

VILE: Análisis estadístico de los parámetros relacionados con el grupo de entonación

ALBALÁ, M.J.*; BATTANER, E.**; CARRANZA, M.+;
DE-LA-MOTA, C.+; GIL, J.*; LLISTERRI, J.+;
MACHUCA, M.J.+; MADRIGAL, N.+;
MARQUINA, M.+; MARRERO, V.#; RIERA, M.+; RÍOS, A+
* *Consejo Superior de Investigaciones Científicas*
** *Universidad Rey Juan Carlos, Madrid*
+ *Universitat Autònoma de Barcelona*
Universidad Nacional de Educación a Distancia

1. Introducción

El proyecto VILE¹ aborda el estudio acústico de la variación inter e intralocutor en español para aplicar los resultados a la mejora de sistemas automáticos de reconocimiento, identificación y verificación del hablante y a las prácticas propias de la fonética forense. El trabajo realizado hasta ahora se ha centrado en el análisis de los parámetros acústicos relacionados con los elementos segmentales y suprasegmentales del habla para establecer la variación fonética que se da en las producciones sonoras de un mismo hablante, comparándolas con las diferencias que se encuentran en la realización de las mismas unidades fonéticas por diferentes locutores (Battaner *et al.*, 2007).

Los datos estudiados se han extraído de 30 locutores masculinos del corpus Ahumada (Ortega, González y Marrero, 2000); cada uno de ellos fue grabado en tres sesiones y en dos estilos de habla (lectura y habla espontánea). En el primer caso la tarea consistía en la lectura de un texto controlado de 179 palabras y en el segundo caso los sujetos hablaban durante un minuto sobre un tema cotidiano (vacaciones, deportes, coches, cine...). Se ha llevado a cabo un tratamiento estadístico en el que se tienen en cuenta parámetros frecuenciales y temporales. En la presente contribución se analizan los resultados obtenidos para los grupos de entonación. Mientras que en lectura los grupos de entonación analizados corresponden a fragmentos comprendidos entre dos pausas silenciosas, en el habla espontánea no sólo se han considerado los fragmentos limitados por pausas

¹ Variación inter e intralocutor en español (VILE). En línea:
<http://liceu.uab.cat/~joaquim/VILE.html>

silenciosas (que eran poco frecuentes), sino también los limitados por inflexiones tonales, pausas sonoras, y voz rota (*creaky voice*) (Quilis *et al.*, 1993).

En los apartados siguientes se presentan los resultados relacionados con el grupo de entonación: número de sílabas, valor medio y desviación estándar de la F_0 , y parámetros temporales.

2. Análisis estadístico de los datos

En este apartado se presentan los datos analizados para las tres sesiones anteriormente mencionadas, en las que los locutores emplearon dos estilos de habla: lectura y habla espontánea. Las variables fonéticas consideradas para el análisis han sido las siguientes: número de sílabas fonéticas en cada grupo de entonación, el valor medio y la desviación estándar de los valores de la F_0 de cada grupo, y los parámetros relacionados con el *tempo* (la duración de cada grupo, duración total de cada sesión y duración de las pausas sordas y de las pausas sonoras o llenas en cada sesión).

2.1. Número de sílabas en el grupo de entonación

Para evitar los problemas que pueden causar los fenómenos de reducción en habla espontánea a la hora de contabilizar el número de sílabas dentro del grupo de entonación se han considerado las sílabas que articula el hablante (sílabas fonéticas). A modo de ejemplo, una de las frases del corpus constituida por nueve sílabas fonológicas (*era el puente de Santiago*) queda reducida a siete sílabas fonéticas (*eral puente Santiago*), ya que el hablante ha elidido alguna de las sílabas.

Para establecer qué grupo de entonación es más frecuente en cuanto al número de sílabas en cada uno de los estilos analizados se ha llevado a cabo la prueba de χ^2 . El nivel de significación ($p=0.000$) indica que este parámetro es muy significativo si consideramos los dos tipos de habla analizados.

Los resultados muestran que de 3411 grupos analizados en habla espontánea, el que presenta una frecuencia de aparición más alta es el de 5 sílabas, aunque, como se puede observar en la figura 1, los grupos que tienen entre 4 y 9 sílabas también son muy frecuentes. Por el contrario, en lectura se observa una reducción del número de sílabas. Si se consideran los 3135 grupos de entonación analizados, los grupos que presentan una

frecuencia de aparición más alta son los que están constituidos por 3 sílabas, seguidos de los de 4 y 6 sílabas (véase fig.1).

Por otro lado, cabe destacar que en los dos estilos de habla pueden realizarse grupos muy cortos, constituidos por una sola sílaba, mientras que los grupos con un mayor número de sílabas aparecen sólo en habla espontánea (56 sílabas frente a 32 como máximo en lectura). En la figura 1 se puede observar cómo la mayor variación del número de sílabas por grupo de entonación se encuentra en habla espontánea, ya que en la parte derecha del gráfico sólo se muestran las barras de color negro.

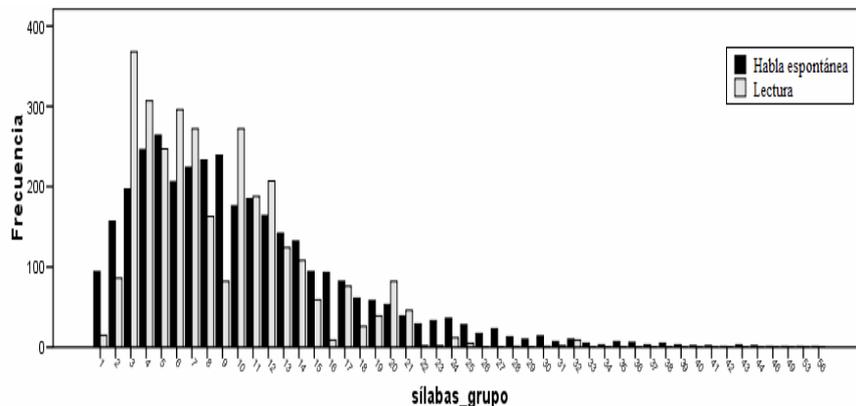


Figura 1. Número de sílabas por grupo de entonación para cada estilo de habla analizado.

2.2. Valor medio y desviación estándar de la F₀ en el grupo de entonación

Para cada grupo de entonación se han analizado los valores medios de frecuencia de la F₀ y la desviación estándar de este parámetro.

Los resultados muestran que el valor de la F₀ es más alto en lectura que en habla espontánea. Teniendo en cuenta todas las sesiones y todos los locutores, en lectura el valor de la F₀ es de 119 Hz y en habla espontánea, de 112 Hz. Este parámetro sirve para identificar los locutores en los dos tipos de habla analizados (p=0.000).

Ahora bien, si se consideran los locutores individualmente en cada una de las tres sesiones analizadas, se observa una tendencia general al mismo comportamiento en los dos tipos de habla.

Sin embargo, la desviación estándar de la F₀ no es un parámetro significativo para diferenciar los estilos de habla analizados (p=0.3). Los datos son muy similares en los dos estilos de habla para todos los locutores

y todas las sesiones (12, 8 Hz en lectura y 12, 7 Hz en habla espontánea). Si se analiza individualmente a cada uno de los hablantes, la desviación de los valores de la F_0 tiende a presentar el mismo comportamiento en las tres sesiones de cada uno de los dos estilos de habla, excepto en cinco de los 30 sujetos.

2.3. Parámetros temporales

Los parámetros temporales analizados son la duración de cada grupo de entonación y, en el caso de la lectura, la duración total de cada sesión. Además, se han considerado las pausas haciendo una distinción entre las sordas (silencios) y las sonoras o llenas. La presencia de una pausa puede deberse a varias causas: la necesidad de respirar, la organización de estructura prosódica o posibles vacilaciones.

Su aparición parece estar condicionada por el estilo de habla. En el habla espontánea, por ejemplo, se producen tanto pausas silenciosas como disfluencias. Las pausas llenas, raras en situaciones de laboratorio, son frecuentes, en cambio, en situaciones comunicativas naturales (Schriberg, 1999).

Esto puede relacionarse con el hecho observado de que existe una relación entre el uso de pausas y la necesidad de planificar léxica y sintácticamente el discurso o de ejecutarlo con propiedad (véase Duez, 1982; Swerts, Wichmann y Beuny, 1996 entre otros).

Los resultados muestran que, si se tratan conjuntamente las tres sesiones de cada estilo de habla, la duración de los grupos de entonación es menor en lectura (1,3 s) que en habla espontánea (1,6 s), y el valor de significación obtenido muestra que este es un parámetro relevante para diferenciar los dos tipos de habla ($p=0.000$).

La duración total de cada una de las sesiones sólo se ha calculado en lectura, ya que en habla espontánea la grabación se interrumpía exactamente al cabo de un minuto, aunque el hablante no hubiera finalizado el discurso. En la tabla 1 se muestran los valores medios de duración para cada sesión teniendo en cuenta todos los locutores.

Se puede observar una tendencia a que la duración sea mayor en la primera sesión que en las siguientes, algo que podría explicarse por el conocimiento del texto a partir de la primera grabación.

	N	Media	Desviación típica
1ª sesión	30	60,8	8,1
2ª sesión	30	58,1	7,3
3ª sesión	30	57,9	6,4
Total	90	59	7,4

Tabla 1. Valores medios y desviación típica (en segundos) de la duración de cada sesión de lectura teniendo en cuenta todos los locutores. (N = número de locutores)

La presencia de pausas es otro parámetro significativo para diferenciar los dos estilos analizados: el 60% de los locutores no realiza ninguna pausa sonora en la lectura del texto y, del 40% que realiza pausas sonoras, la de mayor duración, posee un valor de 0,6 segundos.

En cuanto a los silencios, cabe decir que en lectura estos duran más en la primera sesión (13 s) que en las siguientes (11 s); si los hablantes ya conocen el corpus de lectura, pueden anticipar el texto y las pausas se reducen. Esta diferencia entre las sesiones es muy significativa ($p=0.000$) en lectura, pero no en habla espontánea. La primera vez que lee el texto, el hablante no está familiarizado con él, mientras que en todas las sesiones en que interviene de forma espontánea las posibilidades de anticipar la colocación de pausas en el discurso son las mismas.

Si se considera cada locutor individualmente, el valor de significación obtenido ($p=0.000$) muestra que la duración total de los silencios es un parámetro para caracterizar la variación interlocutor, tanto en lectura como en habla espontánea. A modo de ejemplo, la duración de las pausas en lectura va desde 7 s en el locutor que presenta el valor más bajo hasta 20 s en el que presenta el más elevado; en habla espontánea, el rango oscila entre 8,7 s en el hablante que ofrece valores más bajos y 23,8 s en el que presenta los más altos.

En lo que se refiere a la duración de las pausas sonoras, tanto en habla espontánea como en lectura se va reduciendo desde la primera sesión a la tercera y la diferencia es también muy significativa. Como ya se ha mencionado, la inserción de pausas sonoras es una característica propia del habla espontánea, tal como lo corrobora el hecho de que sólo tres locutores no realizan ninguna pausa de este tipo.

En la figura 2 se pueden observar las duraciones de las pausas sordas y sonoras en cada una de las sesiones y en los dos estilos de habla analizados.

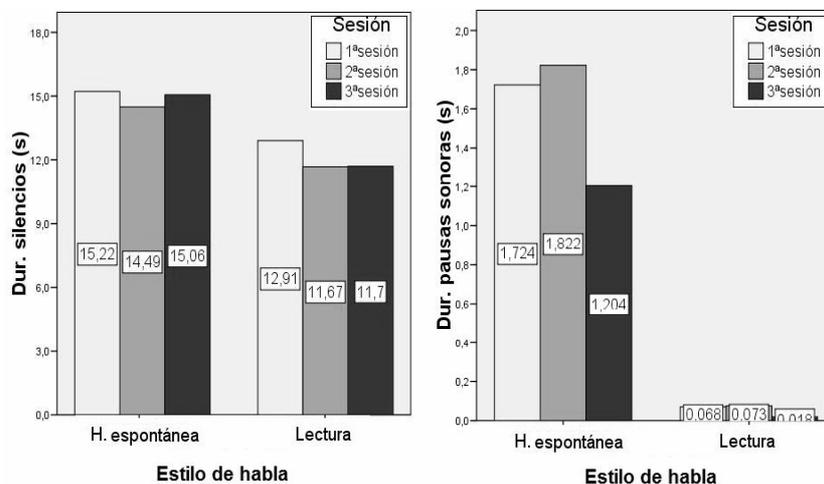


Figura 2. Duración de las pausas sordas y sonoras en cada una de las sesiones analizadas para los diferentes estilos de habla.

3. Conclusiones

Los resultados presentados en este trabajo ayudan a sistematizar las características relacionadas con un tipo de estilo de habla determinado, además de indicar qué parámetros relacionados con el grupo de entonación determinan la variación acústica interlocutor e intralocutor.

El habla espontánea se caracteriza por una mayor dispersión en cuanto al número de sílabas por grupo de entonación, aunque se ha observado que el más frecuente es el de cinco sílabas; una mayor duración de los grupos de entonación y la presencia de pausas sonoras. La lectura se caracteriza por una menor duración del grupo de entonación y la ausencia casi total de pausas sonoras. Sin embargo, los resultados de este experimento muestran una reducción en el número de sílabas de los grupos de entonación más frecuentes, tanto en habla espontánea como en lectura.

Las variaciones entre los locutores analizados vienen determinadas desde el punto de vista acústico por el valor medio de la F_0 y por la duración total de los silencios. Por el contrario, la desviación estándar de la F_0 no es un parámetro relevante a la hora de distinguir a los sujetos entre sí.

Por último, el parámetro que determina la variación en un mismo locutor entre las tres sesiones grabadas, tanto en habla espontánea como en lectura, es la duración de los silencios. Además, debe tenerse en cuenta el tiempo empleado por el hablante en la lectura del texto, con una duración diferente para cada una de las sesiones. El valor medio de la F_0 y la duración de las

pausas sonoras han resultado parámetros relevantes en la variación intralocutor en habla espontánea.

Referencias

- BATTANER, E.; CARBÓ, C.; GIL, J.; LLISTERRI, J.; MACHUCA, M.J.; MADRIGAL, M.; MARRERO, V.; DE LA MOTA, C.; RIERA, M. Y A. RÍOS (2007): "VILE: estudio acústico de la variación inter e intralocutor en español". En *Actas do 3º Congreso Internacional de Fonética Experimental*. Santiago de Compostela: Xunta de Galicia, págs. 157-167.
- ORTEGA, J.; J. GONZÁLEZ Y V. MARRERO (2000): "AHUMADA: a large corpus in Spanish for speaker characterization and identification", *Speech Communication* 31, 2-3, págs. 255-264.
- DUEZ, D (1982) "Silent and non-silent pauses in three speech styles", en *Language and Speech*, vol. 25, part 1, pp. 11-28.
- SWERTS, M., A. WICHMANN & R.-J. BEUNY (1996): "Filled pauses as markers of discourse structure", en ICSLP-1996, 1033-1036, <http://www.ee.columbia.edu/ln/labrosa/proceeds/icslp/1996/vol2/318/a318.pdf>
- SHRIBERG, E. (1999) "Phonetic consequences of speech disfluency", en *Proceedings of the International Congress of Phonetic Sciences (ICPhS-99)*, Vol. I, pp. 619-622 San Francisco, <http://www.speech.sri.com/papers/icphs99-disfluencies.pdf>
- QUILIS, A.- CANTARERO, M.- ESGUEVA, M. (1993) "El grupo fónico y el grupo de entonación en español hablado", *R.F.E.* LXXIII, pp. 55-65.