

**PRODUCCION DEL RASGO ACENTUAL
MEDIANTE SINTESIS DE VOZ**

CAROLINA FIGUERAS
Universidad de Barcelona

MARISA SANTIAGO
Universidad de Barcelona

RESUMEN

En este artículo se describe un método de síntesis del rasgo acentual en un sintetizador de voz PCF 8200 de Philips y se demuestra que los valores aportados por el análisis de la voz no siempre se corresponden con los que es necesario introducir en un sintetizador para que los hablantes reconozcan ese rasgo.

ABSTRACT

In this paper we describe a method of synthesis for stress in a Philips's voice synthesizer PCF 8200 and we prove that the values provided by the analysis of the voice do not always correspond to the ones that have to be introduced in a synthesizer, so that speakers identify this feature.

1. INTRODUCCION

Para intentar reproducir los datos que habíamos obtenido en el análisis de palabras compuestas realizado a través del VISI-PITCH¹ trabajamos con un programa de síntesis de voz² creando archivos de sílabas átonas y tónicas que se adaptaran lo mejor posible a los resultados de nuestra primera investigación.

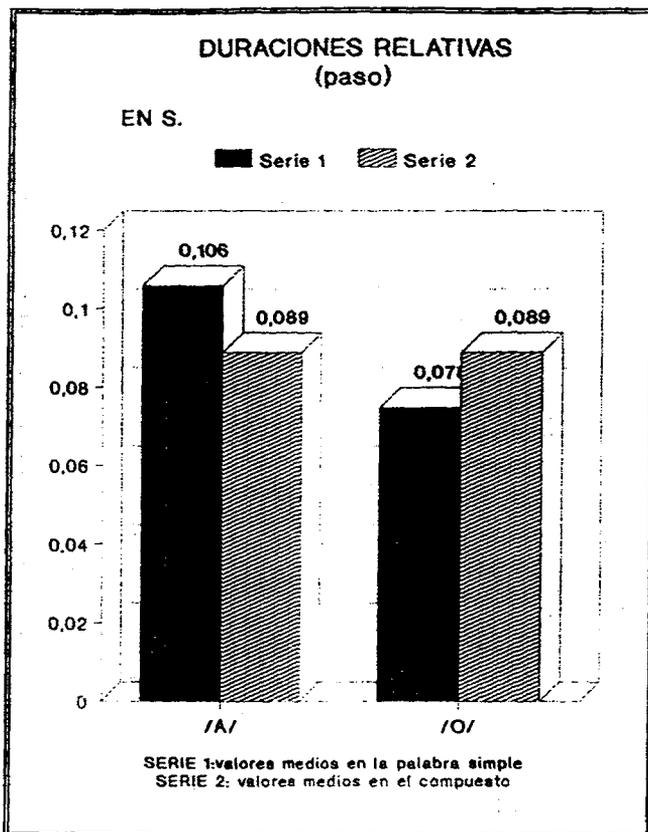
2. PRIMER EXPERIMENTO

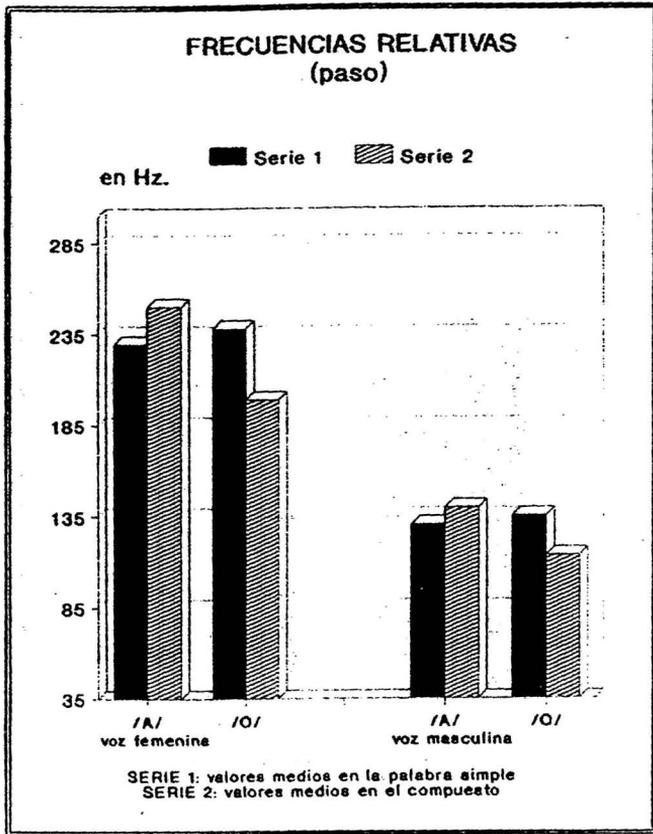
En principio elegimos la palabra **paso** porque el término **paso cebra** fue pronunciado por casi todos los informantes como compuesto, es decir, con pérdida de acento en **paso**. En la palabra simple (**y**, por tanto, llana) los informantes masculinos mantenían casi el mismo tono fundamental en las dos vocales, **o**, como máximo había una diferencia de 12 Hz a favor de la [o]. La palabra compuesta presentaba una átona [o] con valores mucho más bajos (mientras que apenas variaba la [a]). La consecuencia de esta variación era una inversión en la relación entre los valores absolutos de ambos tonos fundamentales. Interpretamos como pérdida acentual este cambio.

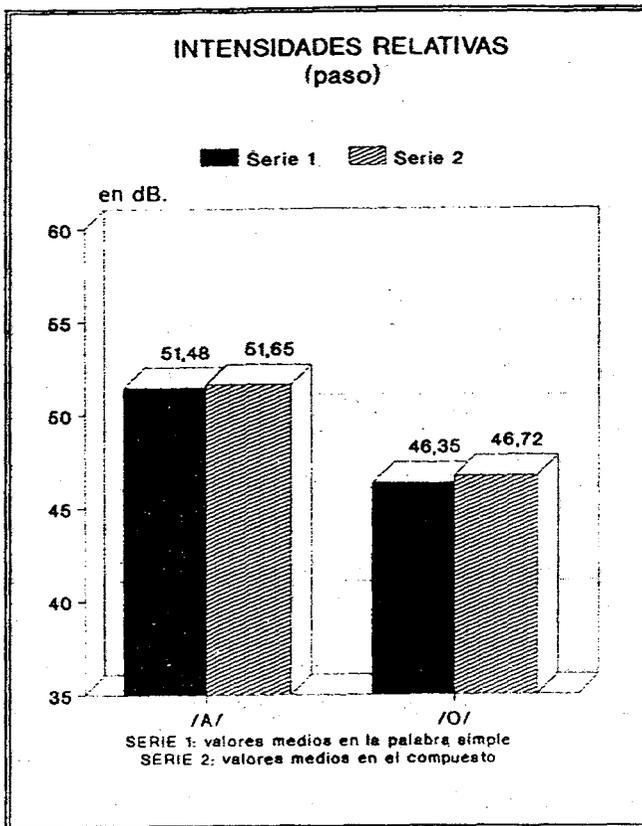
Los tres gráficos que presentamos a continuación muestran los resultados obtenidos en esas mediciones:

¹ Vease nuestro artículo "Investigación sobre la naturaleza del acento a través del VISI-PITCH" en este mismo volumen.

² Hemos utilizado un sintetizador PCF 8200 y el programa editor de unidades fonéticas VOXSY, elaborado por Tecnivoz (Barcelona). Se trata de un sintetizador por formantes, que utiliza cinco formantes con sus correspondientes anchos de banda y permite modificar el tono fundamental, la intensidad y la duración.







Tomando como punto de partida el hecho de que la átona también puede variar, intentamos modificaciones de signo diferente en ambas sílabas para comprobar cuáles se interpretaban como acento. Partimos de un archivo paso neutro, y elevamos el tono fundamental de la [a] 45,2 Hz, 13,4 Hz y 6,1 Hz. Del mismo modo, elevamos el de la [o] con los mismos valores. Probamos también disminuciones del F0 de la [a] con los mismos valores absolutos.

En todos estos archivos, compensábamos la subida o bajada al principio de la vocal con un movimiento equivalente de signo contrario al final de la misma, para mantener el tono fundamental de

Observamos una tendencia a considerar llanas la mayoría de las realizaciones. Quizá tenga que ver con que ésta es la estructura fonética más frecuente en castellano. Descartamos que este hecho dependa de desigualdades en los archivos de las dos sílabas puesto que la vocal [o] de nuestro archivo es dos tramas (unos 0,016s) más larga que [a].

Sólo parece haber unanimidad en considerar agudos los archivos en los que la vocal [o] es sustancialmente más alta que [a] (45,2 Hz), tanto si esta diferencia se debe a una subida de los valores de [o] como si es fruto de la bajada del tono fundamental de [a].

Pasamos, en consecuencia, a intentar reforzar la sensación de acento en la última sílaba, evidentemente mucho más difícil de percibir que el de las llanas. Utilizamos para ello un aumento en la duración de la tónica (manteniendo la modificación anterior sobre el F0). Además, pese a reconocer acento, varios de los informantes afirmaban oír una entonación no neutra (preguntas, exclamaciones...) en las supuestas agudas. Convenía, pues, utilizar combinaciones de duración y frecuencia, para, así, rebajar los valores de estos dos parámetros de modo que la variación fuera la mínima necesaria para interpretar acento.

3. SEGUNDO EXPERIMENTO

Probamos con una nueva palabra de las mismas características fonéticas: *pi-se*. El cambio nos servía para comprobar si todas las vocales se comportan de un modo similar o no. De momento, no hemos observado diferencias importantes a causa del timbre.

La elevación de 45,2 Hz de la [e] se percibe más como entonación que como acento. El resultado de aumentar tres tramas a la [e], además, mejora en algo el efecto. Seguimos haciendo pruebas al respecto y el resultado más natural fue el que llamamos archivo *pi-sef*.

Pisef es un archivo en el que la [i] de un originario [pi] neutro fue reducida en 24 ms. Su amplitud alcanza un máximo de 33dB y su F0 hace una bajada respecto al fundamental de 6,1 Hz. En contraste con esto, la [e] final tiene 56 ms. más que [i], una amplitud máxima de 39 dB, 6 dB más que la átona, y su tono fundamental hace una subida sobre el tono original de 8,5 Hz. Una propuesta que nos pareció que daba muy buen resultado para evitar la sensación de entonación ascendente (que no se elimina usando simplemente valores de compensación) es dejar un par de tramas (cada trama corresponde a 8 ms.) con el mismo tono fundamental y después realizar una nueva bajada ligera en la última trama, unos 6,1 Hz, que da al conjunto un matiz conclusivo. Este archivo fue el resultado más natural de palabra aguda que conseguimos.

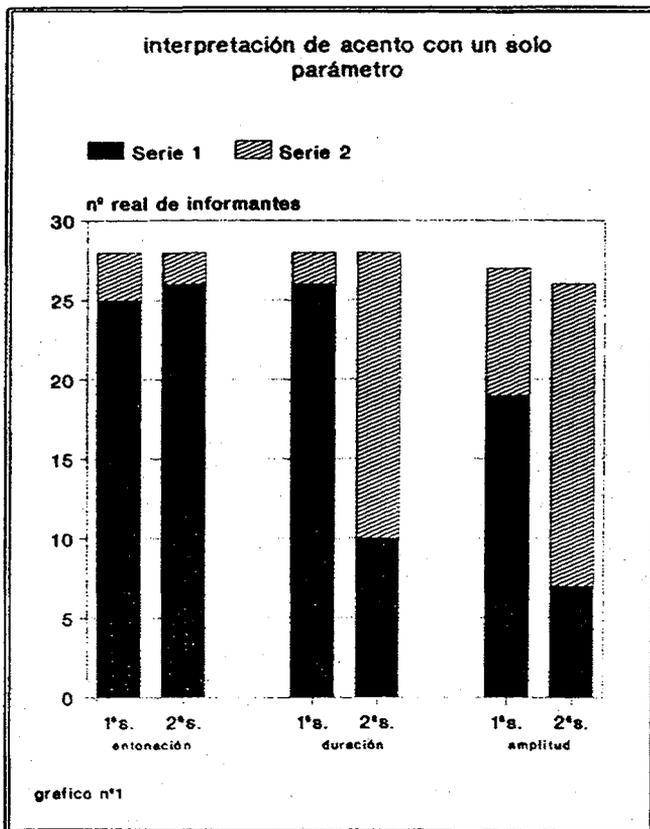
Con estos datos empezamos a construir un test de audición. Para eliminar posibles distorsiones utilizamos dos sílabas iguales [pa]. Llamamos [pa] tónica a la que contenía las variaciones de frecuencia, tiempo y amplitud que hemos descrito en el [se] de pisef, y [pa] átona a la sílaba análoga al [pi] de ese archivo.

Elegimos esta sílaba porque, concatenada consigo misma, da como resultado dos palabras existentes en español (/pápa/ y /papá/) de la misma categoría morfológica e, incluso, del mismo significado, con lo que pretendíamos atenuar aspectos subjetivos en la identificación de una u otra³.

Realizamos varias modificaciones (que iremos detallando a continuación) sobre estos dos archivos y presentamos las diversas concatenaciones de dos archivos [pa] a veintiocho informantes.

³ Enríquez, Casado y Santos en "la percepción..." señalan repetidamente la tendencia a escuchar un nombre antes que una forma verbal, y la influencia que estas cuestiones tienen a la hora de decidir dónde está el acento en una palabra.

Dado que nos interesaba verificar mediante el test el grado de intervención de cada uno de los tres parámetros en la percepción del acento, preparamos también sílabas, a partir de la [pa] átona, en los que uno de ellos (sólo el tiempo, o sólo el F0 o sólo la amplitud) se correspondiera con los valores propios de la tónica, y los combinamos con otra [pa] átona.



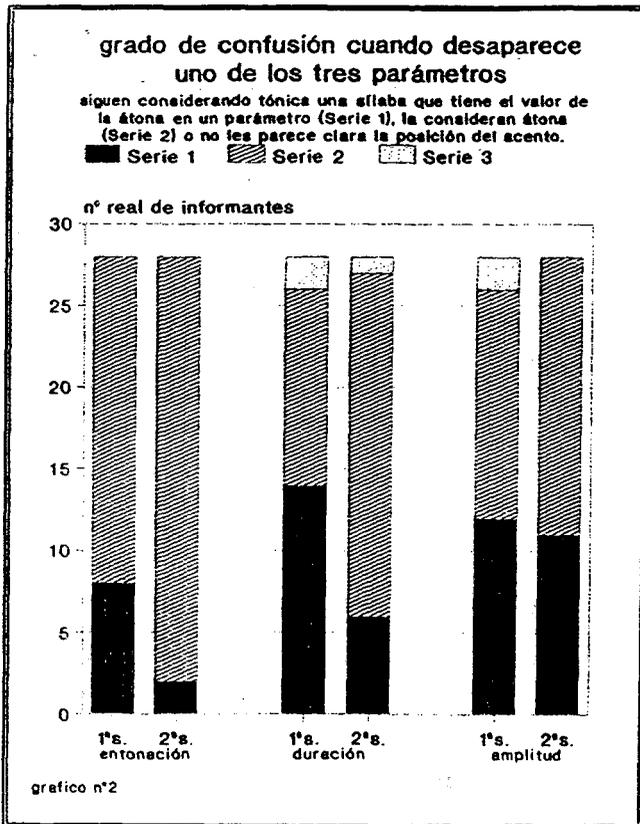
El gráfico nº1 representa los resultados de estas combinaciones. Llama la atención que las palabras con variaciones de amplitud y duración en la sílaba final no son percibidas como agudas por

la mayoría de los informantes. En cambio, una palabra con una modificación en el F0 sí se interpreta como aguda, por 26 de los 28 informantes. En los otros casos, se comprueba la tendencia de que hablábamos a percibir llanas, tanto si es el tiempo como si es el F0 el parámetro modificado. La amplitud no produce resultados tan evidentes, pero, de todos modos, 19 de los 28 informantes siguen reconociéndola como llana.

Se puede observar también que modificar la duración de la segunda sílaba apenas influye en la percepción del acento, al contrario de lo que sucede si se modifica en la primera sílaba.

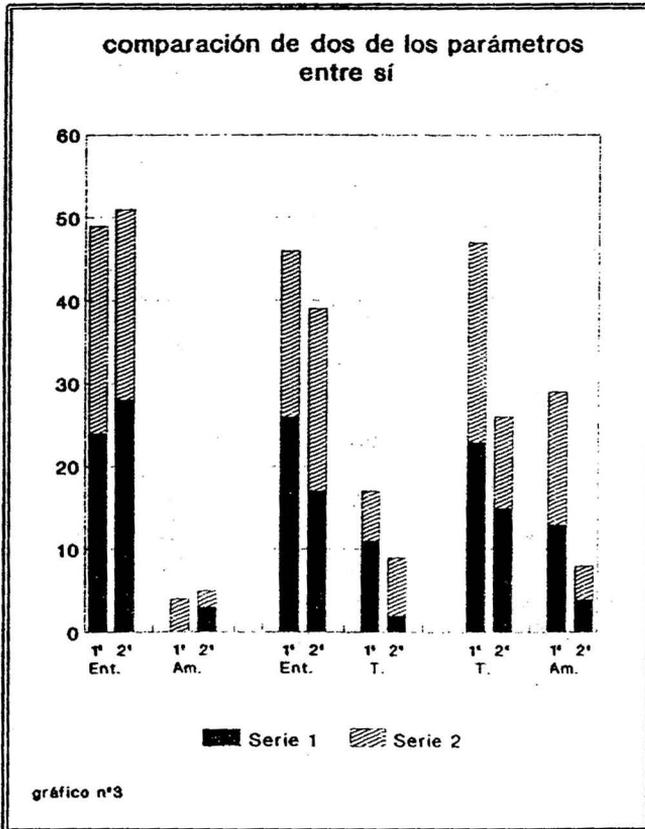
Realizamos también la prueba complementaria, es decir, comprobar de cuál de los tres parámetros podíamos prescindir alterando en menor medida la percepción de acento. Para ello construimos sílabas tónicas con uno de los tres valores propios de las átonas. El gráfico nº2 muestra los resultados de combinar estos archivos con tónicas plenas. Si no hubiera habido modificación de ningún tipo no habría ninguna razón para esperar un número mayor de respuestas a favor de una aguda o a favor de una llana. Eliminar uno de los valores disminuye la probabilidad de considerar esa sílaba como tónica. Los resultados indican que la entonación es prioritaria frente a los otros dos factores, ya que, donde ésta falta, la gran mayoría de informantes no reconoce una sílaba tónica.

La amplitud apenas influye en la percepción de acento. Además, no hay diferencias entre variarla en la primera o en la segunda sílaba. El tiempo, sin ser tan relevante como la frecuencia, sí resulta muy significativo cuando es menor en la segunda sílaba. Esto refuerza la idea de que la duración es útil para crear una palabra llana, pero no consigue por sí misma producir una aguda.



El gráfico nº3 muestra los datos obtenidos al presentar a los informantes dos tipos de combinaciones: por un lado, se concatenan dos sílabas átonas, cada una de las cuales tiene un parámetro propio de la tónica, para comprobar cuál de los dos predomina (Serie 1) ; por otro, y con el mismo fin, se presentan unidos dos archivos tónicos con un parámetro de vocal átona, diferente en cada uno de ellos (Serie 2). Con los casos del primer tipo, el parámetro no implicado en la comparación (es decir, aquel cuyo valor es idéntico en ambas sílabas) presenta valores propios de la sílaba átona, mientras que con el segundo, ese tercer parámetro reforzaría el

acento, puesto que presenta valores propios de sílaba tónica.

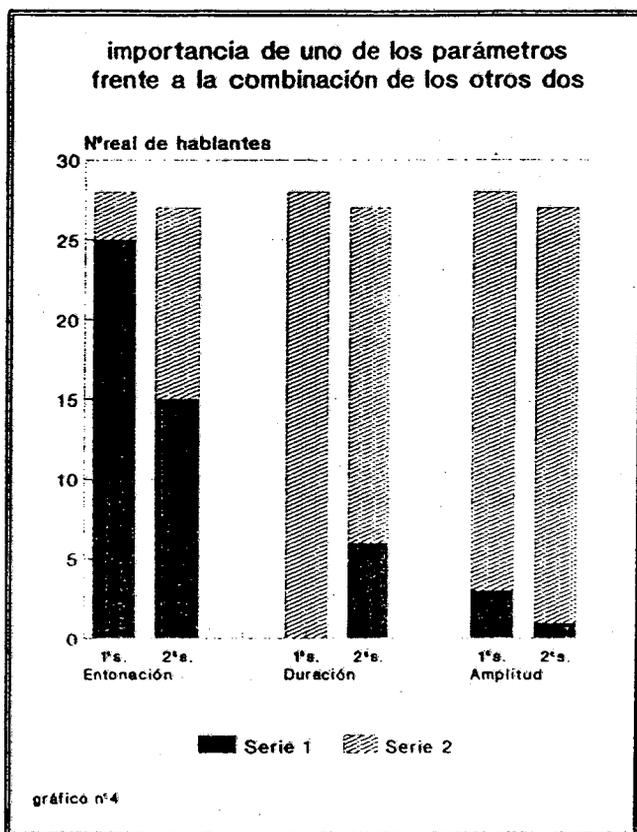


Estos datos nos permiten establecer una gradación clara de los parámetros en función de su preponderancia relativa en la percepción de acento: la entonación es el factor fundamental, seguido por la duración, tanto en posición inicial como final.

En el gráfico n°4 se ha intentado reflejar la importancia de uno de los parámetros, frente a la combinación de los otros dos; es decir, se compara

una sílaba átona a la que se ha aumentado la entonación, frente a otra átona a la que se han aumentado duración y amplitud; lo mismo se ha hecho con los otros dos parámetros.

La entonación predomina ampliamente sobre duración e intensidad, tanto si eso obliga a que la palabra se interprete como aguda o como llana.



4. CONCLUSIONES

Los resultados obtenidos en el test corroboran la idea de que es la frecuencia fundamental el parámetro más significativo en la percepción del acento. La posición de éste no parece influir en la importancia de este valor que, por sí solo, produce más sensación de acento que los otros dos en conjunto (tiempo e intensidad).

La duración refuerza la sensación de acento en una palabra llana, pero resulta insuficiente para crear una aguda. La amplitud no crea por sí misma sensación de acento.

En todo caso, cabe establecer una gradación en la intervención de los tres parámetros: en primer lugar actúa la frecuencia fundamental, seguida de la duración. La intensidad, por último, no parece influir prácticamente nada en el acento, aunque, combinada con los otros dos valores, permite ofrecer mayor naturalidad a la sílaba tónica creada sintéticamente.

Cabe concluir, por tanto, que el acento en español es la combinación de estos tres parámetros, los cuales contribuyen en variable medida a su percepción.

5. BIBLIOGRAFÍA

ENRIQUEZ, E-V., CASADO, C., SANTOS, A. 1989. "La percepción del acento en español", *LEA XI*.

MARTINEZ CELDRAN, E., ROMERA BARRIOS, L. y SALCIOLI GUIDI, V. 1986. "Reconocimiento de voz y síntesis de habla", *Actas del II Congreso de lenguajes naturales y lenguajes formales*.

MARTINEZ CELDRAN, E. 1986. *Fonética*, Teide, Barcelona.