

PROPUESTA DE UN MODELO DE INTENSIDAD VOCÁLICA DEL CASTELLANO Y EL CATALÁN APLICABLE A UN SISTEMA DE CONVERSIÓN DE TEXTO A HABLA

Beatriz Blecua

Universitat de Girona / Universitat Autònoma de Barcelona¹

Vanessa Acín

EUETT La Salle, Universitat Ramon Llull

RESUMEN

La intensidad de los sonidos del castellano y del catalán es un tema poco tratado en la bibliografía. Resulta interesante llenar este vacío no sólo desde un punto de vista descriptivo, sino también para una mejora en los sistemas de síntesis de habla. En este trabajo se presentan dos modelos de intensidad vocálica, uno para el castellano y otro para el catalán, aplicables a sistemas de conversión de texto a habla. Por otra parte, se propone un sistema de análisis de la intensidad de los sonidos.

Palabras clave: intensidad, vocales, síntesis, catalán, castellano.

Área temática: reconocimiento y síntesis de habla

0. INTRODUCCIÓN

La intensidad es uno de los componentes menos tratados en los estudios de fonética acústica, probablemente a causa de los problemas que plantea a la hora de determinarla. En primer lugar, generalmente interesa la diferencia de intensidad entre los sonidos, de forma que los valores obtenidos no son absolutos, sino relativos, y por lo tanto es imprescindible determinar la referencia que se utiliza. Por otra parte, es necesario adoptar una técnica de medición y determinar con exactitud qué se va a medir. Existen distintas formas de calcular la intensidad del sonido: se puede considerar la intensidad en un punto, como puede ser el punto máximo (*peak value*), o tomar una serie de valores en un segmento y calcular el nivel de intensidad media. En este caso es muy importante la acotación del segmento analizado, ya que su duración es relevante. Otra posibilidad para el caso

¹ Dirección Universitat de Girona: Facultat de Lletres, Departament de Filologia i Filosofia
p/ Ferrater i Mora 1, 17071 Girona. Telf: (972)418106
e-mail: blecua@skywalker.udg.es

específico de sonidos vocálicos es tener en cuenta la intensidad de cada uno de los formantes².

La grabación del corpus comporta un problema adicional para el estudio de la intensidad. Parámetros como la frecuencia de los formantes o la duración de un sonido no dependen de aspectos externos al propio proceso de producción del habla una vez controladas las variables, mientras que la intensidad está directamente relacionada con la distancia del informante al micrófono, el volumen de grabación o el énfasis que le imprima el locutor, de forma que cualquier variación en estos aspectos durante la grabación, o incluso entre las distintas sesiones, lleva a resultados erróneos. Resulta imprescindible, por tanto, controlar todos estos parámetros.

El objetivo del presente trabajo es realizar un estudio de la intensidad de las vocales del castellano y del catalán aplicable a un sistema de síntesis de habla. Será imprescindible, en primer lugar, determinar qué variables influyen en las variaciones de intensidad, para poderlas controlar al elaborar el corpus. Una vez analizados los datos, veremos en qué medida influyen estas variables, y de qué forma se combinan unas con otras. Se intentará comprobar también la existencia de una intensidad intrínseca para cada vocal, tal como han afirmado algunos autores³.

Por otra parte, se propone un sistema para medir la intensidad de los sonidos en el que queden controlados los problemas relacionados con la grabación que se han señalado más arriba.

1. PROCEDIMIENTO EXPERIMENTAL

1.1 Elaboración del corpus

1.1.1. Variables

Un factor que parece tener incidencia en la intensidad de las vocales es el acento. A pesar de que el parámetro que estamos tratando influye en que una sílaba se perciba como acentuada o inacentuada, existen estudios que muestran que, al menos para el español, no es tan importante como la duración o la frecuencia fundamental, y que una variación en la intensidad no es suficiente para que se perciba un cambio claro de acentuación⁴. De todos modos, vamos a considerarlo como una de las variables de nuestro corpus. Albalá- Marrero (1993) tienen en cuenta este factor y obtienen una diferencia global de 3,67 dB entre vocales tónicas y átonas del español, produciéndose las mayores diferencias en la [a].

² Borzone de Manrique (1980), pp.95-100.

³ Borzone de Manrique (1980), pp.99. En relación con este aspecto, cita a Black (1949), Fairbanks et al. (1950) y Scia y Beck (1926).

⁴ Enríquez et al. (1989).

Otro aspecto que vamos a tener en cuenta es si la sílaba a la que pertenece la vocal analizada está en posición prepausal o no. Se ha visto que este es un factor que influye de manera notable en parámetros como la duración⁵: en posición prepausal, la vocal es más larga que si no se encuentra ante pausa. Creemos que puede ser una variable decisiva también para la intensidad.

La posición en el enunciado podría suponer también cambios de intensidad: no será igual una vocal que se encuentre en la primera sílaba del enunciado que la que forme parte de las sílabas centrales o finales. Esto está relacionado con la curva melódica, así como con la respiración. El estudio de Albalá-Marrero (1993) concluye, sin embargo, que las vocales en posición inicial absoluta presentan menor intensidad que en interior o final de grupo fónico. Sería interesante comprobar estos datos.

Otra variable que se ha decidido considerar es la longitud del enunciado: la intensidad de los elementos puede variar dependiendo de la duración de la frase en la que se encuentran. Parece que el volumen fónico, es decir, el número de sílabas, incide en variaciones de intensidad ya que la intensidad y la duración de los segmentos disminuye a medida que aumenta el número de sílabas.

Por último, se va a tratar también la estructura silábica, es decir, si la vocal se encuentra en sílaba libre o trabada.

Así, el corpus se ha elaborado a partir de la combinación de cinco variables⁶:

1. vocal en sílaba acentuada o inacentuada (+ ac/ - ac)
2. vocal en posición prepausal o no prepausal (+ prep/ - prep)
3. vocal en sílaba trabada o libre (+ T/ - T)
4. vocal en un enunciado largo o corto (enunciado largo / enunciado corto)
5. vocal en posición inicial, media o final del enunciado (I/M/F)

1.1.2. Corpus

El corpus está formado por frases aisladas, ya que de esta forma es más sencillo conseguir un número representativo de casos de cada vocal combinada con todas las variables. Para diseñarlo se han combinado las cinco variables hasta obtener tres casos de cada combinación para cada una de las lenguas⁷.

⁵ Borzone y Signorini (1983), Marín (1994).

⁶ En el corpus para el estudio del catalán no se ha tenido en cuenta la variable longitud del enunciado. Sin embargo, se ha controlado que todos fueran enunciados cortos.

⁷ El sistema vocálico del castellano consta de 5 vocales [i,e,a,o,u], que pueden aparecer tanto en posición tónica como átona. El catalán tiene un sistema vocálico formado por 8 vocales. 4 de ellas [e,ɛ,a,ɔ] sólo se encuentran en posición tónica, [ə] únicamente en posición átona, y [i,u] en ambas posiciones.

La primera división se ha realizado según la longitud del enunciado, de manera que la mitad son enunciados cortos y la otra mitad enunciados largos. Esta distinción se ha establecido a partir del número de sílabas: se han considerado enunciado corto las frases de menos de 8 sílabas, y enunciado largo las de más de 8 sílabas. En cada uno de los enunciados se ha distinguido entre posición prepausal y no prepausal. Se han tomado como "prepausal" los casos en posición final de enunciado, ya que de esta forma nos aseguramos de que se realizará pausa en el momento de la lectura.

Una vez separados los casos de prepausal y no prepausal se han combinado, para cada una de estas opciones, las variables +/-acento y +/- trabada, y posteriormente, para cada una de las combinaciones resultantes no prepausales, se ha tenido en cuenta la posición en el enunciado: inicial, media y final para los enunciados largos y únicamente inicial y final para los cortos, ya que resultaba muy difícil distinguir tres posiciones en frases de menos de 8 sílabas. Esta variable sólo se ha podido considerar, como es lógico, para sílabas no prepausales, ya que hemos visto que todas las prepausales se encuentran en posición final. Se ha considerado posición inicial la primera sílaba del enunciado. Como posición final se ha tomado la penúltima sílaba, ya que la última siempre es prepausal, y para posición media se han tenido en cuenta sílabas situadas en el centro del enunciado.

Para evitar la posible influencia de la entonación, todas las frases tienen un patrón melódico enunciativo.

El total de realizaciones vocálicas analizadas es de 960 para el castellano y 250 para el catalán⁸

1.2. Informantes

Se han analizado las vocales del corpus producido por cuatro informantes, dos para cada lengua. Todos son varones de edad comprendida entre 20 y 30 años y formación universitaria. Los locutores han sido grabados en su lengua materna: dos de ellos ha leído el corpus catalán y otros dos el castellano.

1.3. Grabaciones

Las grabaciones se han llevado a cabo en la cámara semianecoica de la EUETT La Salle (Universitat Ramon Llull), con un micrófono de condensador y un

⁸ El menor número de casos analizados para el catalán se debe a que, por una parte, el corpus consta únicamente de enunciados cortos, lo que implica que no sólo se reduce a la mitad el número de frases, sino que además se elimina la posición media. Por otra parte, así como en el corpus castellano se han analizado 3 casos por contexto, en el catalán sólo se han tenido en cuenta 2 casos.

DAT. La grabación ha consistido en la lectura del corpus descrito en el apartado anterior. Durante la misma se ha controlado la realización de pausas, es decir, que no se produjeran en medio de un enunciado pero sí entre las distintas frases.

1.4. Análisis acústico

Para evitar los problemas relacionados con la distancia al micrófono, los movimientos del locutor y las variaciones interlocutor, se ha utilizado un sistema especialmente diseñado para este estudio por el CITVAM del Departamento de Acústica de la EUETT La Salle, de la Universitat Ramon Llull, que toma como valor de referencia la intensidad media de la voz del hablante. Este sistema parte de la idea de que para hallar la intensidad media de la voz de un hablante es necesario antes determinar el tiempo mínimo necesario para que se establezca la energía. Para ello se realiza la media aritmética del valor de energía de una serie de muestras, añadiendo sucesivamente las muestras, de forma que hallamos una curva de intensidad que se estabilizará a partir de un tiempo determinado, es decir, que llegará un punto en el eje temporal en que no se producirán oscilaciones de más del 10% con respecto a la media. En la figura 1, T_p representa este tiempo mínimo y E_0 corresponde al valor de energía en el punto medio:

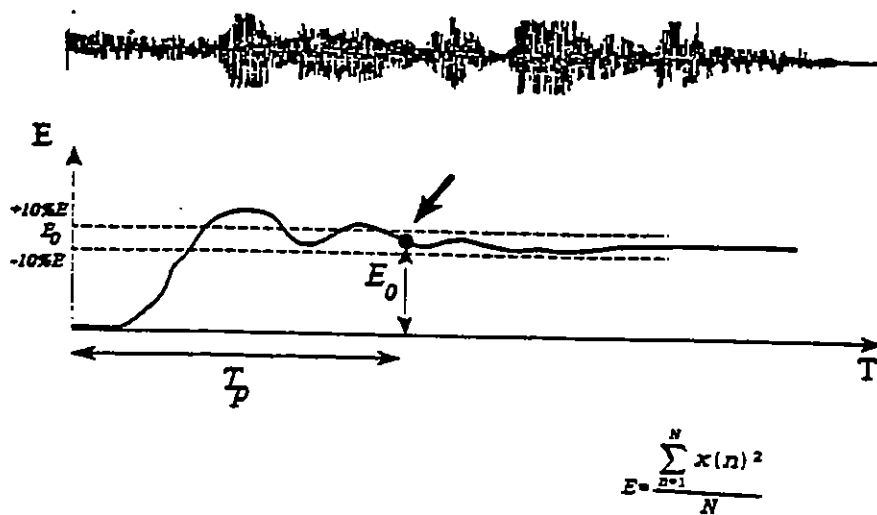


figura 1

figura 1: esquema del procedimiento de determinación de la intensidad media de la voz de un hablante

Por medio de este procedimiento se ha determinado que T_p , el tiempo mínimo en que se estabiliza la curva de intensidad, es de 2 segundos. Esto supone que a partir de una grabación de al menos 2 sg de duración podremos hallar la intensidad

media de la voz de un hablante. Una vez establecido este valor, se determina la energía de la trama que nos interesa, en este caso un segmento vocálico, en relación a dicho valor medio, de forma que las medidas que se obtienen son relativas. Estos valores se expresan en decibelios.

Una vez obtenidas las grabaciones, se han filtrado y digitalizado a una frecuencia de muestreo de 20 KHz a 16 bits mediante el programa SoundScope 16, instalado en un ordenador Macintosh II VX.

El punto de medición de la intensidad ha sido el valor máximo de la curva de energía para cada vocal.

2. TRATAMIENTO ESTADÍSTICO

Los datos han sido tratados estadísticamente con el programa Statview. Se han realizado ANOVA y T-test, que han proporcionado los grados de significación de las diferencias entre variables, las medias de intensidad en dB para cada caso y la desviación estándar.

3. RESULTADOS

Para determinar qué variables influyen en la intensidad vocálica se han realizado t-Test o ANOVA de los distintos valores de cada una de dichas variables, pruebas que muestran la probabilidad de que los conjuntos de datos comparados pertenezcan a un mismo grupo. En los casos en que $p > 0,05$, no se dan diferencias significativas entre los grupos comparados, de forma que se han considerado como un mismo grupo.

Como paso previo, se ha realizado para cada una de las lenguas un t-test para determinar si los locutores se comportan de forma uniforme. Los resultados muestran, tanto en el caso del castellano como del catalán, que no se dan diferencias significativas entre los valores que presentan los dos informantes, de manera que para los dos estudios los datos de los informantes se han analizado de forma conjunta.

3.1. Castellano

En primer lugar se van a presentar los resultados correspondientes a cada variable por separado, y posteriormente se comprobará la influencia de unas variables en otras.

Los resultados de los t-test aplicados a las variables longitud del enunciado y tipo de sílaba muestran que no se dan diferencias significativas entre las vocales que se encuentran en enunciado corto y las que aparecen en enunciado largo, así como tampoco entre las que forman parte de una sílaba libre y las que se hallan en sílaba trabada. No parece, por lo tanto, que estas variables influyan en la intensidad de las vocales del español, al menos de forma aislada.

Entre las vocales en posición prepausal y no prepausal sí se dan, en cambio, diferencias significativas ($p = 0,0001$). En posición prepausal la media de los valores de intensidad normalizada es de 5 dB, es decir, 5 dB por encima del valor de intensidad media del hablante, bastante más baja que en posición no prepausal, cuyo valor medio es de 11,17 dB.

Por otra parte, también la variable acento presenta diferencias significativas al aplicar el t-test. Los valores medios son de 10,61 dB para las acentuadas frente a 8,16 dB de las inacentuadas.

De igual forma, la posición en el enunciado parece influir en la intensidad vocálica. En este caso, el test estadístico utilizado es el ANOVA. Los resultados muestran que se pueden distinguir dos grupos: las vocales en posición inicial y media, por una parte, y las vocales en posición final por otra. Los valores medios para las vocales del primer grupo son 12,8 dB y 12,22 dB, respectivamente, mucho más elevados que los que se encuentran para la posición final, 8,98 dB⁹.

Las fig. 2, 3 y 4 muestran los valores medios de intensidad de cada una de las variables que se han dado diferencias significativas en los tests estadísticos.

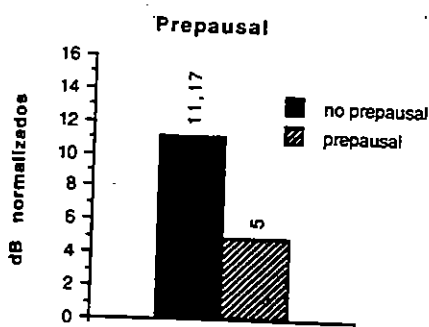


Fig. 2

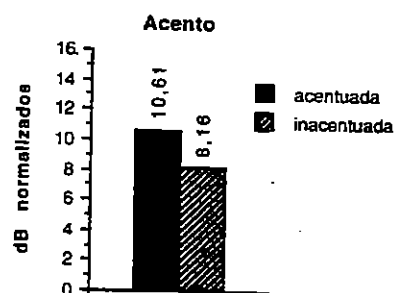


Fig. 3

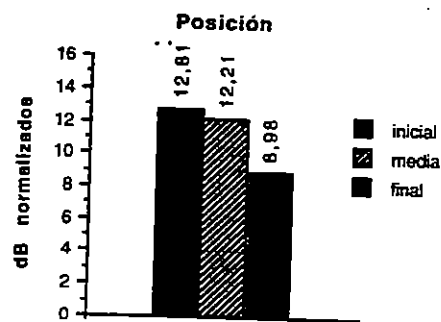


Fig. 4

Fig.2, 3 y 4: valores medios de intensidad relativa según las variables posición prepausal (fig.2), acento (fig.3) y posición en el enunciado (fig.4).

⁹ Para el valor medio en posición final se han excluido los casos prepausales (que son todos final), ya que, al ser "prepausal/no prepausal" una variable que influye en la intensidad vocálica, podría interferir en los resultados. Se trata, por tanto, de una comparación de las tres posiciones para las vocales no prepausales.

Por último, se ha realizado un ANOVA de las cinco vocales con el fin de determinar si se pueden tratar como un mismo grupo, cada una de ellas por separado o como varios grupos. A partir de los resultados, se podría considerar una escala [a] > [i] > [o] > [e] > [u], de mayor a menor intensidad, que se mantiene en prácticamente todas las condiciones. Esto sería una prueba a favor de la existencia de una intensidad intrínseca para cada vocal. Sin embargo, las diferencias entre vocales no son muy elevadas (la diferencia entre el valor medio de [a] y el de [u] es de 2 dB).

Hasta aquí se ha estudiado cómo influye cada variable por separado. Pero es posible que algunas variables en las que no se obtienen diferencias significativas estudiadas por separado, sí que las presenten combinadas con otras variables.

Se ha observado que la variable longitud del enunciado no influye en la intensidad vocálica si se toman todos los valores en conjunto. Sin embargo, al realizar las pruebas estadísticas teniendo en cuenta la posición en el enunciado, los resultados difieren. En posición inicial, la longitud del enunciado no es relevante. En posición final, en cambio, se dan diferencias significativas entre vocales que forman parte de enunciados largos y cortos.

Por otra parte, se ha querido comprobar cómo afecta la combinación de variables que sí daban resultados significativos. Al combinar las variables acento y posición prepausal, se observa que, tanto en sílaba inacentuada como acentuada, las vocales en posición prepausal son más intensas que las que se encuentran en posición no prepausal:

posición prepausal inacentuada:	2,24 dB
posición prepausal acentuada:	7,52 dB
posición no prepausal inacentuada:	10,2 dB
posición no prepausal acentuada:	11,17 dB

Por último, se han comparado las vocales acentuadas e inacentuadas en cada una de las tres posiciones (inicial, media y final). Los resultados muestran que las vocales en sílaba acentuada son más intensas que las que se hallan en sílaba inacentuada, independientemente de la posición en que se encuentren.

En la tabla 1 se presentan los valores medios y el número de casos observados para cada vocal, así como el total, teniendo en cuenta únicamente aquellas variables para las que se obtienen diferencias significativas.

				[a]	[e]	[i]	[o]	[u]	TOTAL			
				n m (dB)	n m (dB)	n m (dB)	n m (dB)	n m (dB)	n m (dB)			
PREPAUSAL	- ACENTO	n m (dB)	ENUNCIADO LARGO	10 0,88	12 1,67	12 2,14	12 0,76	9 1,94	55 1,47			
			ENUNCIADO CORTO	11 4,77	11 3,65	10 3,01	10 2,3	11 1,41	53 3,04			
	+ ACENTO	n m (dB)	ENUNCIADO LARGO	11 6,02	11 7,48	12 6,62	14 8,19	13 6,05	61 6,83			
			ENUNCIADO CORTO	11 7,79	12 8,46	11 8,31	12 9	11 7,69	57 8,27			
NO PREPAUSAL	- ACENTO	n m (dB)	FINAL	enunciado largo	11 10,39	12 8,05	11 8,38	11 8,35	12 6,02	57 8,2		
				enunciado corto	11 10,05	11 9,23	12 8,56	11 11,23	12 12	57 9,55		
			MEDIA	11 13,07	11 12,1	12 12,12	11 11,83	12 9,89	57 11,78			
			INICIAL	21 12,57	22 10,33	23 11,38	24 11,43	22 10,9	112 11,31			
			+ ACENTO	n m (dB)	FINAL	enunciado largo	12 9,18	11 8,88	12 7,86	10 7,46	11 8,28	56 8,36
						enunciado corto	11 11,28	11 8,33	11 11,44	10 9,38	9 8,82	52 9,9
	+ ACENTO	n m (dB)	MEDIA	11 14,5	11 12,71	12 13,02	13 12,5	12 10,64	59 12,64			
			INICIAL	22 15,71	25 14,56	23 14,55	19 14,44	21 12,19	110 14,32			

Tabla 1: intensidad media normalizada de cada una de las vocales del castellano en función de las variables prepausal, acento, posición y longitud del enunciado.

3.2. Catalán

Mediante los mismos métodos estadísticos utilizados para el castellano, se ha determinado qué variables son las que más influyen en la intensidad de las vocales catalanas.

La variable que más afecta al parámetro estudiado es el hecho de que la vocal sea prepausal o no. La media de los valores relativos en posición prepausal es de 8,36 dB, mientras que en posición no prepausal es de 11,76 dB.

La influencia del acento también es destacable: 9,6 dB en vocales inacentuadas frente a 11,6 dB en vocales acentuadas.

Por otra parte, se ha estudiado el efecto que produce la posición inicial o final en la intensidad de la vocal. Debido a que el corpus del catalán consta únicamente de enunciados cortos, no se incluye la posición media. Los resultados del t-test muestran que esta variable afecta en todos los contextos, aunque las diferencias entre estas dos posiciones son más elevadas en las vocales acentuadas que en las inacentuadas.

Las fig. 5, 6 y 7 muestran los valores medios de intensidad de las tres variables que suponen mayores diferencias:

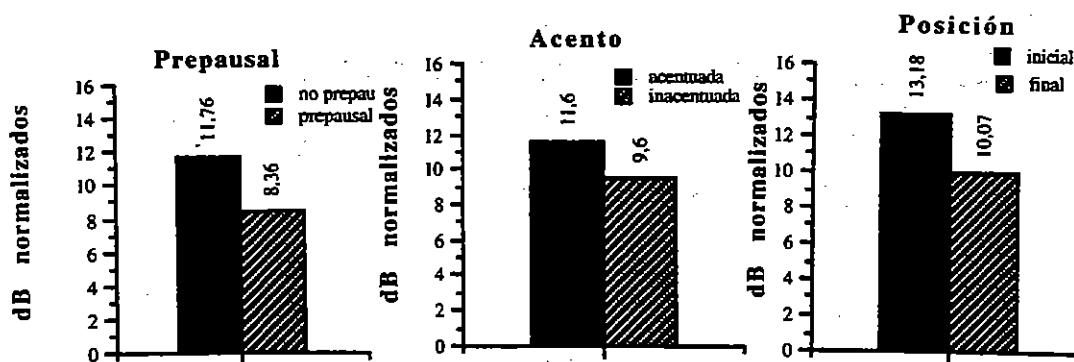


Fig. 5

Fig. 6

Fig. 7

Fig.5,6 y 7: valores medios de intensidad relativa según las variables posición prepausal (fig.5), acento (fig.6) y posición en el enunciado (fig.7).

La estructura silábica tiene una influencia mínima en la intensidad vocálica en la mayoría de los casos. Sin embargo, en determinados contextos se han hallado diferencias significativas entre vocales en sílaba trabada y libre.

En cuanto a las distintas vocales del catalán, dado que su distribución no es la misma (no todas ellas pueden aparecer tanto en sílaba tónica como en átona), parece conveniente diferenciar tres grupos:

grupo 1: [a, e, ε, o, ɔ]

grupo 2: [i, u]

grupo 3: [ə]

El grupo 1 corresponde a las vocales que únicamente aparecen en sílaba acentuada. En el grupo 2 se hallan las vocales que se pueden dar tanto en sílaba acentuada como inacentuada, y el grupo 3 incluye la única vocal que sólo se encuentra en sílaba inacentuada.

Los resultados muestran diferencias significativas en el comportamiento los tres grupos respecto a las distintas variables, pero no en el interior de cada uno de ellos, por lo tanto se ha realizado su estudio por separado.

Al combinar las variables acentuada/no acentuada y prepausal/no prepausal, se observa que las vocales prepausales son menos intensas que las no prepausales, tanto si son inacentuadas como si son acentuadas. Por otra parte, en posición inicial se da más intensidad que en posición final, independientemente de la variable acento.

En la tabla 2 aparecen los valores medios de intensidad para cada grupo de vocales en los contextos en que se han hallado diferencias significativas:

				GRUPO 1		GRUPO 2		GRUPO 3		TOTAL	
				n	m (dB)	n	m (dB)	n	m (dB)	n	m (dB)
PREPAUSAL	- ACENTO	- TRABADA		-		10	4	6	9	16	6,5
		+ TRABADA		-		6	6,62	4	7,6	10	7,11
	+ ACENTO	- TRABADA		20	9,73	8	10		-	28	9,86
		+ TRABADA		58	10,02	10	9,25		-	30	10,8
NO PREPAUSAL	- ACENTO	- TRABADA	FINAL			6	11	6	12	12	11,5
			INICIAL			8	11,5	8	8	16	9,75
		+ TRABADA	FINAL			14	11,5	6	11	20	11,25
			INICIAL			8	11,5	2	13	10	12,25
	+ ACENTO	- TRABADA	FINAL	18	10,88	4	11,65		-	22	11,26
			INICIAL	20	13,7	8	12,24		-	28	12,97
		+ TRABADA	FINAL	14	12	6	9,17		-	20	10,58
			INICIAL	26	15,12	12	14,58		-	38	14,85
PREPAUSAL	- ACENTO	- TRABADA									
		+ TRABADA									
	+ ACENTO	- TRABADA									
		+ TRABADA									
NO PREPAUSAL	- ACENTO	- TRABADA	FINAL								
			INICIAL								
		+ TRABADA	FINAL								
			INICIAL								
	+ ACENTO	- TRABADA	FINAL								
			INICIAL								
		+ TRABADA	FINAL								
			INICIAL								

Tabla 2: intensidad media normalizada de cada grupo de vocales del catalán en función de las variables prepausal, acento, posición y estructura silábica.

4. COMPARACIÓN Y DISCUSIÓN DE LOS RESULTADOS

Hasta este punto se ha realizado una descripción de la intensidad de los sonidos vocálicos del castellano y del catalán teniendo en cuenta una serie de variables. En términos generales, los resultados indican que las mayores diferencias se dan entre vocales prepausales y no prepausales, con una intensidad menor en las primeras. En este aspecto, sin embargo, el catalán presenta diferencias menores entre los dos grupos de vocales (3,4 dB de diferencia) que el castellano (6,17 dB).

La variable acento es otra de las más importantes, y combinada con la primera permite hacer una primera clasificación de los segmentos que estamos estudiando. De menor a mayor intensidad, encontramos la siguiente escala:

- vocales prepausales inacentuadas
- vocales prepausales acentuadas
- vocales no prepausales inacentuadas
- vocales no prepausales acentuadas

También hay que tener en cuenta la posición en el enunciado. En posición inicial la intensidad es mayor que posición final en ambas lenguas.

El aspecto en que encontramos más diferencias entre el catalán y el castellano es en el comportamiento de cada una de las categorías vocálicas. La diferencia entre los sistemas vocálicos tónicos y átonos de ambas lenguas dificulta su comparación. Nos hemos limitado a comparar el efecto de las variables sin tener en cuenta el timbre de la vocal, llegando a la conclusión de que éstas afectan de forma similar en las dos lenguas.

Por otra parte, resulta interesante señalar que los contextos en que aparecen los valores máximo y mínimo son los mismos para ambas lenguas. El valor mínimo se halla en vocales prepausales, inacentuadas y en sílaba libre (en enunciado largo, para el castellano). La intensidad media en esta posición es de 0,27 dB para el castellano (en la vocal [u]) y de 4 dB para el catalán (en el grupo 2, que incluye las vocales [i, u]). El valor máximo aparece en posición no prepausal, acentuada, sílaba trabada y posición inicial, 16,03 dB para el castellano (en la vocal [a]) y 15,12 dB para el catalán (en el grupo 1, que corresponde a las vocales [a, e, ε, o, o]). El rango es, por tanto, de 15,76 dB en castellano y 11,12 dB en catalán¹⁰.

Los resultados del castellano pueden compararse también con el estudio de Albalá-Marrero (1993). Ambos trabajos coinciden en las vocales que presentan la mayor y la menor intensidad media, [a] y [u], respectivamente, mientras que [e, i, o] muestran valores intermedios. Por otra parte, la variable acento influye de la misma forma en los dos estudios: las vocales acentuadas son más intensas que las inacentuadas. Un aspecto en el que difieren es en la variable posición en el enunciado: mientras que en nuestro estudio el valor de intensidad disminuye a medida que avanza la posición en el grupo fónico, para Albalá-Marrero (1993) las vocales en posición inicial absoluta presentan menor intensidad que las que se encuentran en posición interior. Sin embargo, hay que tener en cuenta que los dos estudios están realizados a partir de corpus de distinto tipo y la referencia utilizada para medir la intensidad no es la misma, de forma que las diferencias pueden estar relacionadas con estos factores.

5. MODELOS DE INTENSIDAD VOCÁLICA PARA EL CASTELLANO Y EL CATALÁN

Una vez repasadas las variables que más influyen en la intensidad vocálica, se va a intentar diseñar un modelo que muestre cómo se manifiestan estas

¹⁰ A pesar de que el rango en catalán es menor que en castellano, hay que tener en cuenta que el corpus catalán consta únicamente de enunciados cortos, y es precisamente en los enunciados largos donde aparecen los valores mínimos del castellano.

diferencias y que pueda ser aplicable a un sistema de conversión de texto a habla. En el diseño del modelo de intensidad se han seguido los criterios empleados para otros modelos prosódicos, especialmente de duración¹¹. En este sentido, la primera decisión fue la elección de un modelo de tipo aditivo o de tipo multiplicativo. Se ha optado por la primera opción, que se adapta mejor a las necesidades de este trabajo. En un modelo aditivo se parte de un valor y se van sumando cantidades (pueden ser valores concretos o porcentajes) según se den una serie de condiciones. En nuestro caso, hemos optado por restar en lugar de sumar, ya que los valores máximos resultaban más estables que los mínimos. Las unidades con las que se trabaja son dB relativos al nivel medio de intensidad de la voz del hablante. Se ha partido de los valores de intensidad más elevados para cada vocal, que corresponden a las vocales no prepausales, acentuadas y en posición inicial, y se han ido restando decibelios según los valores de cada variable. Las diferencias entre el modelo propuesto para el castellano y el catalán se hallan en el valor de partida, que en el caso del catalán es el mismo para todas las vocales (las diferencias entre los tres grupos de vocales se añaden al final) y en el del castellano depende de la vocal (se parte del valor máximo de cada vocal). Esto es debido a que en castellano las variaciones de intensidad en los distintos contextos afectan de forma similar a todas las vocales. Sin embargo, dado que cada vocal tiene su propia intensidad intrínseca, el valor de partida será distinto para cada una. En catalán, las variables no influyen de la misma forma en todas las vocales, sino que a cada grupo le afectan en distintas proporciones. Se ha considerado más sencillo para su posterior implementación en un sistema de conversión texto a habla partir de un mismo valor y controlar el grupo al que pertenece la vocal al final del proceso.

Las fig. 8 y 9 representan los modelos para cada una de las lenguas:

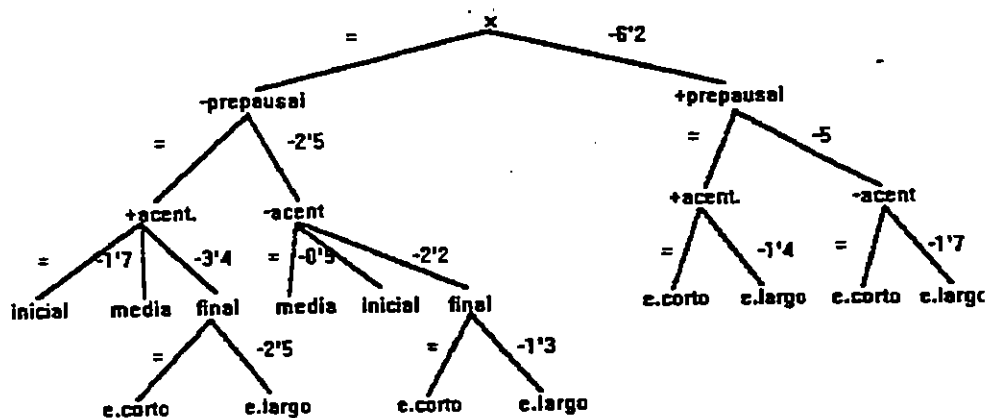


Figura 8: modelo de intensidad vocálica para el castellano

¹¹ Marín (1994)

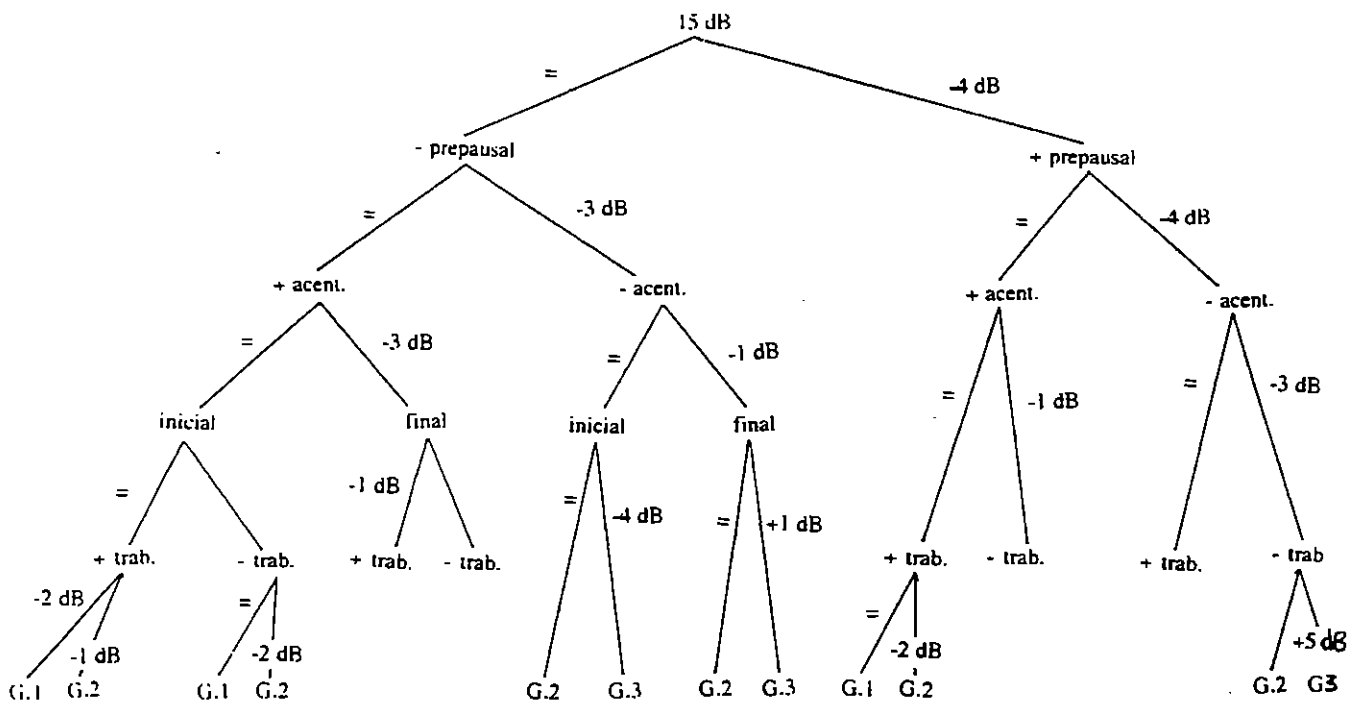


Figura 9: modelo de intensidad vocálica para el catalán

Se parte del nodo superior, y cada vez que hay un nodo se opta por uno de los caminos. La rama izquierda de cada nodo mantiene el valor de intensidad, mientras que las ramas de la derecha producen sustracciones o adiciones de dB. Al final de cada rama del organigrama se obtiene un valor en dB correspondiente a la vocal en el contexto elegido. Este valor es la diferencia de intensidad, expresada en dB, de dicha vocal en relación a la energía media de la voz del hablante.

6. CONCLUSIÓN

En términos generales, la intensidad de las vocales varía en función de los mismos factores en castellano y en catalán. Se ha demostrado que las variables que más influyen en la intensidad vocálica son el acento, la posición en el enunciado (inicial o final) y el hecho de hallarse en posición prepausal o no. Por otra parte, no se comportan de igual modo las distintas vocales en ambas lenguas, pero los sistemas vocálicos tónico y átono son también muy diferentes, factor que podría influir en estas diferencias.

Los modelos propuestos están diseñados para ser aplicados a un sistema de conversión texto-habla. Sería necesario, sin embargo, un estudio de la intensidad de las consonantes de ambas lenguas que complementase la presente propuesta.

REFERENCIAS

- ALBALÁ, M.J. - MARRERO, V. (1993) *La intensidad de los sonidos españoles*, en prensa.
- BORZONE DE MANRIQUE, A.M. (1980) *Manual de fonética acústica*. Buenos Aires, Hachette Universidad.
- BORZONE, A.M. - SIGNORINI, A. (1983) "Segmental duration and rythm in Spanish" *Journal of Phonetics* 11 pp.117-128.
- ENRIQUEZ - CASADO - SANTOS (1989) "La percepción del acento en español", *Lingüística Española Actual* XI pp.241-269.
- FRY, D.B. (1955) "Duration and intensity as physical correlates of linguistic stress" *JASA* 27 pp.765-768.
- LLISTERRI, J. (1984) *Anàlisi i síntesi de les vocals del català*. Memoria de licenciatura. U.A.B.
- MARÍN, R.(1994) "Diseño y evaluación de un modelo de duración vocálica del español para la síntesis de habla", Actas del X Congreso de la SEPLN, Córdoba.