PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994) "Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in *Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*. Córdoba, 20-22 de julio de 1994.

http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

PARAMETRIZACIÓN DE LAS PAUSAS ORTOGRÁFICAS EN CASTELLANO. APLICACIÓN A UN CONVERSOR DE TEXTO A HABLA

PUIGVÍ, D. (1) - JIMÉNEZ, D. (2) - FERNÁNDEZ, J.M. (1)

(1) Departamento de Filología Española, Edificio B,

Universitat Autònoma de Barcelona

08193, Bellaterra, Barcelona.

(2) Departamento de Acústica EUETT La Salle,

Passeig de la Bonanova, 8,

08022, Barcelona.

Resumen: Con el presente estudio se pretende establecer una serie de reglas de duración para las pausas ortográficas en castellano, con el fin de aplicarlas a un conversor de texto a habla. El método se basa en el análisis oscilográfico de las producciones sonoras obtenidas en la lectura de un texto literario, realizada por dos informantes castellano-hablantes, que tienen el castellano como primera lengua y como lengua de uso habitual. Las reglas han sido implementadas en un sistema de conversión de texto a habla. Finalmente, se evaluan las mejoras en la calidad del sistema introducidas por el nuevo algoritmo.

Abstract: The present paper establishes a set of duration rules for ortographic pauses in Spanish to be applied to a text-to-speech system. The method is based on oscillographic analysis of the utterances of two speakers reading a literary text. The rules were implemented in a text-to-speech system. Finally, the improvements in the system's quality introduced by the new algorithm are evaluated.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

INTRODUCCIÓN

El estudio de las pausas, aunque se trata de un elemento ligado al proceso de producción del habla y parece que su análisis debería ser abordado principalmente por la fonética, puede enfocarse desde campos muy diversos (como la psicolingüística y la sociolingüística). Dada esta interdisciplinariedad, autores como O'Connell-Kowal (1980), proponen la creación de una nueva ciencia: la pausología.

Dentro del campo psicolingüístico, se atiende a las causas que originan la aparición de la pausa en el discurso, relacionándola con la organización mental del enunciado. Por ello, se distinguen dos tipos fundamentales de pausa: "empty pauses" -o pausas vacías- (Goldman-Eisler, 1972) y "filled pauses" -o pausas que podrían denominarse llenas- (Boomer, 1965). Se entiende por "empty pauses" aquellas en las que no existe producción sonora alguna, identificables con las también llamadas "pausas fisiológicas", realizadas para respirar. Por otra parte, en las "filled pauses" sí existe producción sonora; éstas equivalen a las "hesitation pauses" o "pausas de dubitación".

Sin embargo, la pausa es ante todo un fenómeno fonético, consistente en un período de silencio, de duración variable, que interrumpe la cadena fónica pero que no tiene función articulatoria, es decir, que no corresponde a la fase de oclusión de las consonantes oclusivas, africadas y vibrantes (Caldognetto, 1992).

En el campo de los estudios sobre el castellano se define la pausa como un momento de silencio, y se relaciona con la intención del hablante, a la vez que se menciona el carácter fisiológico de la pausa.

En ninguno de estos enfoques se establece una tipología de duraciones de la pausa que permita proponer una serie de reglas para su implementación en un sistema de conversión de texto a habla. Así, atendiendo a la diferenciación en la duración de la pausa, sólo se distingue

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

entre pausas ortográficas (aquellas indicadas mediante un signo de puntuación) mayores y menores, encontrándose las primeras a nivel discursivo - de mayor duración - y las segundas en el interior de la frase - de menor duración -.

El estudio que aquí se presenta intentará aplicar a un sistema general de síntesis un conjunto de reglas sobre la duración de las pausas ortográficas en castellano. Dichas reglas tendrán que ser aplicadas automáticamente por un conversor de texto a habla mediante programa, de manera que posteriormente a la introducción del texto por el usuario, el algoritmo de síntesis ordene a la placa de control del sistema realizar la pausa con la duración adecuada. En la figura 1 se muestra el proceso seguido para realizar dicho estudio.

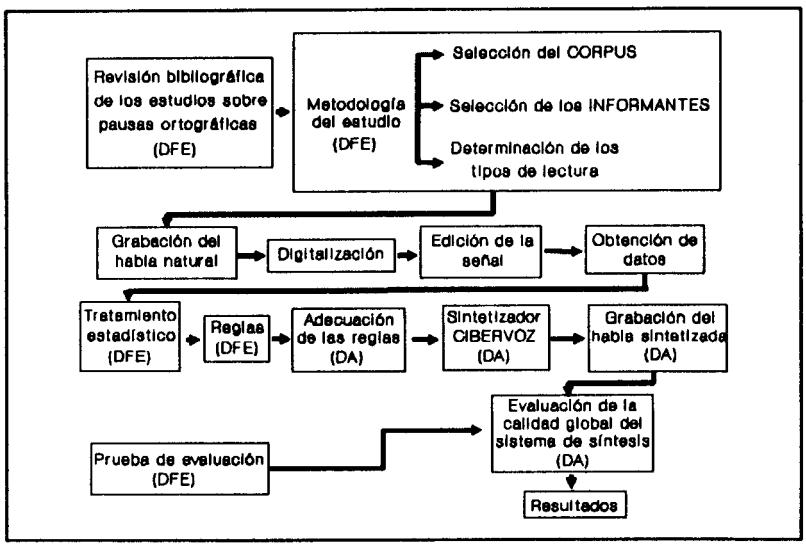

Fig. 1. ***Diagrama de bloques mostrando el proceso seguido en la realización del estudio.***

(DA: Departamento de Acústica. DFE: Departamento de Filología Española)

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994) "Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

MÉTODO

Como corpus fue utilizado un fragmento de un texto literario, perteneciente a una obra de L. Alas "Clarín", titulada **Zurita**, publicada, junto a otras narraciones, en un volumen, bajo el nombre genérico de **Narraciones Breves**. Se trata de un texto que contiene 517 palabras y toda la variedad de signos de puntuación posible. Este texto fue leído por dos informantes de sexo diferente, bilingües catalán-castellano con un elevado grado de dominancia castellana, sin rasgos dialectales marcados, universitarios residentes en Barcelona, y familiarizados con la tarea de leer textos en voz alta en condiciones de grabación. Sus edades estaban comprendidas entre los 20 y 30 años. El reducido número de informantes es debido a que el estudio tiene una finalidad más cualitativa que cuantitativa. Estos informantes leyeron los textos a tres velocidades de elocución diferentes (normal, lenta y rápida), con el fin de estudiar la duración de la pausa en relación a la velocidad de elocución, aunque después los resultados fueron agrupados para simplificar las reglas del conversor. Se pidió a los informantes que leyeran "de una manera normal" y a partir de esta lectura realizaron las otras dos, aumentando o reduciendo su velocidad de elocución, consiguiendo así una lectura rápida y otra lenta, respectivamente.

Las grabaciones se llevaron a cabo en la cámara semi-anecoica del Laboratorio de Fonética de la U.A.B. El material utilizado para realizar dichas grabaciones fue una platina TASCAM 112, TEAC Professional Division, con una respuesta en frecuencia de 25 Hz a 18 KHz +/- 3 dBs; una mesa de mezclas TASCAM M-106, y un micrófono SENNHEISER MKH 20 P 48 U-3.

La duración de la pausa se ha obtenido a partir de un análisis oscilográfico con el programa MacSpeech-Lab II. Los datos obtenidos mediante este procedimiento se han tratado estadísticamente con el objeto de determinar la duración media de cada tipo de pausa en

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

función del signo de puntuación en el texto.

A partir de dicha clasificación se establecieron unas reglas aplicables al módulo preprocesador del conversor.

En la figura 2 se muestra el diagrama de bloques del conversor CIBERVOZ ™, utilizado en el estudio, donde se representan los módulos -entre ellos el preprocesador-, la información dinámica y el conjunto de tablas y reglas necesarias para su función.

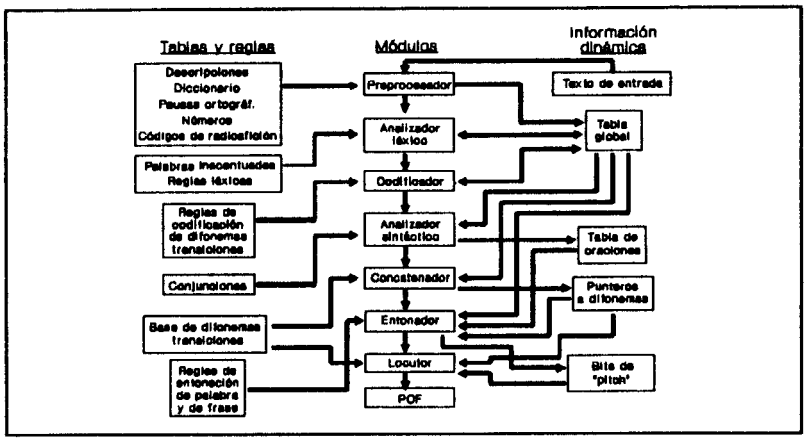

Fig. 2. *Algoritmo de síntesis de habla utilizado en el conversor CIBERVOZ ™*

RESULTADOS

A. PROPUESTA DE REGLAS

Una vez efectuado el análisis estadístico sobre los resultados correspondientes a cada clase de pausas asociadas a signos de puntuación, se diseñó un conjunto de reglas para aplicar sistemáticamente al módulo preprocesador -módulo inicial del algoritmo de síntesis de habla mostrado en la figura 2-.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

En la propuesta de reglas se diferenció entre <segmento largo> -aquel que iguala o supera en número de sílabas al que normativamente se ha considerado como el grupo fónico estándar para el castellano, es decir, de 8 a 10 sílabas- y <segmento corto> -aquel que no supera las 8 sílabas- (Navarro, 1918). A continuación se presentan las reglas de duración obtenidas para las pausas marcadas ortográficamente:

1. Punto y seguido precedido de un segmento largo: 824 ms.
2. Punto y seguido precedido de un segmento corto: 667 ms.
3. Entre dos fragmentos del discurso separados por un punto y aparte y antes de un segmento interrogativo largo: 873 ms.
4. Después de un segmento interrogativo largo: 748 ms.
5. Punto y coma precedido de un segmento largo: 711 ms.
6. Punto y coma precedido de un segmento corto: 591 ms.
7. Puntos suspensivos: 565 ms.
8. Después del segundo signo de exclamación: 468 ms.
9. Coma anterior a un nexo copulativo: 462 ms.
10. En una aposición entre comas, la primera coma: 401 ms; la segunda: 416 ms.
11. En una oración parentética, entre guiones, paréntesis o comas, la primera pausa: 321 ms; la segunda: 558 ms.
12. Coma después de una construcción con participio absoluto: 472 ms.
13. Coma que divide los segmentos en una enumeración con nexos distributivos: 429 ms.
14. Coma situada entre los miembros de una enumeración: 481 ms.
15. Coma que precede a una subordinada relativa: 463 ms.
16. Coma que precede a una subordinada adverbial: 407 ms; la que la sigue: 413 ms.
17. Dos puntos: 638 ms.
18. Antes de un segmento interrogativo corto: 284 ms; después: 561 ms.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

19. Coma que sigue a un SPrep locativo: 549 ms.

20. Coma que sigue a un adverbio acabado en "-mente", situado a comienzo de frase: 448 ms.

21. Coma situada entre dos oraciones yuxtapuestas: 553 ms.

22. Coma anterior a una oración comparativa: 560 ms; la posterior: 515 ms.

La duración de la pausa del primer signo exclamativo no se contempla porque viene determinada por el signo ortográfico anterior.

Se observa que para aplicar las reglas referentes a la coma, los paréntesis y los guiones, con los valores anotados, sería necesario llevar a cabo un análisis sintáctico automático del texto ("parsing") excesivamente complejo, dado el estado actual del módulo de análisis del conversor utilizado. Por otra parte, los valores que aparecen están comprendidos entre 400-500 ms. Este intervalo notablemente pequeño no justifica la necesidad de un analizador sintáctico que adjudique valores numéricos exactos, disminuyendo por otra parte, la velocidad de proceso del texto. Por tanto, se puede tomar un valor medio para la coma, el guión y el paréntesis de 450 ms, que cubre correctamente todos los casos. Esta agrupación permite utilizar en la síntesis los valores que se muestran en la figura 3.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

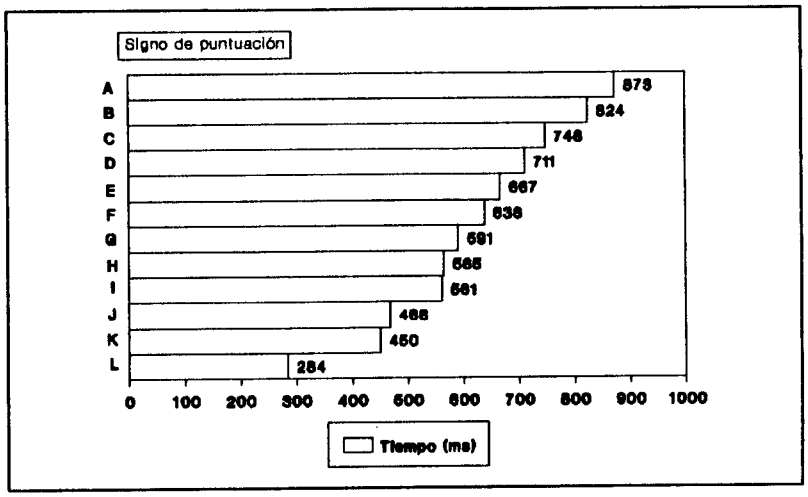

Fig. 3. *Histograma que muestra la duración de la pausa asociada al signo de puntuación*

Los códigos utilizados en el eje de ordenadas se explican a continuación.

A: Punto y aparte / 1ª interrogación en un segmento interrogativo largo
B: Punto y seguido precedido de un segmento largo
C: 2ª interrogación en un segmento interrogativo largo
D: Punto y coma precedido de un segmento largo
E: Punto y seguido precedido de un segmento corto
F: Dos puntos
G: Punto y coma precedido de un segmento corto
H: Puntos suspensivos
I: 2ª interrogación en un segmento interrogativo corto
J: Exclamación
K: Coma / guión / paréntesis
L: 1ª interrogación en un segmento interrogativo corto

B. EVALUACIÓN

Los datos presentados en la figura 3 ratifican la división propuesta tradicionalmente (RAE, 1973), según la cual existe una gradación entre las duraciones del punto, punto y coma y

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

coma, estableciendo además los valores en milisegundos. En este mismo cuadro se presentan valores para otros signos de puntuación no considerados hasta el momento.

A partir de las reglas obtenidas se ha configurado el *software* de control necesario, incluyéndolo dentro del sistema de conversión automática de texto a habla CIBERVOZ™ (ver figura 2).

Para validar la calidad del sistema de síntesis así modificado utilizamos a la prueba propuesta por Fernández (1992). Esta prueba permite evaluar la calidad global del sistema, mediante la respuesta a una serie de preguntas que indirectamente intentan evaluar la adecuación de elementos segmentales y suprasegmentales, así como la impresión global que recibe el usuario del sistema estudiado. La prueba se llevó a cabo con un grupo de 10 personas, castellano-hablantes de nivel cultural alto, que puntuaron (por pares de opciones dadas) aspectos relevantes de tres versiones de habla. Para cada una de estas opciones se dió una escala de valor de 10 puntos, tomando el 5 como punto medio y el 1 y el 10 como extremos de la bipolaridad. Las tres versiones antes mencionadas corresponden a habla sintetizada sin incluir reglas sobre la duración de las pausas ortográficas, habla sintetizada incluyendo las reglas de duración, y como referencia de las anteriores un habla natural pre-grabada. En cada una de las condiciones se reproducía el mismo texto.

La prueba está dividida internamente en tres bloques. En el primer bloque se pide a los informantes que opinen sobre el habla que acaban de escuchar. Los resultados se presentan en la figura 4. En esta figura y en las siguientes se dan las puntuaciones medias obtenidas a partir de los 10 informantes.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

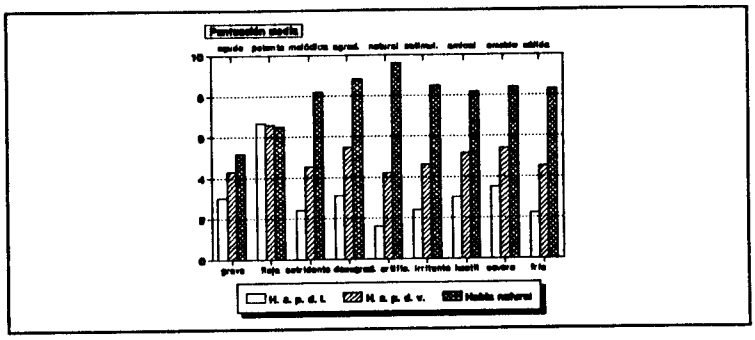

Fig. 4. *Histograma en el que se muestran las puntuaciones medias asociadas a cada pareja de opciones para los tres tipos de habla analizados (bloque 1 de la prueba). (H.s.p.d.i: Habla sintetizada con pausas de duración invariable. H.s.p.d.v: Habla sintetizada con pausas de duración variable según el signo de puntuación)*

En la figura 4 se puede observar una mejora notable, una vez aplicado el sistema de reglas propuesto, puesto que en la mayoría de las opciones presentadas los valores se acercan a los términos positivos, esto es, "más amical", "más amable", "más cálida",...

El segundo bloque de la prueba se refiere a la manera de leer el texto en las tres condiciones. Los resultados se presentan en la figura 5.

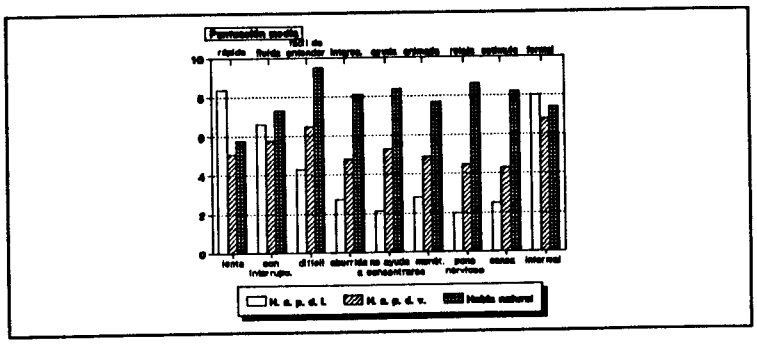

Fig. 5. *Histograma en el que se muestran las puntuaciones medias asociadas a cada pareja de opciones para los tres tipos de habla analizados (bloque 2 de la prueba). (H.s.p.d.i: Habla sintetizada con pausas de duración invariable. H.s.p.d.v: Habla sintetizada con pausas de duración variable según el signo de puntuación)*

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

Por lo que respecta a la figura 5, se observa nuevamente una mejora en el sistema de síntesis con pausas de duración variable: el habla así sintetizada se percibe como más relejada, estimulante y formal, entre otras opciones.

El tercer bloque de la prueba valora el conjunto del sistema de síntesis desde el punto de vista del usuario. Los resultados se muestran en la figura 6.

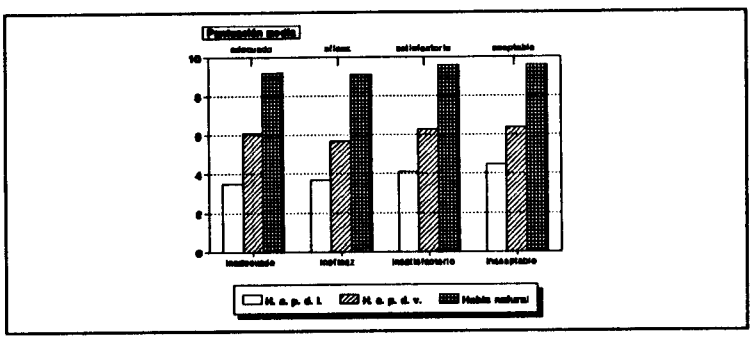

Fig. 6. *Histograma en el que se muestran las puntuaciones medias asociadas a cada pareja de opciones para los tres tipos de habla analizados (bloque 3 de la prueba). (H.s.p.d.i: Habla sintetizada con pausas de duración invariable. H.s.p.d.v: Habla sintetizada con pausas de duración variable según el signo de puntuación)*

El histograma que se muestra en la figura 6 es un resumen de los anteriores, mostrándose el nuevo sistema más adecuado, eficaz, satisfactorio y aceptable, respecto al sistema anterior. No obstante, en ningún caso se consigue la calidad que ofrece el habla natural.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

CONCLUSIÓN

A la vista de los resultados presentados, se puede concluir que las reglas propuestas en este estudio optimizan el funcionamiento del conversor. Como se observa en la figura 6, los resultados globales obtenidos para el habla sintetizada con pausas de duración variable según el signo de puntuación mejoran notablemente los del habla sintetizada con asignación de pausas invariable -valor único para cualquier signo de puntuación-. No obstante, no se consigue aún la calidad del habla natural, porque las pausas no son evaluables separadamente del resto de los aspectos de la prosodia. Por consiguiente, el objetivo de futuros estudios será mejorar el módulo prosódico, a fin de acercarse a una calidad más cercana a la del habla natural.

AGRADECIMIENTOS

Al Dr. Joaquim Llisterri, a Juan M. Garrido y a Francesc Gudayol, por sus continuos consejos y la ayuda prestada en todo momento.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994)
"Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf

REFERENCIAS

Boomer, D.S. (1965). "Hesitation and grammatical encoding". **Language and Speech**, 8.

Caldognetto, E.M. (1992). "Le pause quali indici diagnostici per lo stile del parlato spontaneo". **Quaderni del Centro di Studio per le Ricerche di Fonetica**, XI.

Fernández, N. (1992). "Un test de evaluación subjetiva de la calidad de la síntesis". Universitat Autònoma de Barcelona, Departament de Filologia Espanyola, ms no publicado.

Goldman-Eisler, F. (1972). "Pauses, clauses, sentences". **Language and Speech**, 15.

Jiménez, D. (1992). "Estudio y parametrización de las pausas ortográficas en castellano", Departamento de Acústica EUETT La Salle, ms no publicado.

Navarro Tomás, T. (1918). **Manual de pronunciación española**. Madrid: CSIC (Publicaciones de la Revista de Filología Española III). 1989. 23 edición.

O'Connell, D. - Kowal, S. "Prospectus for a science of pausology" en **Temporal variables in Speech**, H. W. Dechert - M. Raupach Eds., Ed. Mouton, 1980.

Puigví D. - Fernández J.M. (1992). "Estudi de la durada i la situació de les pauses en castellà". Universitat Autònoma de Barcelona, Departament de Filologia Espanyola, ms no publicado.

REAL ACADEMIA ESPAÑOLA (1973). **Esbozo para una nueva gramática de la lengua española**. Madrid. Duodécima reimpresión: enero, 1989.

PUIGVÍ, D.- JIMÉNEZ, D.- FERNÁNDEZ, J. M. (1994) "Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto a habla", in Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 de julio de 1994.
http://liceu.uab.es/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf