

# ANÁLISIS INSTRUMENTAL DE LA DURACIÓN DE LAS VOCALES EN ESPAÑOL

*Mary Hely Cuenca*

In this paper we describe an experimental analysis of vowel duration in Spanish. Some factors that seem relevant in other studies for characterizing vowel duration were analyzed here using speech data from one speaker by means of carrier sentences. These factors are: vowel identity, lexical stress, open versus closed syllables, final versus non-final position in the word, number of syllables and segments in the word, number of segments in the syllable, syllable structure, syllable duration, word duration and the effect of the postvocalic consonant. Our results show that Spanish vowels have intrinsic duration and there are some contextual factors, as final position in the word, stressed syllables, number of syllables and segments in the word, word duration, that influence vowel lengthening or shortening. For open versus closed syllables, number of segments in the syllable, syllable structure, syllable duration and manner of postvocalic consonant, significant differences were not found.

## 1. Introducción

El objetivo fundamental de este trabajo consiste en llevar a cabo una aproximación a la caracterización de la duración de las vocales en español. Para ello resulta necesario, en primer lugar, determinar qué factores influyen y en qué medida en la cantidad vocálica. Se ha extraído una serie de variables de otros trabajos que han analizado el fenómeno de la duración en español, si bien se han ignorado otras como su posición en la frase o el registro de habla, que requerirían estudios específicos dada su complejidad.

### *1.1. Factores que influyen en la duración vocálica*

En esta primera aproximación al tema, hemos seleccionado como factores que pueden influir en la duración de las vocales aquéllos que aparecen de forma más recurrente en la bibliografía tanto para el español como para otras lenguas. Dejando de lado la duración intrínseca de cada vocal, las variables que hemos tenido en cuenta en nuestro análisis son las siguientes:

- Posición final/no final en la palabra
- Condición acentuada/inacentuada
- Tipo silábico abierto/cerrado

- Consonante siguiente<sup>1</sup>
- Número de segmentos en la sílaba
- Estructura silábica
- Posición con respecto a la sílaba acentuada
- Número de segmentos en la palabra
- Número de sílabas en la palabra
- Duración de la sílaba
- Duración de la palabra

## 2. Protocolo experimental

### 2.1. Corpus

Hemos seleccionado un cuerpo principal de 340 palabras con distinto número de sílaba y patrón acentual, donde la vocal acentuada fue siempre [a] u [o], con el objeto de englobar el máximo número de contextos posibles y poder evaluar así las incidencias de distintos factores en la duración vocálica. Las palabras objeto de investigación fueron insertadas en la frase sostén “Diga la palabra \_\_\_\_\_ de nuevo”, creándose contextos homogéneos para todas ellas, y han sido presentadas en tarjetas al azar a un informante para su lectura a una velocidad normal. Asimismo, hemos tratado de asignar un indicador de velocidad a todas las frases para corregir las posibles desviaciones que se produjesen en las duraciones por mayor o menor velocidad de producción. Para ello, se ha medido la duración de la única parte fija en las frases portadoras (“Diga la palabra”, de nuevo”), a la que se ha restado la duración de la palabra y pausas ocasionales antes y después de la palabra objeto. Seguidamente se ha calculado la duración media de la parte fija (1066 ms.) y se ha relativizado todas éstas en función de la media. Este valor, que hemos denominado coeficiente de velocidad, ha servido para dividir la duración de consonantes, vocales, sílabas y palabras, y así relativizar los datos. Los resultados obtenidos han sido confrontados con los de otras vocales con menos ocurrencias, lo que hace un total de 42 palabras<sup>2</sup> y 1404 sílabas.

<sup>1</sup> Apoyándonos en Sonnenschein (1925:124): “... any sounds which may precede the vowel-sound of the syllable are negligible, so far as an effect upon ear is concerned; the effective part of a syllable from the psychological point of view is its vowel-sound and any consonant-sound or sounds that may follow this vowel-sound within the syllable. The explanation probably is that the initial consonant-sounds of syllables (if any), though they occupy time in utterance and form part of the syllable, are thrown into obscurity by the superior sonority of the vowel-sounds that follow them and so produce little or no effect upon the ear - i.e., upon the mind; whereas there is nothing within the syllable that can weaken the effect of a consonant-sound that stands at the end of it.”, en esta muestra hemos ignorado la influencia de la consonante en ataque silábico.

<sup>2</sup> La [a] fue la vocal acentuada en 202 de esas palabras, la [o] en 138, la [e] en 26, la [i] en 21, la [u] en 19 y distintos díptongos en 18 de ellas.

## 2.2. Informante

El informante es un hablante nativo castellano, culto, representante de la variedad estándar de la lengua, es decir, sin rasgos dialectales marcados.

## 2.3. Grabación

La muestra sonora fue grabada en su totalidad en el estudio de grabación de la Facultad de Ciencias de la Información de la Universidad de Sevilla mediante una mesa de mezclas Total Audio Concepts (TAC) Scorpion II, sin ecualización, sobre una cinta DAT (Digital Audio Tape) a 48 Khz. en una platina DAT Sony DTC -57 ES, y micrófonos Sony F-730 situados a unos 24 - 30 cms. de la boca del informante.

## 2.4. Análisis acústico

El análisis acústico del corpus se realizó con un Kay DSP- Sonagraph (modelo 5500) en el Laboratorio de Fonética de la Universidad de Sevilla, que permitió la comprobación auditiva y en la onda oscilográfica de la segmentación y medición de las duraciones realizadas sobre sonogramas de banda ancha.

## 3. Resultados

### 3.1. Duración intrínseca

El número total de ocurrencias vocálicas ha sido de 1404 en esta muestra, desigualmente repartidas entre las vocales puras y diptongos, lo que está directamente relacionado con el criterio de selección del corpus:

vocal	Nº casos	%
[a]	586	41.73
[e]	189	13.46
[i]	167	11.89
[o]	351	25.00
[u]	54	3.84
diptongos	57	4.05

Una vez obtenidos los datos sobre duración vocálica, se realizaron diversos análisis de regresión mediante el programa informático de análisis estadístico "CSS Statistics", junto con distribuciones de frecuencia para cada vocal.

Del análisis de casos se obtuvieron las siguientes duraciones medias y desviaciones típicas en milisegundos para cada una de las vocales puras:

Vocal	Dur. media	Desv. típica	Nº casos
[a]	73.65	21.55	586
[e]	65.58	21.21	189
[i]	58.71	20.14	167
[o]	70.94	20.34	351
[u]	55.66	15.47	54

Es incuestionable la mayor duración intrínseca de la [a], de acuerdo con los resultados de Navarro Tomás (1916) o Monroy (1980):

Estudio	Dur. [a]	Dur. [e]	Dur. [i]	Dur. [o]	Dur. [u]
N. Tomás (1916)	126.7	122.4	117.2	123.2	119.5
Monroy (1980)	72.6	71.2	69.0	64.3	53.1
Acentuada	76.0	67.4	75.9	69.1	66.7
Total	73.6	65.5	58.7	70.9	55.6

Como puede apreciarse, nuestros resultados no coinciden con el orden de perceptibilidad establecido por Monroy (1980); no obstante, sí concuerdan con la regla formulada por Meyer<sup>3</sup> y generalizada por autores como Navarro Tomás (1916) y Marí (1994-95): en las vocales del español, a una mayor obertura corresponde una mayor duración, constatándose la existencia de tres grupos diferenciados con respecto a la cantidad: [a] > [o], [e] > [i], [u].

Con respecto a los diptongos, éstos presentaron valores medios de duración mayores que los de las vocales puras pero, como han señalado Santos et al. (1988:80): “menores que la suma de sus dos componentes si estuvieran en hiato o si se tomaran aislados en el mismo contorno.”

<sup>3</sup> Los resultados de Navarro Tomás (1916) y Monroy (1980) incluyen tan sólo ocurrencias acentuadas de las vocales de ahí que aportemos los valores medios de duración tanto para el conjunto total de vocales como sólo para las acentuadas.

<sup>4</sup> Navarro Tomás (1916:403): la vocal es “tanto más breve cuanto mayor es la elevación de la lengua requerida para su articulación.”

### 3.2. Efectos en la duración vocálica de los distintos factores

Con el fin de averiguar qué factores pueden incidir en la duración vocálica, se han aplicado correlaciones entre la duración vocálica y otras variables, obteniendo los siguientes resultados:

1. Posición final/no final. 0.4748
2. Acentuación. 0.1154
3. Tipo abierto/cerrado. 0.0043
4. Número de sílabas en la palabra. - 0.2141
5. Número de segmentos en la palabra. - 0.2493
6. Número de segmentos en la sílaba. - 0.0928
7. Posición con respecto a la síl. acentuada. 0.3001
8. Duración de la sílaba. 0.3127
9. Duración de la palabra. - 0.1186

La correlación más fuerte resultó ser la de posición final/no final en la palabra, seguida lógicamente de la duración de la sílaba, así como de las correlaciones negativas esclarecidas entre la duración de la vocal y el número de segmentos o sílabas en la palabra, indicios de una probable compresión vocálica.

Una vez comprobada la normalidad hemos aplicado el análisis tipo ANOVA para examinar qué factores pueden explicar las varianzas y, por tanto, determinar cómo intervienen en la duración de la vocal. En primer lugar, aplicamos un análisis factor a factor y los resultados obtenidos se exponen a continuación:

En el estudio de las incidencias de la condición final/no final de la sílaba donde aparecía la vocal, los resultados son:

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1*	1	134149	1402	423.362	316.871	0.00000

Por tanto, este factor ha resultado ser altamente significativo a la hora de explicar la variabilidad en la duración del conjunto total de vocales, dando lugar a duraciones medias para cada grupo muy dispares:

Final/no final	Dur. media vocal	Desviación típica	Número de casos
Final	85.52	18.90	422
No final	64.20	21.25	982
Total	70.61	22.77	1404

Las sílabas no finales lógicamente son más numerosas que las finales y la diferencia entre las respectivas medias son muy acusadas, con desviaciones típicas inferiores a la del conjunto general, lo que significa que pueda hablarse de agrupamientos especiales para cada uno de los grupos. Las vocales en sílaba final de palabra son notablemente más largas que las vocales en posición no final, acorde con la regla formulada por Santos et al. (1988:78) en su afán por alcanzar una mayor inteligibilidad y naturalidad en el habla sintetizada: "Si una vocal no se encuentra en sílaba final de palabra se acorta."

El segundo factor analizado es el de la acentuación, cuyos resultados son:

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1*	1	7819.64	1402	513.469	15.2290	0.00010

Este factor, considerado de forma aislada, ha resultado también significativo a la hora de explicar la duración del conjunto total de vocales a un nivel de confianza del 99%, como se puede apreciar comparando las duraciones medias y desviaciones típicas para los dos grupos:

Acentuada/inacentuada	Dur. media vocal	Desviación típica	Número de casos
Acentuada	74.20	20.62	424
Inacentuada	69.05	23.48	980
Total	70.61	22.77	1404

Se puede observar cómo las vocales inacentuadas son mucho más numerosas que las acentuadas y las duraciones medias no tan dispares como para el primer factor, si bien ligeramente superiores en las vocales acentuadas; las desviaciones típicas son similares a la

del conjunto total de vocales. En general, en el sistema lingüístico español las vocales son más largas cuando acentuadas, lo que se muestra de acuerdo con la regla formulada por Santos et al. (1988:78): "... se acorta la vocal si está en una sílaba no acentuada." y los resultados de Marín (1994-95:219), en cuyo estudio el valor medio de duración de la vocal acentuada fue de 70.17 ms. y el de la vocal inacentuada 58.31 ms.<sup>5</sup>

El tercer factor analizado es el de tipo silábico abierto/cerrado, y los resultados de la aplicación del análisis ANOVA son los siguientes:

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	131.064	1404	518.953	0.25255	0.61536

Contrariamente a lo sucedido con los factores anteriores y al igual que en Marín (1994-95:220)<sup>6</sup>, este factor no ha resultado significativo para el conjunto total de vocales, como puede apreciarse en las medias, casi idénticas, si bien las desviaciones típicas muestran valores más dispares:

Abierto/Cerrado	Dur. media vocal	Desviación típica	Número de casos
Abierto	70.82	23.91	944
Cerrado	70.17	20.25	460
Total	70.61	22.77	1404

Las vocales [a, i, u] resultaron más largas en sílaba cerrada, frente a las vocales [e, o], que presentaron duraciones medias superiores en sílaba abierta.

Por otra parte, tanto para el conjunto total de vocales como para cada una de ellas, fue también relevante la correlación negativa entre la duración de la vocal y el número de sílabas<sup>7</sup> y segmentos en la palabra, es decir, cuanto mayor es el número de sílabas o segmentos en la palabra, mayor es la reducción en la duración de la vocal.

Con el fin de comprender mejor la influencia de estos tres factores de forma conjunta, comparando sus pesos relativos en la varianza de las variables y efectos interrelacionados, se ha recurrido al análisis multifactorial tipo MANOVA. Los resultados obtenidos son los siguientes:

<sup>5</sup> Los resultados de Marín (1994-95) corresponden a un corpus de lectura formado por dos textos extraídos del editorial de un periódico.

<sup>6</sup> Las vocales que se encuentran en sílaba cerrada (64.36 ms.) presentan prácticamente la misma duración que las que aparecen en sílaba abierta (64.17 ms.).

<sup>7</sup> Santos et al. (1988:77) lo llaman "acortamiento polisilábico".

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1*	1	39448.4	1396	408.226	96.6338	0.000000
2*	1	12709.9	1396	408.226	31.1344	0.000000
3*	1	11449.7	1396	408.226	28.0474	0.000000
12*	1	1971.6	1396	408.226	4.8298	0.028135
13*	1	6583.2	1396	408.226	16.1265	0.000062
23*	1	1980.6	1396	408.226	4.8518	0.027780
123	1	5.1	1396	408.226	0.0127	0.910294

Como se puede apreciar, los tres factores aisladamente explican la mayor proporción de la varianza, significativos a un nivel de confianza superior al 99%. No obstante, excepto la suma de ellos tres, las distintas combinaciones posibles resultaron igualmente significativas, sobre todo la de posición final/no final + tipo abierto/cerrado. Es más, el tipo silábico abierto/cerrado, que el análisis unifactorial revelaba como no significativo, al encuadrarlo en este nuevo análisis cobra significación, así como su interrelación con los otros dos factores restantes. En estos tres factores obtenemos grupos distintos con las duraciones medias y desviaciones típicas siguientes:

Final/no final Acentuada/inacentuada Abierta/cerrada	Dur. media vocal	Desviación típica	Nº
Final Acentuada Abierta	106.88	23.11	14
Final Acentuada Cerrada	83.13	16.57	121
Final Inacentuada Abierta	86.32	18.89	271
Final Inacentuada Cerrada	71.28	15.77	16
No final Acentuada Abierta	73.69	21.96	103
No final Acentuada Cerrada	66.20	17.13	186
No final Inacentuada Abierta	61.82	21.54	556
No final Inacentuada Cerrada	63.98	22.49	137
Todos los grupos	70.61	22.77	1404

En general, los grupos finales son los que presentan duraciones medias más elevadas, destacando sobre todo el grupo de vocales “final acentuada abierta”<sup>8</sup>; por el contrario el valor medio de duración mínimo resultó ser para el grupo de vocales “no final inacentuada

<sup>8</sup> El escaso número de ocurrencias se explica porque, como dicen D’Introno et al. (1995:157): “Las palabras agudas raramente terminan en vocal.”

abierta”<sup>9</sup> para el conjunto total de vocales y para las vocales [a, i, u], frente a la vocal [e] que presentó su valor mínimo en el grupo de vocales “no final acentuada cerrada” y la vocal [o] que presentó su valor mínimo en el grupo de vocales “no final inacentuada cerrada”. Por tanto, todos los valores mínimos se dieron en posición no final y teniendo en cuenta que los archifonemas representan más de la mitad de las consonantes en coda silábica en esta muestra (73.26%), quizás éstos sean los responsables de las diferencias en las vocales [e] y [o]. Sin embargo, los datos no son suficientemente esclarecedores, con la excepción de la influencia del elemento reductor nasal.<sup>10</sup>

Debido al excesivo número de categorías generadas por los otros factores estudiados, no consideramos conveniente aplicar un análisis tipo ANOVA. Sin embargo, es especialmente significativa la correlación negativa entre el número de sílabas en la palabra y la duración de la vocal, es decir, cuanto mayor es el número de sílabas en la palabra, mayor es la reducción en la duración de la vocal:

Nº sílabas en la palabra	Dur. media vocal	Desviación típica	Nº de casos
1	79.60	23.72	8
2	78.26	22.09	203
3	72.88	23.46	470
4	70.15	21.53	377
5	64.49	20.41	200
6	62.78	21.92	102
7	60.40	26.45	35
8	----	----	---
9	55.11	17.99	9
Todos grupos	70.61	22.77	1404

También significativo, y de igual peso, ha sido la correlación entre la duración de la vocal y el número de segmentos en la palabra:

<sup>9</sup> Categoría más numerosa, con 556 casos, lo que ha representado el 39.71% del total.

<sup>10</sup> Igualmente en Monroy (1980:34): “... en español, a diferencia de lo observado en inglés, no parece darse un orden prioritario en las consonantes postvocálicas en función de su efecto sobre la vocal ... De dar alguna prioridad a los segmentos consonánticos, el máximo reductor en general parece ser la n, según vimos en los primeros resultados.”

Nº segmentos en la palabra	Dur. media vocal	Desviación típica	Nº de casos
2	97.33	48.71	2
3	97.82	37.13	14
4	81.47	24.95	42
5	78.56	19.36	152
6	75.92	23.47	169
7	63.96	23.31	261
8	71.47	20.11	201
9	69.12	21.06	217
10	65.52	21.30	96
11	61.99	19.36	85
12	69.65	32.53	16
13	61.45	21.36	57
14	66.10	24.68	43
15	56.06	18.86	26
16	48.38	18.23	7
17	63.04	15.74	7
18	55.11	17.99	9
Todos grupos	70.61	22.77	1404

Poco o nada significativa resultó la correlación entre la duración de la vocal y el número de segmentos en la sílaba. Por el contrario, el efecto de la posición de la vocal con respecto a la sílaba acentuada en la duración de la vocal ha resultado ser el tercero en importancia:

Posición acento	Dur. media	Desv. estándar	Nº casos
- 6	64.17	5.01	3
- 5	56.10	17.76	11
- 4	63.59	20.43	26
- 3	62.00	23.59	70
- 2	65.67	24.61	174
- 1	63.05	20.90	321
0	74.14	20.66	424
1	77.04	23.31	287
2	81.37	20.79	82
3	75.67	19.70	5
4	65.75	00.00	1

Las vocales preacentuales, más numerosas<sup>11</sup>, presentaron una duración media de 63.58 ms., con un desviación típica de 22.22 ms. y un rango de 143 ms., con un valor mínimo de 18 ms. y un valor máximo de 161 ms. Las vocales postacentuales fueron más largas, con una duración media de 77 ms. y una desviación típica escasamente superior a la de las vocales preacentuales (22.74 ms.). Su rango o recorrido fue de 157 ms., con un valor mínimo de 10 ms. y un valor máximo de 167 ms. Estos resultados coinciden con los de Simões (1991:191) y Macarrón et al. (1991:619):

Stressed vowels were not found to be the longest within the word, but poststressed ones, the shortest being prestressed vowels ... This may give some new insight on the nature of stress in Spanish, since it is generally assumed that stressed vowels have longer duration than unstressed..., an idea that is contradicted by our results.

Por último, se han examinado las características de la vocal según la estructura silábica: el conjunto total de vocales respondió en un 55.27% de los casos a la estructura silábica CV y en un 25.28% a la estructura CVC. Sin embargo, no se han apreciado diferencias por el número de segmentos en la sílaba, siendo tan sólo significativo el hecho de que la vocal contase con una mayor duración en estructuras silábicas cerradas, reduciéndose además las desviaciones típicas dentro de los grupos más frecuentes con respecto a la del conjunto general:

<sup>11</sup> 605 ocurrencias de vocales preacentuales frente a 375 de vocales postacentuales.

Estructura silábica	Duración media vocal	Desviación típica	Número de casos
V	79.31	28.81	123
CV	69.69	22.91	776
VC	73.19	24.83	83
CCV	66.66	21.56	46
CVC	69.96	19.20	355
VCC	88.28	00.00	1
CCVC	62.41	14.87	18
CVCC	53.17	12.75	2

Podemos afirmar, por tanto, que las diferencias no son estadísticamente significativas.

#### 4. Conclusiones

A partir de los resultados obtenidos en este estudio, hemos extraído las conclusiones siguientes:

Podemos diferenciar tres grupos vocálicos ordenados de mayor a menor duración: [a] > [o]/[e] > [i]/[u]. Ello nos permite relacionar un aspecto acústico con otro articulatorio: a una mayor abertura corresponde una mayor duración de la vocal. A la luz de estos resultados debemos concluir que las vocales se diferencian acústicamente por su duración.

La posición final/no final en la palabra resulta ser un factor que influye claramente en la duración vocálica: las vocales en posición final presentan una mayor duración que las vocales en posición no final. De hecho, el valor máximo de duración media para el conjunto total de vocales corresponde al grupo de vocales “final acentuada abierta”, mientras que el valor mínimo corresponde al grupo de vocales “no final inacentuada abierta”.

Asimismo el acento ejerce una notable influencia en la duración de la vocal: las vocales acentuadas presentan una mayor duración que las vocales inacentuadas.

Se han constatado efectos de compresión temporal en la duración de la vocal por efecto del número de sílabas y el número de segmentos en la palabra.

Los resultados obtenidos en lo que se refiere a la influencia del número de segmentos en la sílaba y la estructura silábica no indican que éstos sean factores que afecten a la duración de los sonidos vocálicos.

Finalmente, no ha sido tan clara la correlación entre la duración de la vocal y la posición con respecto a la sílaba acentuada, si bien ha resultado obvia la mayor duración de las vocales postacentuales frente a las preacentuales.

## 5. Bibliografía

- F. D'Introno, E. del Teso, R. Weston: *Fonética y fonología actual del español* (Madrid, 1995)
- A. Macarrón, G. Escalada y M.A. Rodríguez: "Generation of Duration Rules for a Spanish Text-to-Speech Synthesizer", EURO\_SPEECH (Génova, 1991) 617-620
- R. Marín: "La duración vocálica en español", *Estudios de Lingüística* nº 10 (1994-95) 213-226
- R. Monroy: *Aspectos fonéticos de las vocales españolas* (Madrid, 1980)
- T. Navarro Tomás: "Cantidad de las vocales acentuadas", *RFE* III (1916) 387-407
- A. Santos, P. Muñoz y M. Martínez: "Diseño y evaluación de reglas de duración en la conversión de texto a voz", *SEPLN* nº 6 (1988) 71-92
- A. E. Sonnenschein: *What is Rhythm?* (Oxford, 1925)