA BATTANER
Universidad Nacional de Educación a Distancia

CARME CARBÓ
Universitat Autònoma de Barcelona

JUANA GIL
Universidad Nacional de Educación a Distancia

JOAQUIM LLISTERRI
Universitat Autònoma de Barcelona

MARÍA JESÚS MACHUCA
Universitat Autònoma de Barcelona

VICTORIA MARRERO
Universidad Nacional de Educación a Distancia

CARME DE LA MOTA
Universitat Autònoma de Barcelona

MONTSERRAT RIERA
Universitat Autònoma de Barcelona

ANTONIO RÍOS
Universitat Autònoma de Barcelona

6/G
G
resúmenes de
comunicaciones

512

[Paneles de investigación]

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), cuya finalidad es el estudio acústico de la variación fonética en español, con objeto de aplicar los resultados al reconocimiento automático de locutor y a la práctica de la fonética forense (Battaner *et al.*, 2003; Marrero *et al.*, 2003).

1.- Objetivos

VILE pretende alcanzar tres objetivos: (a) caracterizar acústicamente 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; (b) obtener el conocimiento fonético necesario para la mejora de los sistemas de identificación o verificación automáticas del locutor; y (c) dotar a los espe-

cialistas en fonética forense de nuevos datos acústicos que permitan comparar, con un mayor grado de certeza, locutores dubitados e indubitados.

2.- Delimitación de los fenómenos fonéticos considerados

Para delimitar los fenómenos fonéticos que se consideran en el proyecto, se ha realizado una revisión bibliográfica de la que se desprende: (a) la dificultad de encontrar estudios de fonética acústica del español que aborden en profundidad la cuestión del reconocimiento del hablante; (b) la mención recurrente, en la bibliografía sobre otras lenguas, de parámetros como la frecuencia del fundamental (F_0), el espectro medio (LTAS) y la estructura formántica; y (c) la dificultad de establecer una prioridad entre estos parámetros.

3.- Selección de los materiales para el análisis

Teniendo en cuenta la necesidad de reutilizar recursos, se ha llevado también a cabo un análisis de los corpus orales existentes en español, valorando su utilidad en relación con los objetivos del proyecto. En general, los datos disponibles para el nivel segmental son ampliamente suficientes para un estudio como el propuesto. En cambio, en el ámbito suprasegmental la representación es menor, aunque corpus como EUROM 1 y su subconjunto MULTEXT contienen cierta información en el nivel oracional, limitada, sin embargo, a tareas de lectura. Sólo el corpus AHUMADA ofrece muestras de conversación espontánea, así como muestras de variación temporal, con textos leídos a tres ritmos diferentes. La principal carencia detectada sería un corpus de diálogos espontáneos orales, cuya obtención quedaría para un futuro proyecto.

4.- Análisis acústico

El análisis acústico de los fenómenos fonéticos anteriormente mencionados se está llevando a cabo en una muestra de materiales seleccionada a partir de los corpus que se han descrito. Se toman en consideración, además, otros parámetros como el valor medio y la desviación típica de la F_0 en el grupo fónico, las características espectrales de sonidos turbulentos como /s/ y, en la medida de lo posible, variables temporales relativas a la velocidad de habla.

El tratamiento de los resultados se abordará desde la perspectiva de la variación, tanto entre hablantes (variación interlocutores) como en un mismo hablante (variación intralocutor), considerando la relación entre ambos tipos de variabilidad y su relevancia para el reconocimiento automático de locutor y la fonética forense.

6/G
G
resúmenes de
comunicaciones

REFERENCIAS

BATTANER, E.- GIL, J.- MARRERO, V.- LLISTERRI, J.- CARBÓ, C.- MACHUCA, M.J.- de la MOTA, C. - RÍOS, A. (2003) "VILE: Estudio acústico de la variación inter e intralocutor en español", in *SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense*. Barcelona, 10 y 11 de abril de 2003. Barcelona: SEAF, Sociedad Española de Acústica Forense. pp. 59-70.

MARRERO, V.- GIL, J.- BATTANER, E. (2003) "Inter-Speaker Variation in Spanish. An Experimental and Acoustic Preliminary Approach", in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003. pp. 703-706.

Fundación CNSE para la supresión de las barreras de comunicación. Centro de Recursos "Juan Luis Marroquín"

Departamento de Investigación y Materiales
del Centro de Recursos para la Comunidad Sorda
"Juan Luis Marroquín" (FUNDACIÓN CNSE)

[Paneles de investigación]

6/6
G
resúmenes de
comunicaciones

514

¿QUÉ ES LA FUNDACIÓN CNSE?

Es una Entidad de ámbito estatal con personalidad jurídica propia, sin ánimo de lucro, constituida por la Confederación Estatal de personas Sordas de España (CNSE). Está inscrita en el Protectorado de Fundaciones del MECD. Su objetivo es fomentar la Supresión de Barreras de Comunicación que encuentran las personas sordas en la sociedad

DESTINATARIOS/AS

- Directos/as: Personas sordas, familias con miembros sordos, investigadores/as, profesionales relacionados con la Comunidad Sorda.
- Indirectos/as: Entidades del Movimiento Asociativo de Personas Sordas, entidades relacionadas con la Comunidad Sorda, entidades sociales, educativas, sanitarias, agentes sociales (sindicatos, empresas...).

ESTRUCTURA DE LA FUNDACIÓN CNSE

- Centro de Recursos para la Comunidad Sorda "Juan Luis Marroquín".
- Área de Empleo.
- Área de Nuevas Tecnologías.
- Áreas Transversales (Coordinación Intérpretes de LSE y Distribución de materiales).

VILE: ESTUDIO ACÚSTICO DE LA VARIACIÓN INTER E INTRALOCUTOR EN ESPAÑOL

ELENA BATTANER², CARME CARBÓ¹, JUANA GIL², JOAQUIM LLISTERRI¹, MARÍA MACHUCA¹, NATALIA MADRIGAL¹, VICTORIA MARRERO², CARME DE LA MOTA¹,
MONTserrat RIERA¹, ANTONIO RÍOS¹

¹ *Departamento de Filología Española, Universidad Autónoma de Barcelona*

² *Departamento de Lengua Española y Lingüística General, Universidad Nacional de Educación a Distancia*

1. OBJETIVOS

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), cuya finalidad es el estudio acústico de la variación fonética en español, con objeto de aplicar los resultados al reconocimiento automático de locutor y a la práctica de la fonética judicial (Battaner *et al.* 2003; Marrero *et al.* 2003)¹.

El proyecto pretende alcanzar tres objetivos: (a) caracterizar acústicamente 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; (b) obtener el conocimiento fonético necesario para la mejora de los sistemas de identificación o verificación automáticas del locutor; y (c) dotar a los especialistas en fonética judicial de nuevos datos acústicos que permitan comparar, con un mayor grado de certeza, locutores dubitados e indubitados.

2. DELIMITACIÓN DE LOS FENÓMENOS FONÉTICOS CONSIDERADOS

En el conjunto de la investigación en torno a la identificación del hablante se pueden diferenciar tres tipos de estudios: (a) los centrados en el reconocimiento visual de los espectrogramas (por ej., Tosi *et al.* 1972); (b) los que tratan el reconocimiento perceptivo del hablante (por ej., Pollack *et al.* 1954; Compton 1963; Stevens *et al.* 1968; Hollien *et al.* 1982; Kuwabara y Takagi 1991; Kreiman y Papcun 1991; Pisoni 1993); y (c) los que se ocupan del reconocimiento automático del habla. Este último grupo de trabajos constituye la principal base bibliográfica del proyecto VILE.

Con respecto a los parámetros acústicos descritos en la bibliografía, y aunque éstos se han agrupado en torno a distintos criterios (véase Doddington 1985 y Kuwabara 1995 para un primer acercamiento), parece necesario partir de la ya clásica enumeración de Stevens (1971). Este autor destacó ciertos factores, básicamente fisiológicos, que son susceptibles de emplearse en la identificación y discriminación de hablantes:

¹ Puede encontrarse más información sobre el proyecto en <http://liceu.uab.es/~joaquim/VILE.html>

(1) Factores referidos a la fuente:

Los dos aspectos referidos a la fuente señalados por Stevens son la *frecuencia fundamental* (F_0) *media* del locutor y la *forma de la onda glotal*, que difiere substancialmente de hablante a hablante.

(2) Factores referidos a los resonadores:

(2.1) *Frecuencias formánticas*. Si se consideran los valores medios de las frecuencias formánticas en un número de vocales suficientemente amplio, se obtiene un indicio de la longitud media del tracto vocal del hablante. Para este propósito es particularmente útil el valor medio del F3, puesto que este no cambia de modo notable de vocal a vocal y proporciona una indicación más precisa de la longitud del tracto vocal que el F1 y el F2: a medida que aumenta la longitud del tracto vocal, disminuye la frecuencia del formante.

(2.2) En cuanto a la *anchura de banda de los formantes*, resulta especialmente interesante comparar los de la vocal [i], que parecen variar poco en un mismo locutor y, sin embargo, presentan claras diferencias en el caso de varios locutores.

(2.3) *Sonidos turbulentos*. En el caso de la [s], se producen algunas diferencias intralocutor, pero claramente menos marcadas que las que se dan entre distintos locutores. Las resonancias de alta frecuencia del tracto vocal que se ven excitadas en la producción de un sonido turbulento como [s] dependen de la forma de las cavidades anteriores a la constricción y del modo en el que la lengua y el paladar se disponen en la parte inmediatamente posterior a la constricción.

(2.4) *Consonantes nasales*. Al igual que en el caso anterior, las consonantes nasales parecen presentar diferencias espectrales más marcadas entre hablantes que en el mismo locutor, si bien el propio Stevens (1971) reconoce que las muestras analizadas para estudiar la variabilidad de las nasales fueron muy próximas en el tiempo. Probablemente todas estas características se vieran alteradas con muestras más distanciadas temporalmente.

Ya con posterioridad a la temprana fecha del trabajo de Stevens, y a medida que se desarrollan las técnicas de análisis, algunos autores como Hollien (1990, 1991), aún manteniendo ciertos aspectos de la división de Stevens, añaden otros rasgos que podrían ser decisivos en la discriminación entre hablantes:

(3) El *espectro medio* (*LTAS, Long Time Averaged Spectrum*), especialmente útil con datos normalizados obtenidos en laboratorio. En el sistema de Hollien es el resultado del análisis de 40 parámetros extraídos de la señal. Otra de sus particularidades es que el espectro medio es muy resistente a los efectos del estrés al que pueda estar sometido el hablante.

(4) El *vector de los formantes vocálicos*. Es un parámetro muy importante para la identificación de los hablantes, porque el tracto vocal individual presenta estabilidad y porque estos rasgos son muy resistentes a la distorsión y a las interferencias (véase también el trabajo pionero de Ladefoged y Broadbent 1957). Hollien eligió dos parámetros para configurar su vector: las frecuencias centrales de los *tres primeros formantes*, que parecen ser muy reveladoras, especialmente si se estudian al menos tres vocales, [a, i, u] y la sílaba [na], y la distancia entre estos tres primeros formantes (F1/F2, F2/F3), que el hablante no puede alterar a voluntad (Tosi *et al.* 1972).

(5) El *vector temporal*. Pese a que no existen muchos trabajos sobre este factor (véase, por ejemplo, Johnson *et al.* 1984), en buena lógica el vector temporal puede contener informaciones muy importantes para la identificación. Las medidas empleadas para la definición de este vector incluyen: el tiempo total de habla, definido como el tiempo en milisegundos que lleva producir la emisión de un conjunto dado de sílabas; la proporción del tiempo de habla, definido como la medida del tiempo total durante el cual existe energía acústica en una emisión; la proporción de los intervalos de silencio; la velocidad de habla, obtenida a partir de las sílabas completadas durante un periodo de tiempo fijado; la proporción de la duración consonante / vocal, esto es, la relación entre el tiempo destinado a la producción de la consonante y el destinado a la vocal en la emisión de una combinación CV.

(6) El *vector de frecuencia fundamental* (F_0). En el trabajo de Hollien y colaboradores, este vector implica la medida de 30 parámetros diferentes, por lo que los resultados, según el autor, son más fiables que los ofrecidos por estudios anteriores.

Sin embargo, la casi unanimidad que se aprecia en la bibliografía con relación a los parámetros que deben considerarse en la identificación del locutor no es tal respecto a la prioridad que se les otorga. Esta prioridad puede ser atribuida a la *frecuencia fundamental* (Compton 1963; Wolf 1972; Matsumoto *et al.* 1973; Brown 1981; van Dommelen 1987), al *espectro medio* (Bordone-Sacerdote y Sacerdote 1969; Doherty 1976; Hollien y Majewski 1974, 1977; Furui 1986; Pittam 1987), o a la *estructura espectral*, bien sea la que reflejan las frecuencias formánticas absolutas (Shearme y Holmes 1959; Miller 1964; Carrell 1984; Kuwabara y Ohgushi 1987; Kuwabara y Takagi 1991), bien sea de las trayectorias formánticas (Ingram *et al.* 1996).

Por lo que se refiere a qué formantes o qué distancias entre formantes son las más informativas, tampoco hay coincidencia en las posturas. Por una parte, Hollien (1990), Kuwabara y Takagi (1991) y Kreiman y Papcun (1991) coinciden en la importancia de los tres primeros formantes, aunque Hollien se inclina por considerar relevantes la distancia entre F1 y F2 y entre F2 y F3 mientras que Kreiman y Papcun señalan únicamente la que se encuentra entre F1 y F2. Por otra, desde la perspectiva de Furui (1986) y de Ramón *et al.* (2000) la información más representativa de la individualidad del hablante estará localizada entre los 2.5 kHz. y los 3.5 kHz.

Otros autores otorgan la misma prioridad a la F_0 y a la estructura formántica (La Riviere 1975), a las variables temporales (Pruzansky 1963; Wolf 1972; Doherty y Hollien 1978; Brown 1981; Johnson *et al.* 1984), o al contorno tonal (Atal 1972; van Dommelen 1987). Frente a las opiniones anteriormente citadas, se encuentran las de investigadores como Gobl (1989) o Kuwabara y Sagisaka (1995) que no consideran factible establecer una prioridad, dado que la importancia de cada parámetro puede diferir de hablante a hablante y depende también de la naturaleza de las muestras analizadas.

El apartado 4 de este trabajo recoge el conjunto de parámetros y de variables que se ha establecido en el marco del proyecto VILE para el estudio de la variación inter e intralocutor en lengua española.

3. SELECCIÓN DE LOS MATERIALES PARA EL ANÁLISIS

Paralelamente al estudio bibliográfico, se ha llevado a cabo un análisis de los corpus orales existentes en español (Battaner *et al.* 2003). Con el fin de reutilizar recursos ya existentes, el análisis se centró en los siguientes corpus: Albayzín (Casacuberta *et al.* 1992), EUROM 1 (Chan *et al.* 1995), Multext (Campione y Véronis 1998), Gaudí, que incluye el subcorpus Ahumada (Ortega *et al.* 1998, 2000), y SpeechDat (Draxler *et al.* 1998).

En general, los datos disponibles para el nivel segmental son ampliamente suficientes para el estudio propuesto. Sin embargo, dadas las exigencias del proyecto y los resultados obtenidos tras la revisión bibliográfica, la representación del ámbito suprasegmental era un requisito imprescindible. Algunos corpus como EUROM 1 y su subconjunto Multext incluyen cierta variación entonativa, aunque limitada a tareas de lectura. En cambio, el corpus Ahumada ofrece muestras de conversación espontánea; este subconjunto, además, también posee una importante representación de la variación temporal en la tarea de lectura, ya que ésta se realizó a tres velocidades distintas y en tres sesiones independientes.

Por tal motivo, el corpus seleccionado para el análisis acústico es el de Gaudí-Ahumada, desarrollado por el Departamento de Ingeniería Audiovisual y Comunicaciones de la Escuela Universitaria de Ingenieros de Telecomunicación de la Universidad Politécnica de Madrid en colaboración con el Laboratorio de Acústica e Imagen del Servicio de Policía Judicial de la Dirección General de la Guardia Civil y el Departamento de Lengua Española de la Universidad Nacional de Educación a Distancia (Ortega *et al.* 1998, 2000). Ahumada recoge los siguientes estilos de habla, procedentes de 224 locutores masculinos y 231 locutores femeninos:

- (1) Lectura de 10 frases fonéticamente equilibradas: 409 fonemas.
- (2) Lectura de un texto de 179 palabras, también equilibrado (712 fonemas), a tres velocidades distintas (normal, rápida y lenta).
- (3) Lectura de un texto específico para cada locutor.
- (4) Habla espontánea (explicación sobre un tema libre no planificado con antelación o descripciones de una lámina ya preparada –cuadros o dibujos–) durante más de un minuto.

Para cada uno de estos estilos de habla se realizaron seis grabaciones diferentes: tres sesiones en la sala de grabación con micrófono y tres sesiones por teléfono. Las sesiones microfónicas se grabaron con un intervalo de 20 a 40 días entre ellas.

El material de análisis seleccionado para el estudio de la variación interlocutor e intralocutor del proyecto VILE está constituido por las grabaciones microfónicas correspondientes a muestras de habla de dos tipos: una, en tarea de lectura y contextos controlados (apartado 2, texto de 179 palabras, a velocidad normal); la otra, a las tres sesiones de habla espontánea.

El análisis se realiza a partir de una selección de las grabaciones de 30 locutores masculinos. Se han eliminado 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.

4. ANÁLISIS ACÚSTICO

De acuerdo con las tendencias que se han puesto de manifiesto en la revisión bibliográfica presentada en el apartado 1, los parámetros acústicos considerados en el proyecto están relacionados con la fuente, con la acción de los resonadores y con el vector temporal, tal como se expone a continuación.

(1) Parámetros referidos a la fuente.

Se segmenta el corpus de lectura o de habla espontánea en grupos fónicos (segmento comprendido entre dos silencios) y de cada uno de ellos se extrae el valor medio de la F_0 y su fluctuación (desviación estándar). En la figura 1 puede observarse un ejemplo de segmentación de un grupo fónico y el resultado de la extracción automática del valor medio y de la desviación estándar de la F_0 (ventana superior izquierda).

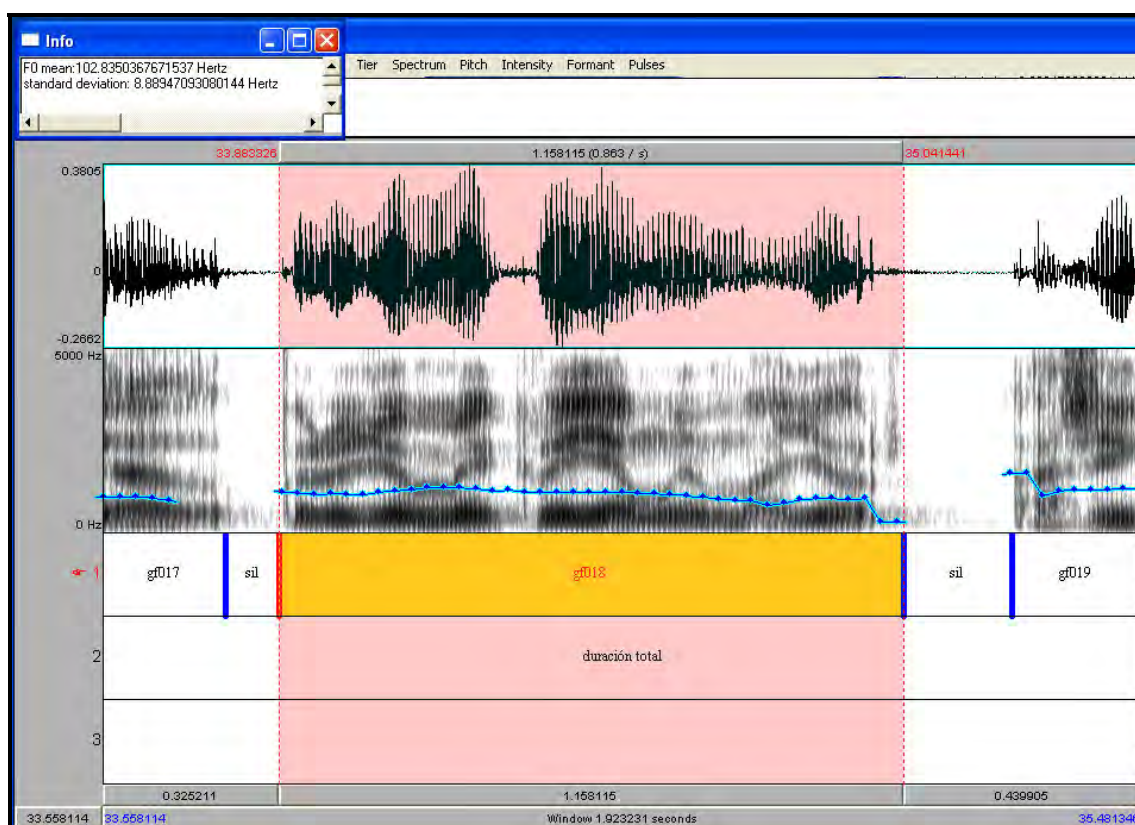


Figura 1. Segmentación y de extracción automática de valores de F_0 en el grupo fónico *pero yo me creía muy mayor* procedente del corpus de lectura.

(2) Parámetros referidos a los resonadores.

Se segmentan en el corpus de lectura las vocales tónicas y átonas precedidas de oclusiva sorda o /s/ y seguidas de oclusiva sorda o /s/, cuya estructura silábica sea CV \$

BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2004) "VILE: Estudio acústico de la variación inter e intralocutor en español", in *6º Congreso de Lingüística General*. Santiago de Compostela, 3-7 de mayo de 2004. Área de Lingüística Xeral, Universidade de Santiago de Compostela.

C. También se segmentan 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 analizan el valor de la F_0 , las frecuencias centrales de los cuatro primeros formantes y los anchos de banda (véase la figura 2). Posteriormente, a partir de estos valores, se extrae la distancia entre los valores de frecuencia de F_0 , F1, F2, F3 y F4. En los segmentos fricativos alveolares sordos se tiene en cuenta el pico de mayor intensidad que aparece en la banda de frecuencia entre 0 y 8 kHz.

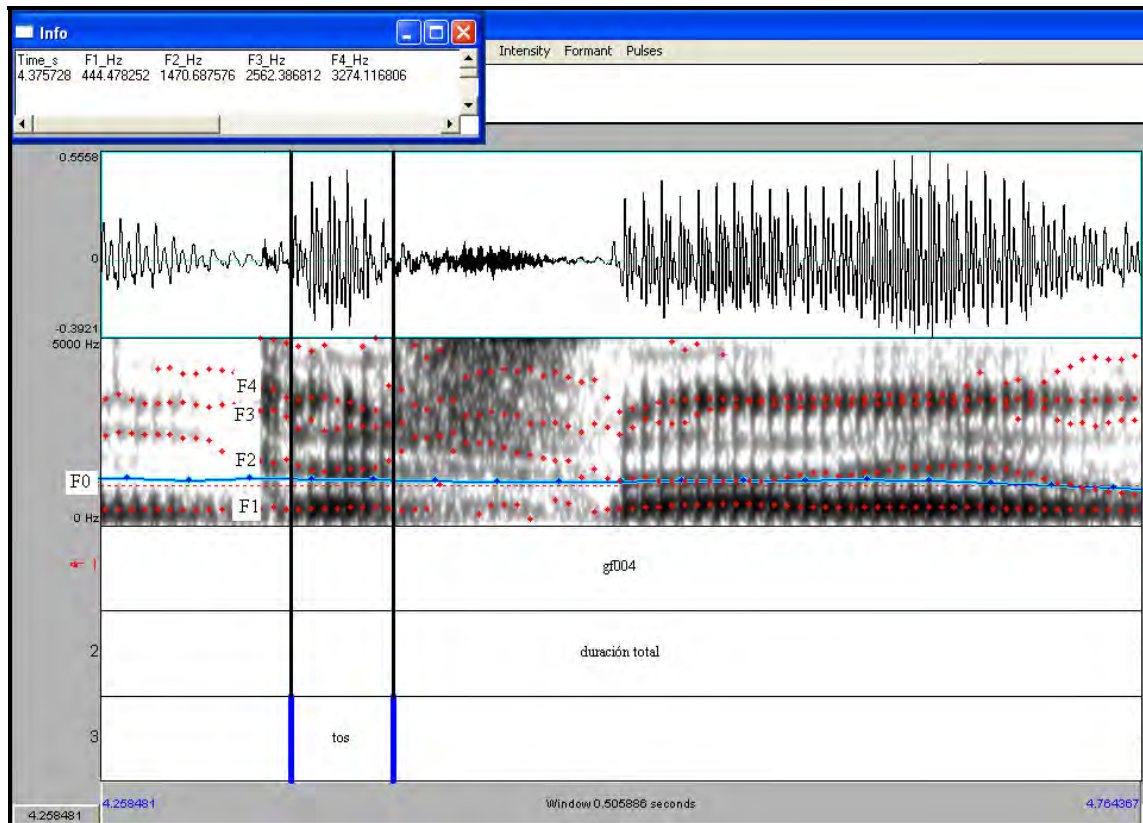


Figura 2. Segmentación y extracción automática de los valores de los cuatro primeros formantes de la vocal [o] del grupo fónico *ese viento suave vuela* procedente del corpus de lectura.

(3) Parámetros temporales.

En cuanto a los parámetros temporales, se consideran el tiempo total de habla, la proporción de habla, la proporción de silencios, la velocidad de elocución y la velocidad de articulación.

(3.1) El tiempo total de habla corresponde a la duración total de cada una de las sesiones analizadas.

(3.2) La proporción de habla se define como la duración de todos los grupos fónicos de una determinada sesión dividida entre la duración total de dicha sesión.

BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2004) "VILE: Estudio acústico de la variación inter e intralocutor en español", in *6º Congreso de Lingüística General*. Santiago de Compostela, 3-7 de mayo de 2004. Área de Lingüística Xeral, Universidade de Santiago de Compostela.

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

(3.3) La proporción de silencios se calcula dividiendo la duración total de los silencios de una determinada sesión por la duración total de dicha sesión.

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

(3.4) La velocidad de elocución consiste en el tiempo total de habla dividido entre el número de sílabas de cada una de las sesiones analizadas.

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

(3.5) La velocidad de articulación 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 utiliza para el análisis de todos los parámetros estudiados, mientras que el corpus de habla espontánea se emplea únicamente para el estudio de los parámetros temporales y de los parámetros relacionados con la fuente.

Finalmente, la herramienta seleccionada para el análisis acústico del corpus ha sido la versión 4.1 del programa *Praat* (<http://www.praat.org>), desarrollado por Paul Boersma y David Weenink en el Instituto de Fonética de la Universidad de Ámsterdam.

5. CONSIDERACIONES FINALES

Una vez finalizado el análisis acústico de los parámetros descritos en el apartado 4, el tratamiento estadístico de los resultados se abordará desde la perspectiva de la variación, tanto entre hablantes (variación interlocutores) como en un mismo hablante (variación intralocutor). Tal como se ha determinado en los objetivos del proyecto VILE, se tomará en consideración la relación entre ambos tipos de variabilidad, con el fin de evaluar la relevancia de los distintos elementos analizados para el reconocimiento automático de locutor y para la fonética judicial.

Un experimento preliminar llevado a cabo por Marrero *et al.* (2003) ha permitido realizar algunas observaciones en lo que se refiere a las variaciones entre hablantes en un texto leído. Un primer resultado destacable es que se aprecian variaciones en el límite inferior de [s] superiores a 1.000Hz para el mismo contexto fonético. En segundo lugar, se obtuvieron también diferencias entre locutores que superaban los 200 Hz en el F2 y los 100 Hz en el F1 de las vocales.

En este mismo experimento se abordó también el análisis experimental de los parámetros fonéticos que inciden en la identificación de los locutores. Los primeros resultados mostraron que el primero de ellos parece ser el F₀ y el segundo el LTAS. Por otra parte, mientras que un hablante con una elevada variación en los valores de

frecuencia de F2 y una variación reducida en los valores de F1 se identificó con un alto porcentaje de aciertos, otro con poca variación en F2 y gran variabilidad en el F1 presentó serios problemas de identificación. Cabe añadir, además, que la fricativa alveolar sorda [s] no constituyó, en este experimento, un indicio excesivamente relevante para la identificación del locutor.

Las discrepancias observadas entre los diversos autores que han abordado la variación fonética con objeto de determinar los parámetros acústicos que aportan datos respecto a la individualidad de un hablante, junto con la escasez de este tipo de estudios para el español, constituyen dos argumentos que justifican la necesidad de llevar a cabo un proyecto como VILE. Aunque, por el momento, la labor que se está realizando puede considerarse como investigación básica, en una segunda etapa se contemplará la viabilidad de incorporar el conocimiento fonético obtenido a sistemas de identificación y verificación automática del locutor y se explorará también su uso en el contexto de la fonética judicial.

REFERENCIAS BIBLIOGRÁFICAS

- ATAL, B.S. (1972): "Automatic Speaker recognition based on pitch contours". *Journal of the Acoustical Society of America*. 52. 1687-1697.
- BATTANER, E., J. GIL, V. MARRERO, J. LLISTERRI, C. CARBÓ, M.J. MACHUCA, C. DE LA MOTA & A. RÍOS (2003) "VILE: Estudio acústico de la variación inter e intralocutor en español". *SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense*. Barcelona, 10 y 11 de abril de 2003. Barcelona: SEAF, Sociedad Española de Acústica Forense. 59-70. http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_SEAF03.pdf [Consulta: 16/04/2004]
- BORDONE-SACERDOTE, C. & G.G. SACERDOTE (1969): "Some spectral properties of individual voices". *Acustica*. 21. 199-210.
- BROWN, R. (1981): "An experimental study of the relative importance of acoustic parameters for auditory speaker recognition". *Language and Speech*. 24: 4. 295-310.
- CAMPIONE, E. & J. VÉRONIS (1998): "A Multilingual Prosodic Database". *ICSLP'98, Proceedings of the 5th International Conferenc on Spoken Language Processing*. 30th November-4th December 1998, Sydney, Australia. 7. 3163-3166. <http://www.up.univ-mrs.fr/~veronis/pdf/1998icslp-database.pdf> [Consulta: 16/04/2004]
- CARRELL, T.D. (1984): "Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification". *Research on Speech Perception Technical Report* (Indiana University Speech Laboratory). 5.
- CASACUBERTA, F., R. GARCÍA, J. LLISTERRI, C. NADEU, J.M. PARDO & A. RUBIO, A. (1992): "Desarrollo de corpus para investigación en tecnologías del habla (Albayzín)". *Procesamiento del Lenguaje Natural*. 12. 35-42. <http://www.sepln.org/revistaSEPLN/revista/12/12-Pag35.pdf> [Consulta: 16/04/2004]
- COMPTON, A.J. (1963): "Effects of filtering and vocal duration upon the identification of speakers aurally". *Journal of the Acoustical Society of America*. 35. 1748-1752.
- CHAN, D., A. FOURCIN, D. GIBBON, B. GRANSTRÖM, M. HUCKVALE, G. KOKKINAKIS, J. KVALE, L. LAMEL, B. LINDBERG, A. MORENO, J. MOUROPOULOS, F. SENIA, I. TRANCOSO, C. VELD & J. ZEILIGER (1995): "EUROM- A Spoken Language Resource for the EU". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*. Madrid, Spain, 18-21 September, 1995. 1. 867-870.
- DODDINGTON, G.R. (1985): "Speaker recognition. Identifying people by their voices". *Proceedings of the IEEE*. 73. 1651-1664.

BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2004) "VILE: Estudio acústico de la variación inter e intralocutor en español", in *6º Congreso de Lingüística General*. Santiago de Compostela, 3-7 de mayo de 2004. Área de Lingüística Xeral, Universidade de Santiago de Compostela.

- DOHERTY, E. & H. HOLLIEN (1978): "Multiple factor speaker identification of normal and distorted speech". *Journal of Phonetics*. 6. 1-8.
- DOHERTY, E.T. (1976): "An evaluation of selected acoustic parameters for use in speaker identification". *Journal of Phonetics*. 4. 321-326.
- DOMMELEN, W.A. VAN (1997): "The contribution of speech rhythm and pitch to speaker recognition". *Language and Speech*. 30: 4. 325-338.
- DRAXLER, C.H.- V. VAN DEN HEUVEL & H. TROPF (1998): "SpeechDat Experiences in Creating Large Multilingual Speech Databases for Teleservices". *Proceedings of the First International Conference on Language Resources and Evaluation*. May 28 - 30, 1998, Granada, Spain. European Language Resources Association. I. 361-366.
- FURUI, S. (1986): "Research on individuality features in speech waves and automatic speaker recognition techniques". *Speech Communication*. 5:2. 183-197.
- GOBL, C. (1989): "A preliminary study of acoustic voice quality correlates". *STL-Quarterly Progress Status Report*. 4. 9-22.
- HOLLIEN, H. & W. MAJEWSKI (1977): "Speaker identification by long-term spectra under normal and distorted speech". *Journal of the Acoustical Society of America*. 62. 975-980.
- HOLLIEN, H. (1990): *The Acoustics of Crime. The New Science of Forensic Phonetics*. Plenum: Nueva York.
- HOLLIEN, H. (1991): "The profile approach to speaker identification". *Actes du XII^{ème} Congrès International des Sciences Phonétiques (Aix-en-Provence, 1991)*. Université de Provence: Aix-en-Provence. 396-401.
- HOLLIEN, H., W. MAJEWSKI & E.T. DOHERTY (1982): "Perceptual identification of voice under normal, stress and disguise speaking conditions". *Journal of Phonetics*. 10. 139-148.
- INGRAM, J.C.L., R. PRANDOLINI & S. ONG (1996): "Formant trajectories as indices of phonetic variation for speaker identification". *Forensic Linguistics*. 3:1. 129-145.
- JOHNSON, C.C., H. HOLLIEN & J.W. HICKS JR. (1984): "Speaker identification utilizing selected temporal speech features". *Journal of Phonetics*. 12. 319-327.
- KREIMAN, J. & G. PAPCUN (1991): "Comparing discrimination and recognition of unfamiliar voices". *Speech Communication*. 10. 265-275.
- KUWABARA, H. & K. OHGUSHI (1987): "Contributions of vocal tract resonants frequencies and bandwidths to the personal perception of speech". *Acustica*. 63. 21-128.
- KUWABARA, H. & T. TAKAGI (1991): "Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method". *Speech Communication*. 10. 491-495.
- KUWABARA, H. & Y. SAGISAKA (1995): "Acoustic characteristics of speaker individuality: Control and conversion". *Speech Communication*. 16. 165-173.
- LADEFOGED, P. & D.E. BROADBENT (1957): "Information conveyed by vowels". *Journal of the Acoustical Society of America*. 29. 98-104.
- LARIVIERE, C. (1975): "Contribution of fundamental frequency and formant frequencies to speaker identification". *Phonetica*. 31. 185-197.
- MAJEWSKI, W. & H. HOLLIEN (1974): "Euclidean distance between long-term spectra as a criterion for speaker identification". *Proceedings of the Speech Communication Seminar, Estocolmo*. 303-310.
- MARRERO, V., J. GIL & E. BATTANER (2003): "Inter-Speaker Variation in Spanish. An Experimental and Acoustic Preliminary Approach". *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, 3-9 August 2003. 703-706.
http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_ICPhS03.pdf
 [Consulta: 16/04/2004]
- MATSUMOTO, H., S. HIKI, T. STONE & T. NIMURA (1973): "Multidimensional representation of personal quality of vowels and its acoustical correlates". *IEEE Transactions on Audio and Electroacoustics*. AU-21. 428-436.
- MILLER, J.E. (1964): "Decapitation and recapitation: a study of voice quality". *Journal of the Acoustical Society of America*. 36.
- ORTEGA, J., J. GONZÁLEZ Y V. MARRERO (2000): "AHUMADA: A large corpus in Spanish for speaker characterization and identification". *Speech Communication*. 31: 2-3. 255-264.

- <http://www.atvs.diac.upm.es/publicaciones/docs/Ort00a.pdf> [Consulta: 16/04/2004]
- ORTEGA, J., J. GONZÁLEZ, V. MARRERO, J.J. DÍAZ, J. LUCENA Y J.A.G. SÁNCHEZ (1998): "AHUMADA: A Large Corpus in Spanish for Speaker Identification and Verification". *Proceedings of ICASSP-98*. IEEE International Conference on Acoustics Speech and Signal Processing. 773-776. <http://www.atvs.diac.upm.es/publicaciones/docs/Ort98b.pdf> [Consulta: 16/04/2004]
- PISONI, D.B. (1993): "Long-term memory in speech perception: some new findings on talker variability, speaking rate and perceptual learning". *Speech Communication*. 13. 109-125.
- PITTAM, J. (1987): "The long-term spectral measurement of voice quality as a social and personality marker: a review". *Language and Speech*. 30. 1-13.
- POLLACK, I., J.M. PICKETT & W.H. SUMBY (1954): "On the identification of speakers by voice". *Journal of the Acoustical Society of America*. 26. 403-412.
- PRUZANSKY, S. (1963): "Pattern matching procedure for automatic for automatic talker recognition". *Journal of the Acoustical Society of America*. 35. 354-358.
- RAMÓN, J.L., V. GARCERÁN, M. CANTERAS & J.A. SÁNCHEZ MOLERO (2000): "Parametric speaker verification with linear prediction and cepstrum using the envelope of voice and discriminant analysis". *Actas del I Congreso de la Sociedad Española de Acústica Forense*. Madrid, 5 y 6 de octubre de 2000. 169-181.
- SHEARME, J.N. & J.N. HOLMES (1959): "An experiment concerning the recognition of voices". *Language and Speech* 2, 123-131.
- STEVENS, K. (1971): "Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds". *Proceedings of the Seventh International Congress of Phonetic Sciences*. ed. por R. Charbonneau & A. Rigault. The Hague: Mouton. 206-232.
- STEVENS, K.N., C.E. WILLIAMS, J.R. CARBONELL & B. WOODS (1968): "Speaker identification and authentication: a comparison of spectrographic and auditory presentation of speech materials". *Journal of the Acoustical Society of America*. 44. 1596-1607.
- TOSI, O., OYER, H., LASHBROOK, W., PEDREY, C., NICHOL, J. Y NASH, W. (1972): "Experiment on voice identification". *Journal of the Acoustical Society of America* 51, 2030-2043.
- WOLF, J.J. (1972): "Efficient acoustic parameters for speaker recognition". *Journal of the Acoustical Society of America*. 51. 2044-2056.