

Evaluación del conversor de texto en habla con modelo HNS, y comparación con el modelo MBROLA

Jon Sanchez, Eva Navas, Ibon Saratxaga, Inma Hernáez

Euskal Herriko Unibertsitatea
{ion,eva,ibon,inma}@bips.bi.ehu.es

Abstract

This paper describes the test performed to compare two different speech synthesis motors. 22 people took part in the experiment, in 7 different tests: vowels, consonants, consonant groups, sentences, text and global quality. The results can help improving the speech synthesis quality.

Laburpena

Lan honetan ahozko sintesirako garatutako motoreak konparatzeko egindako frogak deskribatzen dira. Esperimentuan hogeita bi lagunek parte hartu dute, zazpi froga desberdinetan: bokalak, kontsonanteak eta kontsonante taldeak, esaldiak (zentsudunak eta zentzugabeak), testuak eta kalitate orokorra. Emaitzak interesgarriak dira sintesi motore bien kalitatea lantzeko.

Keywords: TTS, HNS, MBROLA, subjective tests

Hitz gakoak: Testua hots bihurketa, HNS, MBROLA, froga subjektiboak

1. Introducción

Hoy en día gracias al avance de las Tecnologías del Habla es cada vez más habitual recibir informaciones mediante interfaces orales. Sistemas telefónicos de información, megafonías en aeropuertos y estaciones de tren, sistemas GPS de guiado por voz, sistemas de ayuda para discapacitados visuales... todos ellos utilizan la voz para transmitir información. Dependiendo de la complejidad del sistema, en ocasiones dicha información puede transmitirse mediante grabaciones concatenadas, si bien en otras ocasiones, la variabilidad de la información a transmitir requerirá sistemas de Conversión de Texto en Habla.

Para la utilización de dichos sistemas, es importante que la calidad de la voz sea suficiente, de manera que los usuarios se sientan cómodos al oír, especialmente si han de utilizarse esos sistemas durante periodos de tiempo largos. Esta medida de la calidad se traduce en dos aspectos concretos de la voz la inteligibilidad y la naturalidad.

Hay que tener en cuenta que, si bien existen métodos cuantitativos para evaluar esta calidad, al final que una voz sea comprensible o no, o agradable o no, depende en última medida del usuario del sistema, que ha de oír y entenderla, y saber si le gusta o no. Por esto, es necesario plantear sistemas subjetivos de evaluación de la calidad.

El objetivo de este trabajo es la realización de una prueba subjetiva de calidad de la voz generada por el sistema AhoTTS (Hernáez et al., 2001). Para ello se realizará el estudio de un sistema de evaluación, y su migración a una plataforma web que permita el acceso y

ejecución remotos, así como el almacenamiento de los resultados en bases de datos MySQL. También se aborda el reclutamiento de evaluadores, la evaluación en sí misma, y el análisis de los resultados obtenidos.

En la sección 2 de este documento se describen las pruebas que se realizan, y la implementación de dichas pruebas. Esto incluye también las condiciones que han cumplido los evaluadores que han realizado las pruebas.

En la sección 3 se da cuenta de los resultados obtenidos en las distintas pruebas. También se menciona el desarrollo del sistema de procesado de datos en la evaluación.

2. Diseño de las pruebas

Para realizar la evaluación se le van presentando al evaluador una serie de estímulos sonoros. Dado que están diseñados para evaluar un aspecto concreto del habla, hay tres tipos de estímulos diferentes: grupos segmentales, frases completas, o textos. La naturaleza de las pruebas se describe en el apartado "Pruebas a realizar".

Estas pruebas fueron originariamente diseñadas para que los usuarios pudieran realizarlas utilizando un reproductor de cassette y apuntando los resultados en papel, si bien actualmente se dispone de medios más eficientes. Para llegar hasta un número mayor de potenciales evaluadores, se ha utilizado un sistema basado en Internet. Del trabajo realizado para la configuración del sistema web, adaptación de las pruebas de evaluación al nuevo medio, así como los problemas aparecidos en el proceso, que ha sido necesario resolver, se trata en el punto 2.2. Por último, las personas que han de realizar la evaluación deben

cumplir una serie de requisitos, que también se detallan en este capítulo.

2.1. Pruebas a realizar

Todas las pruebas de que constará esta evaluación fueron en principio ideadas para otros idiomas. Dado que nosotros debemos evaluar la calidad de un Conversor de Texto en Habla en euskera, es necesario utilizar una versión adaptada a dicho idioma. Para ello se utiliza la adaptación que se desarrolló en (Hernández, 1999), en la cual se estudiaron, además de los estímulos a generar, las opciones a presentar al evaluador, los alófonos que se evalúan en cada prueba, y las matrices de confusión a generar para la exposición de resultados.

2.1.1. Prueba de inteligibilidad de consonantes

Esta prueba está basada en las DRT¹ (Voiers, 1984) y MRT² (FairBanks, 1958), con textos adaptados al euskera. La idea de estas pruebas consiste básicamente en presentar al evaluador estímulos sonoros consistentes en palabra monosilábicas, de estructura CVC, CV, o, según modificaciones posteriores, VCV³. Para cada estímulo oído, el evaluador puede anotar lo que ha oído (prueba abierta) o bien elegir entre diferentes opciones (prueba cerrada). En nuestro caso se utiliza este segundo tipo de prueba, presentando al evaluador 4 opciones para elegir, de las cuales una corresponde realmente con la palabra escuchada. Las otras tres difieren solamente en un alófono, precisamente el sonido que se va a evaluar. Los distintos sonidos se evalúan tanto en posición central (estructura VCV) como en posiciones inicial y final (estructura CVC).

Los sonidos evaluados son los siguientes⁴:

En posición central: m, p, B, f, T, t, D, n, rr, l, s, s', J, tS, jj, S, c, k, G, L, r, ts, ts', x

En posición inicial: m, p, b, f, T, t, d, n, rr, l, s, s', J, tS, S, c, gj, k, g, L

En posición final: B, p, f, t, D, n, rr, l, s, s', ts, ts', tS, jj, S, k, G

Para realizar estas evaluaciones, a cada evaluador se le presenta una batería de 220 estímulos a evaluar, dividida en 11 sesiones de 20 estímulos. De ellos, 100 estímulos se refieren a evaluaciones en posición central, 71 a evaluaciones en posición inicial, y 49 a evaluaciones en posición final.

2.1.2. Prueba de inteligibilidad de grupos consonánticos

De la misma manera que en el apartado anterior se evaluaban consonantes, y dado que el entorno específico usado (fuera del habla real) nos da información sobre las consonantes aisladas exclusivamente, se desarrolla esta otra prueba, más corta que la anterior pero igualmente basada en MRT y DRT. En ella se presentan al evaluador 5 series de 5 estímulos, en las cuales puede elegir entre 4 opciones distintas, difiriendo éstas en el grupo consonántico inicial.

Los grupos evaluados en esta prueba son: br, bl, dr, fr, fl, gr, gl, kr, kl, pr, pl, tr, tl

2.1.3. Prueba de inteligibilidad de vocales

En esta prueba, basada al igual que las anteriores en DRT y MRT, evalúa combinaciones de dos vocales.

Se presentan al evaluador 5 sesiones de 5 estímulos cada una, presentando cada uno de los estímulos una estructura Consonante – Vocal – Vocal.

Las combinaciones de vocales evaluadas en esta prueba son las siguientes: ae, ai, ao, au, ei, eo, eu, ia, ii, oi, ui.

2.1.4. Prueba de inteligibilidad de palabras en frases con sentido

Dado que las pruebas anteriores se realizan con estímulos muy cortos, y son de selección cerrada, se realiza esta otra prueba, con estímulos más similares al habla real, donde se evalúa la inteligibilidad de palabras en frases con sentido. Para este tipo de pruebas se utilizan normalmente las denominadas Frases Psicoacústicas de Harvard (Egan, 1948), que son oraciones con sentido, con estructura sintáctica correcta y variada, y fonéticamente equilibradas. Para esta prueba en particular se ha utilizado la batería de frases realizada para el euskera en (Hernández, 99), obtenidas en base a textos de la revista 'Argia' y equilibradas fonéticamente.

Se presentan al evaluador 10 sesiones con 10 estímulos cada una. Cada estímulo es una frase completa. En este caso, en lugar de ofrecer al evaluador, como en las pruebas anteriores, una serie de estímulos para que elija, se le ofrece un hueco en blanco en el que debe escribir la frase que ha oído.

2.1.5. Prueba de inteligibilidad de palabras en frases sin sentido

En la prueba de inteligibilidad de palabras en frases con sentido el evaluador puede apoyarse en la información contextual de las oraciones, para rellenar lagunas de comprensión en las frases. Para poder evaluar la influencia de este hecho en la comprensión

¹ Diagnostic Rhyme Test

² Modified Rhyme Test

³ Donde C se refiere a consonante y V a vocal.

⁴ Se utiliza como código SAMPA de referencia el desarrollado en (Aholab, 2004)

suele utilizarse otra prueba que se conoce como Frases Semánticamente Anómalas de Haskins (Nye and Gaitenby, 1974) . Las frases que componen la prueba son sintácticamente correctas, pero son semánticamente incorrectas, por lo que la información extraíble del contexto disminuye mucho. De nuevo se utilizará para esta prueba de evaluación la adaptación al euskera realizada en (Hernández, 1999). Las frases estímulo se encuentran, al igual que en el caso de la prueba anterior, equilibradas fonéticamente.

2.1.6. Prueba de comprensión de textos

Con esta prueba se pretende evaluar la capacidad del sistema de síntesis en la transmisión de la información. Se trata de que el oyente comprenda los conceptos que se pretenden transmitir con el texto, por lo que los resultados obtenidos son un buen indicador de la calidad global del sistema. La prueba original descrita por Pisoni (1987) comprendía 15 textos con sus correspondientes preguntas. En este caso, para mantener el tiempo de realización de la prueba de comprensión dentro de unos límites, se ha optado por limitar a tres el número de textos (obtenidos de Hernández, 1999), con su correspondiente batería de preguntas.

Para cada texto se realizan 5 preguntas de comprensión, y cada una de las preguntas presenta 4 posibles respuestas, entre las que el evaluador debe elegir.

Se presentan al evaluador 50 estímulos de este tipo, repartidos en 5 sesiones de 10 frases cada una.

2.1.7. Prueba de calidad global

Una vez se han completado todas las pruebas de inteligibilidad, con esta prueba de calidad global se pretende pulsar la opinión del evaluador respecto a la posibilidad de utilizar un sistema de síntesis de voz, centrándose en aspectos tales como la naturalidad y la aceptabilidad del sistema por parte del usuario.

Se presenta al evaluador un estímulo en forma de texto, y varias preguntas en las que se ofrece dos valoraciones contrapuestas. Las opciones entre las que se puede elegir son cinco, estableciendo una gradación entre ambos valores extremos contrapuestos. Por último, ya sin estímulo, se le hacen preguntas al evaluador para medir la utilidad en general de los sistemas de síntesis de voz en diferentes supuestos. Asimismo, también se deja al evaluador un espacio para que pueda opinar sobre la prueba realizada, o sobre el sistema de síntesis.

2.2. Adaptación del sistema para su utilización vía Internet

Para llevar a cabo la evaluación se ha recurrido a la realización de las pruebas a través de Internet, dado que la red de redes, al estar muy extendida, permite llegar a más evaluadores potenciales. Para ello se ha utilizado la

plataforma descrita en (Martín et al., 2004) de manera que todas las pruebas que se han descrito en el apartado anterior puedan realizarse en base a páginas web.

La plataforma seleccionada permite el diseño de pruebas de evaluación vía web, en las que se presentan siempre estímulos sonoros para evaluar. Esta evaluación se puede realizar de tres maneras:

1. En el tipo de formulario número 1, se presentan al evaluador 2 estímulos, debiendo elegir cual de ellos prefiere (por resultarle más agradable, ajustarse más a lo pedido, o el criterio que se quiera establecer). Este formulario no es necesario para la prueba que tratamos en este trabajo.

2. En el segundo tipo de formulario, para cada estímulo de la prueba se dan al usuario varias opciones entre las que debe elegir, tal y como se ve en la Figura 1 (en nuestro caso, las distintas opciones serán las variantes entre las que el evaluador debe elegir⁵, tanto en las Pruebas de Inteligibilidad de Elementos Segmentales, como en la Prueba de Comprensión de Textos y en la Prueba de Calidad Global).

3. Este tercer tipo de formulario provee al evaluador de un espacio en blanco donde transcribir lo que ha oído en el estímulo correspondiente, tal y como se muestra en la Figura 2. En nuestro caso se utilizará para realizar la Prueba de Inteligibilidad de Palabras en Frases, tanto con sentido como sin sentido.

⁵ Existen dos tipos diferentes de formulario tipo 2. En el primero de ellos, las opciones entre las que el evaluador puede elegir son las mismas para todos los estímulos de la sesión. En el segundo, se puede configurar, utilizando adecuadamente el fichero npunt.txt correspondiente, una batería de opciones para cada uno de los estímulos. Es este segundo tipo de formulario 2 el que debemos utilizar.

Figura 1: Formulario de evaluación tipo 2

Figura 2: Formulario de evaluación tipo 3

Una vez que el evaluador ha completado el formulario, y pulsado el botón “Bidali” presente en todos los formularios (ver Figura 1 y Figura 2), los datos se envían al servidor utilizando un software PHP, el cual accede a una base de datos relacional MySQL (Gilmore, 2001) que almacena los resultados en

distintas tablas. Adicionalmente, los resultados se almacenan también en ficheros del sistema.

La adaptación de los formularios descritos a la prueba de evaluación que se ha de realizar se lleva a cabo de la siguiente manera:

- Las pruebas de inteligibilidad de elementos segmentales (consonantes, grupos consonánticos, grupos de vocales) se realizan mediante formularios de tipo 2, que son los que presentan múltiples opciones. Para ello ha sido necesario hacer ciertas modificaciones al sistema: la primera, que la plataforma estaba prefijada para presentar dos estímulos por fila, lo cual, al ser utilizado en sesiones de 5 estímulos, producía resultados inadecuados. La segunda, que una de las características más destacables de la plataforma, el hecho de aleatorizar tanto el orden de los estímulos de una misma sesión, como el orden de las opciones para un mismo estímulo, estaba produciendo formularios incorrectos al aleatorizar por separado el estímulo y su correspondiente batería de opciones. Ambos problemas se solucionaron con pequeñas modificaciones en los códigos PHP.

- Las pruebas de inteligibilidad de palabras en frases, tanto para el caso de frases con sentido como para el caso de frases sin sentido, se realizan mediante formularios de tipo 3. En este tipo de formularios, el evaluador puede oír el estímulo, mientras junto a su icono aparece un cuadro de texto dispuesto para ser rellenado con lo que se ha escuchado. También para el uso de este tipo de formulario se han necesitado realizar pequeños cambios en la plataforma. En este caso, no en el código PHP, sino en el diseño de las Bases de Datos de MySQL. Uno de los campos de la base de datos estaba dimensionado incorrectamente, de manera que no se almacenaba en su totalidad la frase que el evaluador había tecleado. La solución consistía en configurar correctamente en la base de datos el tamaño del campo que debía almacenar las frases recogidas.

- La prueba de comprensión de textos también se realizó utilizando formularios de tipo 2. En este caso, se realiza una sesión por página y un texto por sesión, de manera que aparecía, junto a cada una de las cinco preguntas, el enlace para lanzar el estímulo, si bien los cinco estímulos eran en realidad el mismo.

- Por último, las pruebas de valoración se realizan mediante un formulario de tipo 2, así como utilizando el formulario que la propia plataforma provee para realizar encuestas como punto final de la evaluación.

El resultado final es una encuesta, que se puede realizar a través de Internet, consistente en 42 formularios: 26 de tipo 2 (11 para la prueba de inteligibilidad de consonantes, 5 para la de grupos

consonánticos, 5 para la de grupos de vocales, 3 para la comprensión de texto, 1 para la consulta de calidad global, y otro formulario más, el primero de todos, que se utiliza a modo de ejemplo de utilización, cuyos resultados no se computan), 15 de tipo 3 (10 para la prueba de inteligibilidad de frases con sentido, y 5 para la de frases sin sentido) y un formulario más que alberga la consulta final al evaluador.

La duración total de la prueba de evaluación se estima en unas dos horas, si bien se observa gran variabilidad entre los distintos evaluadores: mientras algunos han sido capaces de completar la prueba completa en escasos 90 minutos, otros han necesitado más de 4 horas para completarla⁶.

2.3. Perfil del evaluador

En las pruebas de evaluación han participado 22 evaluadores, tanto hombres como mujeres, de edades comprendidas entre 22 y 59 años, procedentes de diferentes ámbitos (educación, incluyendo tanto alumnos como profesores, administración, sanidad, gestión, empresas productivas,...). Las condiciones que se requerían para ser evaluador son las siguientes:

- No tener problemas de audición.
- No tener problemas de comprensión del euskera⁷.
- No tener experiencia previa en el uso de habla sintetizada⁸.

En lo que se refiere a esta última condición, hay que comentar que se realizó, con el fin de comprobar el buen funcionamiento del propio sistema de evaluación, una batería de pruebas de evaluación con sujetos que sí estaban acostumbrados al habla sintética. Estos datos pueden ser útiles de cara a evaluar el efecto aprendizaje.

3. Resultados

Una vez realizadas las evaluaciones a través de Internet, es necesario procesar los resultados obtenidos. La plataforma de evaluación provee dos tipos de

salidas: una en ficheros de texto y otra mediante bases de datos MySQL. La interfaz web Phpmysql (http://www.phpmyadmin.net) permite realizar búsquedas en la base de datos MySQL recogida, y exportarla en varios formatos diferentes. Se ha optado por exportar los datos de los diferentes evaluadores al formato Excel, de manera que en esta hoja de cálculo se automatiza después la obtención de matrices de confusión y recuentos de aciertos y fallos para la exposición de resultados de la evaluación.

En este punto se hace necesario recordar que, como bien indica el propio título del trabajo de investigación, en este trabajo se están evaluando dos métodos diferentes de síntesis, MBROLA y HNS. Es por ello que se han diseñado dos pruebas diferentes, en las que cambia el estímulo. Es decir, se utiliza uno u otro método de síntesis, pero el resto de parámetros (textos, orden de los textos, o incluso entonación⁹) se mantienen iguales.

De cara a los evaluadores, éstos se han dividido en dos grupos, realizando cada uno de los dos grupos una evaluación distinta. De nuevo, los textos y las sesiones son las mismas, pero en lo que se refiere a los estímulos, aquellos que en el primer grupo corresponden a HNS, en el segundo son estímulos del modelo MBROLA, y viceversa. De esta manera, se consigue que todos los evaluadores lleven adelante parte de HNS y parte de MBROLA, consiguiendo que los resultados se obtengan de una muestra mayor de evaluadores. Esto implica que la primera operación de proceso que ha de hacerse en Excel es organizar las bases de datos obtenidas de manera que se consiga agrupar los datos correspondientes a cada uno de los modelos. Una vez organizados los resultados de esa manera, se sigue adelante con las evaluaciones de las distintas pruebas.

3.1. Prueba de inteligibilidad de consonantes

El procesado de datos para la obtención de las matrices de confusión de esta prueba consta de dos partes:

La salida de la base de datos nos da el número de respuesta que el evaluador ha elegido. Es necesario contar esos números, para saber cuántas respuestas ha habido correspondientes a cada opción.

Una vez contadas, se elabora la matriz de confusión sumando los datos de respuestas dadas para cada una de las opciones.

Las matrices de confusión presentan en la columna izquierda el sonido que contenía el estímulo, y que debía ser identificado, y en la fila superior, el sonido que realmente ha sido identificado por el evaluador, de

⁶ Estas duraciones han sido observadas acompañando a los evaluadores durante la prueba de evaluación.

⁷ Si bien en (Hernández, 1999) se recomienda que los evaluadores deberían tener el euskera como lengua materna, para esta evaluación se ha aceptado que euskaldunberri ejerzan de evaluadores, siempre que se de el caso de que utilicen la lengua vasca habitualmente y con fluidez, además de la condición de no presentar problemas en su comprensión.

⁸ Esta condición se impone por el fuerte efecto aprendizaje que se da en los sistemas orales: cuando un usuario se acostumbra a un determinado tipo de voz, es capaz de adaptarse a ella, mejorando sensiblemente los resultados. Por esta misma causa, cada usuario debe realizar la prueba solamente una vez, dado que, de repetirse, el estar el oído más acostumbrado a las voces sintéticas, hará que los resultados mejoren sin que el sistema haya cambiado.

⁹ Para estímulos homólogos de ambas pruebas, se genera una única curva de entonación, y un juego único parámetros prosódicos, generados según (Navas, 2002).

manera que los resultados correctos se muestran en la diagonal principal de la matriz.

A continuación se exponen las matrices de confusión de la prueba de inteligibilidad de

consonantes, tanto en MBROLA como en HNS, y tanto en valores absolutos como relativos, para las consonantes evaluadas en las tres posiciones: inicial, central y final.

3.1.1. MBROLA

	m	p	B	f	T	t	D	n	rr	L	s	s`	J	tS	jj	S	c	k	g	G	L	r	ts	ts`	x
m	37							3		1												0			
p	0	40				1					0							0		0		0			
B		0	37			0	2			0								1		0		2			
f				37		2	0	0		0	2							1		0		0			
T			4	20	17	1					0					0									
t	1	0	0	7		41	0				0							0		0		0			
D			7	1		0	31			1								0		1					
n	4							35		2			0									0			
rr	0							0	42	0										0		0			
l	0							1		53												0	0		
s		0	0	1		0	0	0				39				1		0		1		0			
s`		0	2	0		0	0						38	1		1				0					
J	0							0	0				43								0	0			
tS				0	0	0					0	0		41		0		0							
jj				0	0						0	0			42			0		0					
S		0		0	0	0	0				0	0				40		0		1		0			
c		0				0	0				0					2	40	0		0					
k		0	0			1				0							0	39		0		2			
g																									
G	3	0	0	1		0												0		37		0			
L	0							2		2										7	20	0			
r	0					1	1	0	0	2									0	0		78			
ts				4		0					0	0				0				0			37		
ts`	1			1		0					0	0				0	0					0		39	
x				3		0					0	0				2	0		1						35

Tabla 1: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición central, con estímulos MBROLA. Valores absolutos.

	m	p	b	f	T	t	d	n	rr	l	s	s'	J	tS	S	c	gj	k	g	L
m	41					0		1	0	0								0		
p		18	1			2					0							0		
b	2	0	68	0		1					0							1	1	
f				21	0	0					0							0	0	
T		0		12	18	2					3	3						5		
t				0		20												0	0	
d	13	0		0		0	42			2	1							0	6	
n	4					0		16		0	0									
rr		3				0			17	0									1	
l	0							0		21	0									0
s		0				0				0	20			1				0	0	
s'				0		1						30		0				0	0	
J	1			0				0	0				42							
tS		0				0					2	3		16					0	
S			0	0							4			0	17					
c		0		0		3				1						32		1	5	
gj	0	0		4		1					1	1		2			74	1	0	
k		0	0			0												23	8	
g		2	0	0		4				5		0						1	50	
L	2							5		1								0		34

Tabla 2: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición inicial, con estímulos MBROLA. Valores absolutos.

	B	p	f	t	D	n	rr	l	s	s'	ts	ts'	tS	jj	S	k	G
B	37		2	1		0			0							2	
p		16				0	1	0								3	
f			15	0		0		2	3							0	
t				19		0	2	0							0	0	
D		1		3	8	7	22	2								0	
n						31	0	0	0								
rr				1		1	19	0								0	
l				0		0	0	20	0							0	
s				0		0		0	22							0	
s'			0			0	0			21				0		0	
ts				13		0			0		30					0	
ts'			0	6		0			1	5		31			0	0	
tS				2		0	0		3				16			0	
jj			0	2		5		11	0					44		0	
S			0	0		0		0					0		21	1	
k	6	1		0		0	8									7	
G			0	8		5	1		0							2	26

Tabla 3: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición final, con estímulos MBROLA. Valores absolutos.

	m	p	b	f	T	t	d	n	rr	l	s	s`
m	90,24	0,00	0,00	0,00	0,00	0,00	0,00	7,32	0,00	2,44	0,00	0,00
p	0,00	97,56	0,00	0,00	0,00	2,44	0,00	0,00	0,00	0,00	0,00	0,00
b	0,00	0,00	88,10	0,00	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00
f	0,00	0,00	0,00	88,10	0,00	4,76	0,00	0,00	0,00	0,00	4,76	0,00
T	0,00	0,00	9,52	47,62	40,48	2,38	0,00	0,00	0,00	0,00	0,00	0,00
t	2,04	0,00	0,00	14,29	0,00	83,67	0,00	0,00	0,00	0,00	0,00	0,00
D	0,00	0,00	17,07	2,44	0,00	0,00	75,61	0,00	0,00	2,44	0,00	0,00
n	9,76	0,00	0,00	0,00	0,00	0,00	0,00	85,37	0,00	4,88	0,00	0,00
rr	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
l	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,85	0,00	98,15	0,00	0,00
s	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00	92,86	0,00
s`	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	90,48
J	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tS	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
jj	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
S	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
c	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
k	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00
G	7,32	0,00	0,00	2,44	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	6,45	0,00	6,45	0,00	0,00
r	0,00	0,00	0,00	0,00	0,00	1,22	1,22	0,00	0,00	2,44	0,00	0,00
ts	0,00	0,00	0,00	9,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ts`	2,44	0,00	0,00	2,44	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
x	0,00	0,00	0,00	7,32	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

	J	tS	jj	S	c	k	G	L	r	ts	ts`	x
m	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
b	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	4,76	0,00	0,00	0,00
f	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00
T	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
D	0,00	0,00	0,00	0,00	0,00	0,00	2,44	0,00	0,00	0,00	0,00	0,00
n	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rr	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s	0,00	0,00	0,00	2,38	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00
s`	0,00	2,38	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
J	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tS	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
jj	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
S	0,00	0,00	0,00	97,56	0,00	0,00	2,44	0,00	0,00	0,00	0,00	0,00
c	0,00	0,00	0,00	4,76	95,24	0,00	0,00	0,00	0,00	0,00	0,00	0,00
k	0,00	0,00	0,00	0,00	0,00	92,86	0,00	0,00	4,76	0,00	0,00	0,00
G	0,00	0,00	0,00	0,00	0,00	0,00	90,24	0,00	0,00	0,00	0,00	0,00
L	0,00	0,00	0,00	0,00	0,00	0,00	22,58	64,52	0,00	0,00	0,00	0,00
r	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,12	0,00	0,00	0,00
ts	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	90,24	0,00	0,00
ts`	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,12	0,00
x	0,00	0,00	0,00	4,88	0,00	0,00	2,44	0,00	0,00	0,00	0,00	85,37

Tabla 4: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición central, con estímulos MBROLA. Valores relativos.

	m	p	b	f	T	t	d	n	rr	l	s	s'	J	tS	S	c	gj	k	g	L
m	97,62	0,00	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p	0,00	85,71	4,76	0,00	0,00	9,52	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
b	2,74	0,00	93,15	0,00	0,00	1,37	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,37	1,37	0,00
f	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T	0,00	0,00	0,00	27,91	41,86	4,65	0,00	0,00	0,00	0,00	6,98	6,98	0,00	0,00	0,00	0,00	0,00	11,63	0,00	0,00
t	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
d	20,31	0,00	0,00	0,00	0,00	0,00	65,63	0,00	0,00	3,13	1,56	0,00	0,00	0,00	0,00	0,00	0,00	0,00	9,38	0,00
n	20,00	0,00	0,00	0,00	0,00	0,00	0,00	80,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rr	0,00	14,29	0,00	0,00	0,00	0,00	0,00	0,00	80,95	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,76	0,00	0,00
l	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,24	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00	0,00
s'	0,00	0,00	0,00	0,00	0,00	3,23	0,00	0,00	0,00	0,00	0,00	96,77	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
J	2,33	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	97,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tS	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	9,52	14,29	0,00	76,19	0,00	0,00	0,00	0,00	0,00	0,00
S	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	19,05	0,00	0,00	0,00	80,95	0,00	0,00	0,00	0,00	0,00
c	0,00	0,00	0,00	0,00	0,00	7,14	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	76,19	0,00	2,38	11,90	0,00
gj	0,00	0,00	0,00	4,76	0,00	1,19	0,00	0,00	0,00	0,00	1,19	1,19	0,00	2,38	0,00	0,00	88,10	1,19	0,00	0,00
k	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	74,19	25,81	0,00
g	0,00	3,23	0,00	0,00	0,00	6,45	0,00	0,00	0,00	0,00	8,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,61	80,65
L	4,76	0,00	0,00	0,00	0,00	0,00	0,00	11,90	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	80,95

Tabla 5: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición inicial, con estímulos MBROLA. Valores relativos

	B	p	f	t	D	n	rr	l	s	s'	ts	ts'	tS	jj	S	k	G
B	88,10	0,00	4,76	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,76	0,00
p	0,00	80,00	0,00	0,00	0,00	0,00	5,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	15,00	0,00
f	0,00	0,00	75,00	0,00	0,00	0,00	0,00	10,00	15,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t	0,00	0,00	0,00	90,48	0,00	0,00	9,52	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
D	0,00	2,33	0,00	6,98	18,60	16,28	51,16	4,65	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
n	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rr	0,00	0,00	0,00	4,76	0,00	4,76	90,48	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s'	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ts	0,00	0,00	0,00	30,23	0,00	0,00	0,00	0,00	0,00	0,00	69,77	0,00	0,00	0,00	0,00	0,00	0,00
ts'	0,00	0,00	0,00	13,95	0,00	0,00	0,00	0,00	2,33	11,63	0,00	72,09	0,00	0,00	0,00	0,00	0,00
tS	0,00	0,00	0,00	9,52	0,00	0,00	0,00	0,00	14,29	0,00	0,00	0,00	76,19	0,00	0,00	0,00	0,00
jj	0,00	0,00	0,00	3,23	0,00	8,06	0,00	17,74	0,00	0,00	0,00	0,00	0,00	70,97	0,00	0,00	0,00
S	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,45	4,55	0,00
k	27,27	4,55	0,00	0,00	0,00	0,00	36,36	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	31,82	0,00
G	0,00	0,00	0,00	19,05	0,00	11,90	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,76	61,90

Tabla 6: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición final, con estímulos MBROLA. Valores relativos.

Del estudio de los valores para MBROLA extraemos las siguientes valoraciones:

Se han estudiado 24 alófonos en posición central, de los cuales 21 son correctamente reconocidos al menos en el 80% de los casos. De ellos, solamente 4 consiguen una tasa de reconocimiento del 100% ([rr], [J], [tS], [jj]). El sonido que peores resultados ha obtenido ha sido el fricativo sordo interdental [T], que solamente ha sido identificado correctamente en un 40% de las evaluaciones, mientras que en un 47% de las apariciones ha sido tomado por el sonido fricativo sordo labiodental [f]. También el sonido palatal sonoro lateral [L] ha tenido un porcentaje de aciertos relativamente bajo (64%), siendo confundido en muchas ocasiones por el alófono [G], velar fricativo sonoro aproximante. El porcentaje global de sonidos correctamente identificados es del 89'59%, siendo el nivel de azar¹⁰ del 22%.

De los 20 alófonos estudiados en posición inicial, se encuentra que 17 de ellos han sido reconocidos correctamente en más del 80% de las ocasiones. Entre estos, 3 de ellos han sido reconocidos correctamente el 100% de las veces ([f], [t], [l]). De nuevo el sonido fricativo sordo interdental [T] ha sido el que menos veces ha sido correctamente reconocido,

sólo en un 41% de sus apariciones, y de nuevo ha sido confundido mayoritariamente por el sonido fricativo sordo labiodental [f]. El sonido plosivo sonoro dental [d] ha sido evaluado incorrectamente en un 34% de las ocasiones, siendo con mayor frecuencia confundido por el bilabial sonoro nasal [m]. El porcentaje global de sonidos correctamente identificados es del 83'33%, siendo el nivel de azar del 10%.

Por último, se evalúan en posición final 17 alófonos, de los cuales solamente 9 se han identificado correctamente en más de un 80% de las ocasiones. De ellos, 4 ([n], [l], [s], [s`]) se evalúan correctamente el 100% de las veces. El sonido menos identificado ha sido el representado por el código [D], dental sonoro aproximante, que ha sido equivocado en un 81% de las ocasiones. De hecho, ha sido identificado en más ocasiones como el alveolar sonoro [rr] que como [D]. También el sonido plosivo velar sordo [k] ubicado en posición final ha sido poco reconocido, siendo confundido muy frecuentemente con el bilabial sonoro aproximante [B] o con el alveolar sonoro [rr]. La tasa de reconocimientos correctos se coloca en el 74'22% para esta prueba, con un nivel de azar del 17%.

¹⁰ Entendiendo como tal, el nivel de aciertos cuando se responde al azar a la prueba de evaluación.

3.1.2. HNS

	m	p	B	f	T	t	D	n	rr	l	s	s'	J	tS	jj	S	c	k	G	L	r	ts	ts'	x
m	43							0		0											0			
p	0	42				0					0							1	0		0			
B		0	41			0	0			0								0	0		1			
f				37		2	0	0		0	3							0	0		0			
T			0	10	29	0					0					3								
t	0	0	0	3		38	1				1							1	1		0			
D			4	0		0	39			0								0	0					
n	5							37		0			0								1			
rr	0							0	41	0									0		1			
l	1							2		46											0	2		
s		0	0	0		0	0	0			41					1		0	0		0			
s'		0	0	0		0	0					42	0			0			0					
J	0							0	0				41						0		0			
tS				0	0	1					0	0		42		0		0						
jj				0		0					0	0			42			0	0					
S		0		0		0	0				1	0				42		0	0		0			
c		0				1	0				1					1	39	0	0					
k		1	0			2				0							0	37	1		1			
G	6	0	0	0		0												0	37		0			
L	0							0		0									0	32	0			
r	0					2	7	5	0	3									0		69			
ts				0		1					0	1				0			0			41		
ts'	0			0		0				0	0					0		0			0		43	
x				6		1				0	0					0		0	3					33

Tabla 7: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición central, con estímulos HNS. Valores absolutos.

	m	p	b	f	T	t	d	n	rr	l	s	s`	J	tS	S	c	gj	k	g	L
m	42					0		0	0	0								0		
p		10	5			4					0							2		
b	0	3	68	0		0					0							1	2	
f				16	0	1					1							0	3	
T		4		5	26	1					1	3						1		
t				5		9												0	1	
d	10	0		7		0	49			0	0							0	3	
n	2					0		20		0	0									
rr		1				1			18	1								0		
l	0							0		20	0								1	
s		0				0				0	19			2				0	0	
s`				2		4						15		9				0	2	
J	2			0				1		0			38							
tS		0				0					10	9		2				0		
S			0	0							5			1	15					
c		0		3		0				0						36		0	3	
gj	0	0		3		0					0	5		1			73	2	0	
k		1	1			4												18	8	
g		2	0	2		3				1		0						1	55	
L	2							2		0								0		38

Tabla 8: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición inicial, con estímulos HNS. Valores absolutos.

	B	p	f	t	D	n	rr	l	s	s`	ts	ts`	tS	jj	S	k	G
B	25		4	4		8			0							1	
p		19				1	0	0								2	
f			15	2		0		3	2							0	
t				17		0	2	1							0	1	
D		0		1	9	4	22	5								0	
n						30	2	0	0								
rr				1		1	14	1								4	
l				2		0	0	20	0							0	
s				1		0		0	19							0	
s`			2				0	1		18				0		0	
ts				1		0			0		40					0	
ts`			0	3		0			1	0		36			0	1	
tS				2		0	0		2				17			0	
jj			0	2		0		4	0					58		0	
S			0	0		0		0					1		19	0	
k	2	1		5		0	6									6	
G			2	3		5	0		1							1	30

Tabla 9: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición final, con estímulos HNS. Valores absolutos.

	m	p	B	f	T	t	d	n	rr	l	s	s`
m	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p	0,00	97,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
B	0,00	0,00	97,62	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
f	0,00	0,00	0,00	88,10	0,00	4,76	0,00	0,00	0,00	0,00	7,14	0,00
T	0,00	0,00	0,00	23,81	69,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t	0,00	0,00	0,00	6,67	0,00	84,44	2,22	0,00	0,00	0,00	2,22	0,00
d	0,00	0,00	9,30	0,00	0,00	0,00	90,70	0,00	0,00	0,00	0,00	0,00
n	11,63	0,00	0,00	0,00	0,00	0,00	0,00	86,05	0,00	0,00	0,00	0,00
rr	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	97,62	0,00	0,00	0,00
l	1,96	0,00	0,00	0,00	0,00	0,00	0,00	3,92	0,00	90,20	0,00	0,00
s	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	97,62	0,00
s`	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00
J	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tS	0,00	0,00	0,00	0,00	0,00	2,33	0,00	0,00	0,00	0,00	0,00	0,00
jj	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
S	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,33	0,00
c	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	2,38	0,00
k	0,00	2,38	0,00	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00	0,00
G	13,95	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
r	0,00	0,00	0,00	0,00	0,00	2,33	8,14	5,81	0,00	3,49	0,00	0,00
ts	0,00	0,00	0,00	0,00	0,00	2,33	0,00	0,00	0,00	0,00	0,00	2,33
ts`	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
x	0,00	0,00	0,00	13,95	0,00	2,33	0,00	0,00	0,00	0,00	0,00	0,00

	J	tS	jj	S	c	k	G	L	r	ts	ts`	x
m	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p	0,00	0,00	0,00	0,00	0,00	2,33	0,00	0,00	0,00	0,00	0,00	0,00
B	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	0,00
f	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T	0,00	0,00	0,00	7,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t	0,00	0,00	0,00	0,00	0,00	2,22	2,22	0,00	0,00	0,00	0,00	0,00
d	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
n	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,33	0,00	0,00	0,00
rr	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,38	0,00	0,00	0,00
l	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	3,92	0,00	0,00	0,00
s	0,00	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s`	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
J	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tS	0,00	97,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
jj	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
S	0,00	0,00	0,00	97,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
c	0,00	0,00	0,00	2,38	92,86	0,00	0,00	0,00	0,00	0,00	0,00	0,00
k	0,00	0,00	0,00	0,00	0,00	88,10	2,38	0,00	2,38	0,00	0,00	0,00
G	0,00	0,00	0,00	0,00	0,00	0,00	86,05	0,00	0,00	0,00	0,00	0,00
L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
r	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	80,23	0,00	0,00	0,00
ts	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,35	0,00	0,00
ts`	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
x	0,00	0,00	0,00	0,00	0,00	0,00	6,98	0,00	0,00	0,00	0,00	76,74

Tabla 10: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición central, con estímulos HNS. Valores relativos.

	m	p	b	f	T	t	d	n	rr	l	s	s`	J	tS	S	c	gj	k	g	L
m	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p	0,00	47,62	23,81	0,00	0,00	19,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	9,52	0,00	0,00
b	0,00	4,05	91,89	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,35	2,70	0,00
f	0,00	0,00	0,00	76,19	0,00	4,76	0,00	0,00	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	14,29
T	0,00	9,76	0,00	12,20	63,41	2,44	0,00	0,00	0,00	0,00	2,44	7,32	0,00	0,00	0,00	0,00	0,00	2,44	0,00	0,00
t	0,00	0,00	0,00	33,33	0,00	60,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	6,67	0,00
d	14,49	0,00	0,00	10,14	0,00	0,00	71,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,35
n	9,09	0,00	0,00	0,00	0,00	0,00	0,00	90,91	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rr	0,00	4,76	0,00	0,00	0,00	4,76	0,00	0,00	85,71	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
l	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,24	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,76
s	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	90,48	0,00	0,00	9,52	0,00	0,00	0,00	0,00	0,00	0,00
s`	0,00	0,00	0,00	6,25	0,00	12,50	0,00	0,00	0,00	0,00	0,00	46,88	0,00	28,13	0,00	0,00	0,00	0,00	6,25	0,00
J	4,88	0,00	0,00	0,00	0,00	0,00	0,00	2,44	0,00	0,00	0,00	0,00	92,68	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tS	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	47,62	42,86	0,00	9,52	0,00	0,00	0,00	0,00	0,00	0,00
S	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	23,81	0,00	0,00	4,76	71,43	0,00	0,00	0,00	0,00	0,00
c	0,00	0,00	0,00	7,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	85,71	0,00	0,00	7,14	0,00
gj	0,00	0,00	0,00	3,57	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,95	0,00	1,19	0,00	0,00	86,90	2,38	0,00	0,00
k	0,00	3,13	3,13	0,00	0,00	12,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	56,25	25,00	0,00
g	0,00	3,13	0,00	3,13	0,00	4,69	0,00	0,00	0,00	1,56	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,56	85,94	0,00
L	4,76	0,00	0,00	0,00	0,00	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	90,48

Tabla 11: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición inicial, con estímulos HNS. Valores relativos.

	B	p	f	t	D	n	rr	l	s	s`	ts	ts`	tS	jj	S	k	G
B	59,52	0,00	9,52	9,52	0,00	19,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,38	0,00
p	0,00	86,36	0,00	0,00	0,00	4,55	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	9,09	0,00
f	0,00	0,00	68,18	9,09	0,00	0,00	0,00	13,64	9,09	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t	0,00	0,00	0,00	80,95	0,00	0,00	9,52	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,76	0,00
D	0,00	0,00	0,00	2,44	21,95	9,76	53,66	12,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
n	0,00	0,00	0,00	0,00	0,00	93,75	6,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rr	0,00	0,00	0,00	4,76	0,00	4,76	66,67	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	19,05	0,00
l	0,00	0,00	0,00	9,09	0,00	0,00	0,00	90,91	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s	0,00	0,00	0,00	5,00	0,00	0,00	0,00	0,00	95,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
s`	0,00	0,00	9,52	0,00	0,00	0,00	0,00	4,76	0,00	85,71	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ts	0,00	0,00	0,00	2,44	0,00	0,00	0,00	0,00	0,00	0,00	97,56	0,00	0,00	0,00	0,00	0,00	0,00
ts`	0,00	0,00	0,00	7,32	0,00	0,00	0,00	0,00	2,44	0,00	0,00	87,80	0,00	0,00	0,00	2,44	0,00
tS	0,00	0,00	0,00	9,52	0,00	0,00	0,00	0,00	9,52	0,00	0,00	0,00	80,95	0,00	0,00	0,00	0,00
jj	0,00	0,00	0,00	3,13	0,00	0,00	0,00	6,25	0,00	0,00	0,00	0,00	0,00	90,63	0,00	0,00	0,00
S	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00	0,00	95,00	0,00	0,00
k	10,00	5,00	0,00	25,00	0,00	0,00	30,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	30,00	0,00
G	0,00	0,00	4,76	7,14	0,00	11,90	0,00	0,00	2,38	0,00	0,00	0,00	0,00	0,00	0,00	2,38	71,43

Tabla 12: Matriz de confusión de la prueba de inteligibilidad de consonantes evaluadas en posición final, con estímulos HNS. Valores relativos.

Del estudio de los valores para HNS extraemos las siguientes valoraciones:

De los 24 alófonos estudiados en posición central, 22 son correctamente reconocidos al menos en el 80% de los casos. De ellos, solamente 6 son correctamente reconocidos el 100% de las veces ([m], [s], [J], [jj], [L], [ts']). El sonido que peores resultados ha obtenido ha sido el fricativo sordo interdental [T], que ha sido identificado correctamente en un 69% de las evaluaciones, siendo muchas veces confundido con el sonido fricativo sordo labiodental [f]. El porcentaje global de sonidos correctamente identificados es del 91'62%, siendo el nivel de azar del 23'6%

Se han estudiado 20 alófonos en posición inicial, entre los cuales 11 han sido reconocidos correctamente en más del 80% de las ocasiones, siendo únicamente el sonido [m] reconocido correctamente el 100% de las veces. El sonido [tS], sordo aficado prepalatal frontal, ha sido el que menos veces ha sido correctamente reconocido, sólo en un 9'52% de sus apariciones. Ha sido confundido principalmente con los sonidos sordos alveolares fricativos [s] y [s'], teniendo ambos un peso del 42%, bastante más alto que el índice de aciertos. El porcentaje global de sonidos correctamente identificados es del 78'58%, siendo el nivel de azar del 9'37%

Por último, se evalúan en posición final 17 alófonos. En esta ocasión solamente 12 de los sonidos se han identificado correctamente en más de un 80% de las ocasiones, si bien ninguno ha sido evaluado siempre

correctamente. El sonido menos identificado ha sido el aproximante sonoro dental representado por el código [D], que ha sido equivocado en un 84% de las ocasiones, confundiendo principalmente con el sonido alveolar sonoro [rr]. También la [k] plosiva velar sorda en posición final ha sido poco reconocida (sólo en un 30% de las ocasiones), confundiendo principalmente con los sonidos [rr] y [t]. La tasa de reconocimientos correctos se coloca en el 76'41% para esta prueba, con un nivel de azar del 15'47%.

3.1.3. Comparando ambos modelos

En el caso de la prueba de inteligibilidad de consonantes, ambos modelos se comportan de manera parecida cuando se evalúan consonantes en posición intermedia, destacando ligeramente en modelo HNS (89'6% de acierto en MBROLA frente a 91'6% de acierto en HNS).

En la Figura 3 se puede ver la comparación alófono a alófono entre HNS y MBROLA, para la prueba de inteligibilidad de consonantes en posición intermedia. A primera vista se observa que, si bien la tasa de aciertos es similar en ambos casos, en el modelo MBROLA esos errores son particularmente graves en dos sonidos ([T] y [L]), mientras que el modelo HNS da unos resultados algo más equilibrados. El modelo HNS resulta ser algo peor que el modelo MBROLA solamente para los sonidos [rr], [l], [tS], [c], [k], [G], [r] y [x].

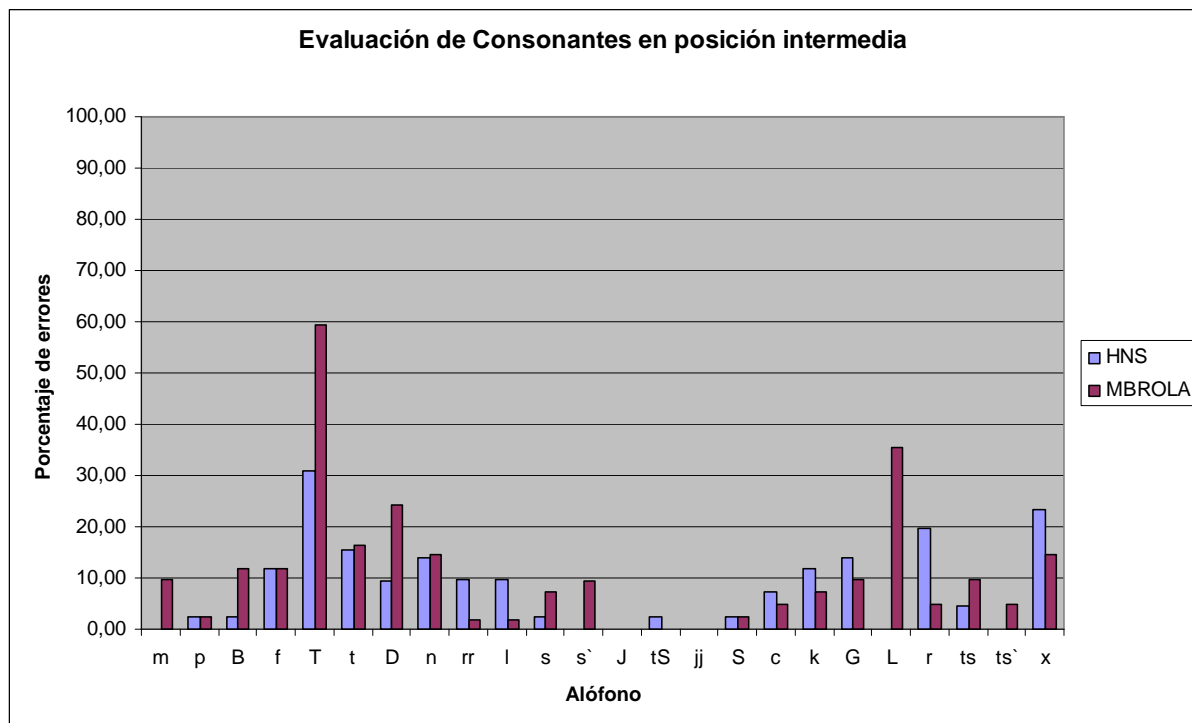


Figura 3: Comparación entre HNS y MBROLA para la prueba de inteligibilidad de consonantes en posición intermedia

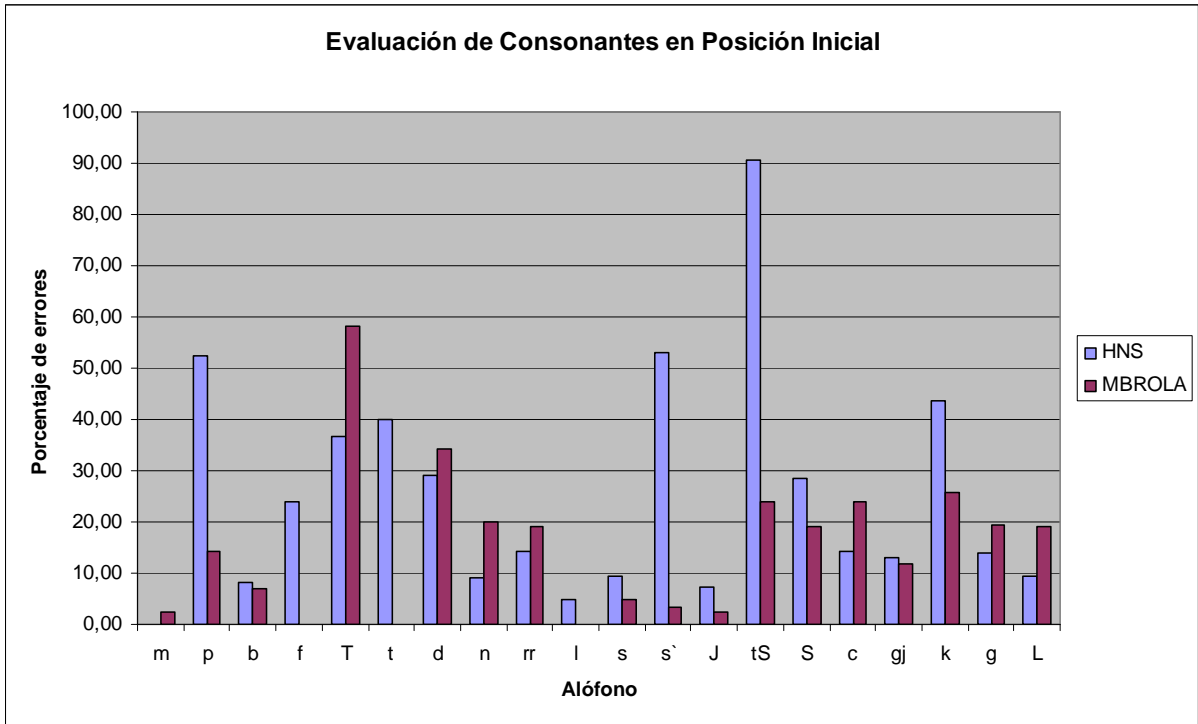


Figura 4: Comparación entre HNS y MBROLA para la prueba de inteligibilidad de consonantes en posición inicial

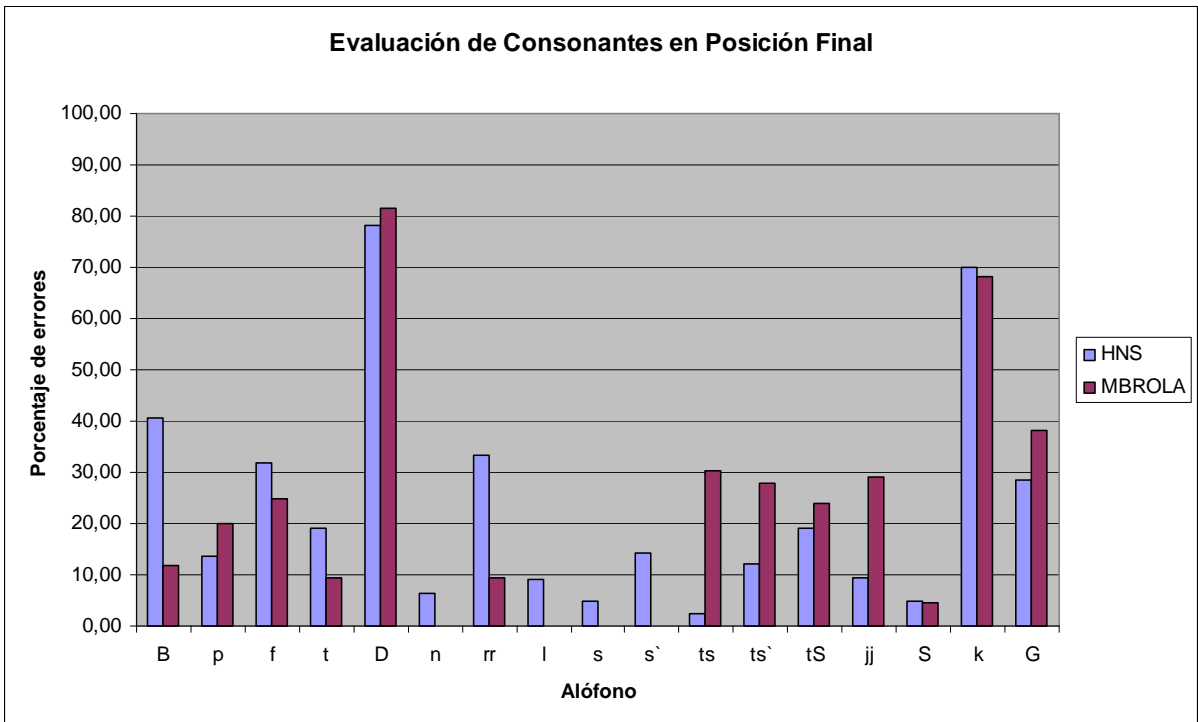


Figura 5: Comparación entre HNS y MBROLA para la prueba de inteligibilidad de consonantes en posición final.

Para la prueba de inteligibilidad de consonantes en posición inicial, se comprueba que el porcentaje de error total es peor para el sistema HNS (78'58% de aciertos) que para el sistema MBROLA (83'33% de aciertos). En la Figura 4 se ve cuál es el comportamiento de cada alófono en cada uno de los dos modelos. Destaca el alto nivel de error de los sonidos [p], [t], [s'] y [tS] para el modelo HNS, y el de [T] para el m

En la prueba de inteligibilidad de consonantes en posición final, se comprueba que el porcentaje de error total es mejor para el sistema HNS (76'41% de aciertos) que para el sistema MBROLA (74'22% de aciertos). La Figura 5 muestra la comparación entre ambos sistemas, mostrando el error cometido en cada alófono para ambos. Destacar que los dos sonidos que peor se reconocen, [D] y [k], se comportan de manera similar en ambos sistemas.

3.2. Prueba de inteligibilidad de grupos consonánticos

El proceso de datos para la obtención de las matrices de confusión de esta prueba consta, al igual que en el caso de la prueba de evaluación de inteligibilidad de consonantes, de dos partes:

- La salida de la base de datos nos da el número de respuesta que el evaluador ha elegido. Es necesario contar esos números, para saber cuántas respuestas ha habido correspondientes a cada opción.

- Una vez contadas, se elabora la matriz de confusión sumando los datos de respuestas dadas para cada una de las opciones.

A continuación se exponen las matrices de confusión de la prueba de inteligibilidad de grupos consonánticos, tanto en MBROLA como en HNS.

3.2.1. MBROLA

	br	bl	dr	fr	fl	gr	gl	kr	kl	pr	pl	tr	tl
br	19	0		0	0	1		0				1	
bl		11		0	0								
dr	5		14			1		5		1		0	
fr	0			18				0		2		1	
fl		0			10				0		0		
gr	3			0		27	0	1	0	0		0	
gl		6				0	14	0	1		0		
kr	0		2	0		7	0	23	0	0		1	
kl		0			1	0	1	0	20		0		
pr	0		1			0		0		13		6	
pl	0	6								0	5		
tr	0		3	0		4		0		0		26	
tl		0							0		2		9

Tabla 13: Matriz de confusión de la prueba de inteligibilidad de grupos consonánticos, con estímulos MBROLA. Valores absolutos.

	br	bl	dr	fr	fl	gr	gl	kr	kl	pr	pl	tr	tl
br	90,48	0,00	0,00	0,00	0,00	4,76	0,00	0,00	0,00	0,00	0,00	4,76	0,00
bl	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
dr	19,23	0,00	53,85	0,00	0,00	3,85	0,00	19,23	0,00	3,85	0,00	0,00	0,00
fr	0,00	0,00	0,00	85,71	0,00	0,00	0,00	0,00	0,00	9,52	0,00	4,76	0,00
fl	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
gr	9,68	0,00	0,00	0,00	0,00	87,10	0,00	3,23	0,00	0,00	0,00	0,00	0,00
gl	0,00	28,57	0,00	0,00	0,00	0,00	66,67	0,00	4,76	0,00	0,00	0,00	0,00
kr	0,00	0,00	6,06	0,00	0,00	21,21	0,00	69,70	0,00	0,00	0,00	3,03	0,00
kl	0,00	0,00	0,00	0,00	4,55	0,00	4,55	0,00	90,91	0,00	0,00	0,00	0,00
pr	0,00	0,00	5,00	0,00	0,00	0,00	0,00	0,00	0,00	65,00	0,00	30,00	0,00
pl	0,00	54,55	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	45,45	0,00	0,00
tr	0,00	0,00	9,09	0,00	0,00	12,12	0,00	0,00	0,00	0,00	0,00	78,79	0,00
tl	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	18,18	0,00	81,82

Tabla 14: Matriz de confusión de la prueba de inteligibilidad de grupos consonánticos, con estímulos MBROLA. Valores relativos.

Se observa que de los 13 grupos estudiados, solamente 2, bl y fl, han sido correctamente identificados en todas sus apariciones. 7 de ellos superan el 80% de éxito en la identificación, y el grupo que más cuesta identificar es el pl, que solo acumula un 45% de identificaciones correctas, mientras que en el resto de las ocasiones ha sido identificado como bl. El porcentaje de acierto de esta prueba ha resultado ser del 77'12%, y el nivel de azar del 20'57%.

3.2.2. HNS

En este caso, de los 13 grupos estudiados, hay 3 (bl, fr, y fl) que han sido correctamente identificados en todas sus apariciones. Y solamente 2 más (gr y gl) que han sido correctamente identificados en más de un 80% de las ocasiones. Llama la atención especialmente el estímulo pl, que siempre que ha aparecido ha sido confundido con bl, si bien los porcentajes de error también son muy altos para tr, que se ha confundido con dr y gr, y para tl, que se confunde con dl y kl. Como se ve, la confusión entre las plosivas en posición inicial es grande.

El índice de acierto para esta prueba llega al 69'02%, con un nivel de azar del 15'47%.

	br	bl	dr	fr	fl	gr	gl	kr	kl	pr	pl	tr	tl
br	15	0		2	0	4		0				0	
bl		10		0	0								
dr	1		16			3		1		1		0	
fr	0			21				0		0		0	
fl		0			11				0		0		
gr	1			0		31	0	0	0	0		0	
gl		2				0	17	0	0		2		
kr	1		1	1		5	0	22	0	0		0	
kl		0			4	0	2	0	18		1		
pr	1		0			0		0		11		10	
pl	0	10								0	0		
tr	2		8	1		9		0		2		8	
tl		3							3		1		3

Tabla 15: Matriz de confusión de la prueba de inteligibilidad de grupos consonánticos, con estímulos HNS. Valores absolutos.

	br	bl	dr	fr	fl	gr	gl	kr	kl	pr	pl	tr	tl
br	71,43	0,00	0,00	9,52	0,00	19,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00
bl	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
dr	4,55	0,00	72,73	0,00	0,00	13,64	0,00	4,55	0,00	4,55	0,00	0,00	0,00
fr	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
fl	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
gr	3,13	0,00	0,00	0,00	0,00	96,88	0,00	0,00	0,00	0,00	0,00	0,00	0,00
gl	0,00	9,52	0,00	0,00	0,00	0,00	80,95	0,00	0,00	0,00	9,52	0,00	0,00
kr	3,33	0,00	3,33	3,33	0,00	16,67	0,00	73,33	0,00	0,00	0,00	0,00	0,00
kl	0,00	0,00	0,00	0,00	16,00	0,00	8,00	0,00	72,00	0,00	4,00	0,00	0,00
pr	4,55	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	50,00	0,00	45,45	0,00
pl	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
tr	6,67	0,00	26,67	3,33	0,00	30,00	0,00	0,00	0,00	6,67	0,00	26,67	0,00
tl	0,00	30,00	0,00	0,00	0,00	0,00	0,00	0,00	30,00	0,00	10,00	0,00	30,00

Tabla 16: Matriz de confusión de la prueba de inteligibilidad de grupos consonánticos, con estímulos HNS. Valores relativos.

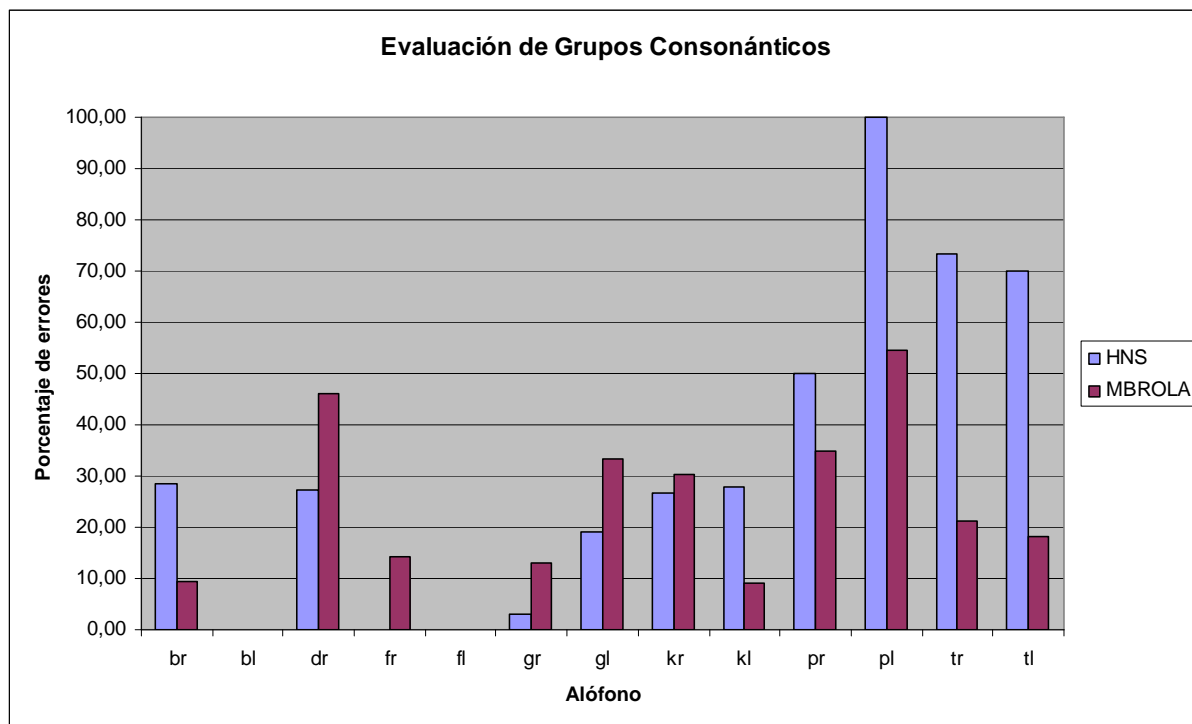


Figura 6: Comparación entre HNS y MBROLA para la prueba de inteligibilidad de grupos consonánticos.

3.2.3. Comparando ambos modelos

Con las pruebas realizadas, el porcentaje de estímulos correctamente identificados es del 77'12% para MBROLA, dato que baja hasta el 69'02% para los estímulos generados con el modelo HNS.

En la Figura 6 podemos observar el comportamiento de ambos modelos para los distintos grupos consonánticos, viendo claramente como el modelo HNS es mejor que MBROLA para dr, fr, gr y gl, pero tiene un comportamiento mucho peor en aquellas combinaciones que comienzan por p y t.

3.3. Prueba de inteligibilidad de combinaciones de vocales

A continuación se exponen las matrices de confusión de la prueba de inteligibilidad de combinaciones de vocales, tanto en MBROLA como en HNS. Para obtenerla se ha procedido, al igual que en los casos anteriores, en dos pasos análogos a los ya comentados.

3.3.1. MBROLA

	ae	ai	ao	au	ei	eo	eu	ia	ii	oi	ui
ae	18	4	0	0			0				
ai	0	64	0	0	0		0			0	0
ao	0	0	9	1							
au	0	0	0	30	0	0	0				
ei	0	0		0	41	0	0			0	0
eo					0	10	0				
eu		0		0	0	0	21				
ia								11			
ii		0						0	10		
oi		0		0	0					19	1
ui		0			0					0	11

Tabla 17: Matriz de confusión de la prueba de inteligibilidad de combinaciones de vocales con estímulos MBROLA. Valores absolutos.

	ae	ai	ao	au	ei	eo	eu	ia	ii	oi	ui
ae	81,82	18,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ai	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ao	0,00	0,00	90,00	10,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
au	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ei	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
eo	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
eu	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
ia	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
ii	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00
oi	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	95,00	5,00
ui	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00

Tabla 18: Matriz de confusión de la prueba de inteligibilidad de combinaciones de vocales con estímulos MBROLA. Valores relativos.

3.3.2. HNS

	ae	ai	ao	au	ei	eo	eu	ia	ii	oi	ui
ae	17	3	0	0			0				
ai	0	62	0	0	0		0			0	0
ao	0	0	6	5							
au	0	0	0	33	0	0	0				
ei	0	0		0	43	0	0			0	0
eo					0	10	1				
eu		0		0	0	0	21				
ia								10			
ii		0						0	11		
oi		0		0	0					22	0
ui		0			0					1	9

Tabla 19: Matriz de confusión de la prueba de inteligibilidad de combinaciones de vocales con estímulos HNS. Valores absolutos.

Según se aprecia en la tabla 18, la mayoría de las combinaciones de vocales han sido correctamente identificadas, apareciendo errores solamente en ae (este es el más relevante, siempre que ha sido incorrectamente identificada se ha confundido con ai), ao y oi. El porcentaje de estímulos correctamente identificados ha sido del 97'60%, y el nivel de azar ha resultado ser del 12'77%

	ae	ai	ao	au	ei	eo	eu	ia	ii	oi	ui
ae	85,00	15,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ai	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ao	0,00	0,00	54,55	45,45	0,00	0,00	0,00	0,00	0,00	0,00	0,00
au	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ei	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
eo	0,00	0,00	0,00	0,00	0,00	90,91	9,09	0,00	0,00	0,00	0,00
eu	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
ia	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
ii	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00
oi	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
ui	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	10,00	90,00

Tabla 20: Matriz de confusión de la prueba de inteligibilidad de combinaciones de vocales con estímulos HNS. Valores relativos.

En este caso, 7 de las 11 combinaciones evaluadas se reconocen correctamente siempre, y solamente una (ao) tiene un índice de reconocimiento menor del 80%. Este error viene dado por el hecho de que en un 45% de las ocasiones se ha confundido con la opción au. La tasa de reconocimiento global de esta prueba es del 96'06%, y el nivel de azar ha resultado ser del 12%.

3.3.3. Comparando ambos modelos

En la Figura 7 comprobamos que los errores en la prueba de inteligibilidad de combinaciones de vocales son pequeños: menores del 20% en ambos modelos, con excepción de la combinación ao para HNS, que se confunde frecuentemente con au.

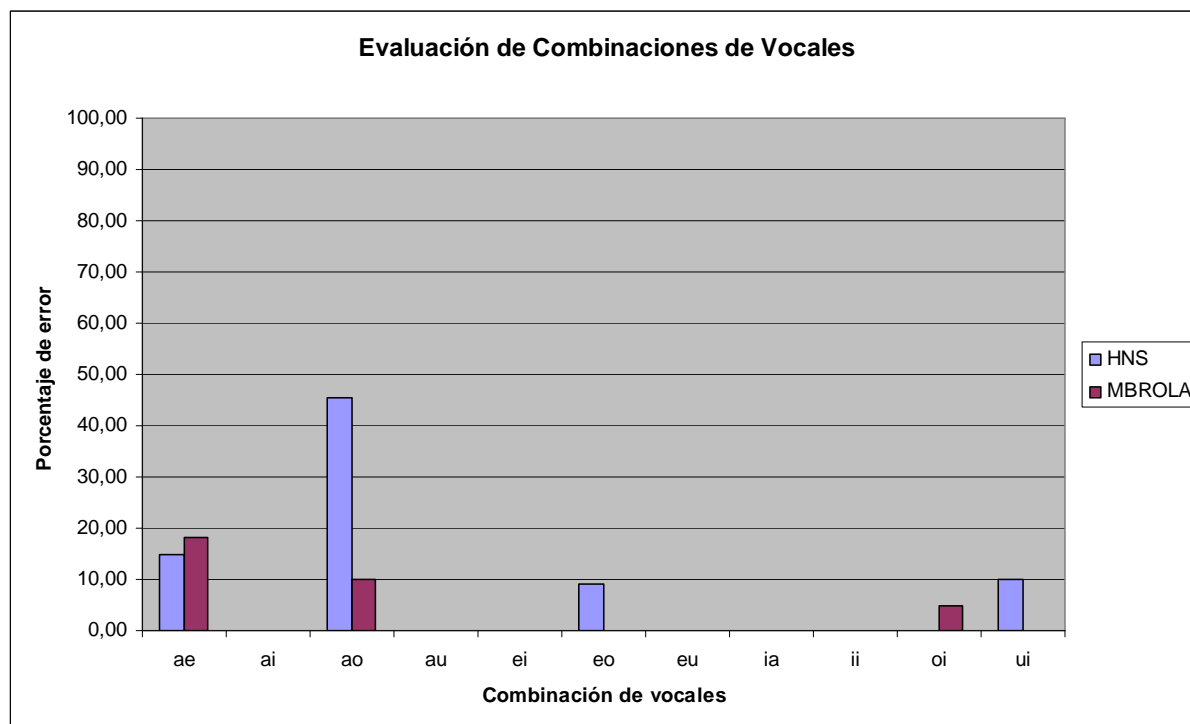


Figura 7: Comparación entre HNS y MBROLA para la prueba de inteligibilidad de combinaciones de vocales.

3.4. Prueba de inteligibilidad de palabras en frases con sentido

Para obtener los resultados correspondientes a esta prueba se ha desarrollado un sistema que automáticamente es capaz de contar cuántas veces se ha identificado cada palabra de manera correcta. Para ello, y utilizando ficheros intermedios de texto, se ubican en una hoja de cálculo de Excel tanto los textos de los estímulos como los textos que los evaluadores han introducido, ubicando una sola palabra en cada celda. Contando cuántas veces aparece cada palabra del texto original en las palabras que los evaluadores han tecleado, tendremos una idea de las palabras entendidas correcta e incorrectamente.

Este método es muy restrictivo, dado que:

a) Se tienen en cuenta errores en todas las palabras, a pesar de que para la comprensión de frases muchas veces solamente es necesario tener en cuenta algunas palabras claves, y el sentido se mantiene aunque alguna partícula pueda perderse.

b) Se toman como errores palabras que simplemente pueden ser interpretaciones diferentes de una misma realización, y que desde el punto de vista de la comprensión se podrían justificar por el contexto y darse por correctas (por ejemplo: si en el texto que ha generado el estímulo se puede leer *etorri z gero*, y alguno de los evaluadores lo transcribiera como *etorri ezker*, se considera un error, a pesar de que el contexto justificaría la diferente transcripción).

c) Algunos errores que nada tienen que ver con la comprensión del texto, tales como fallos ortográficos o tipográficos, también computan como palabras mal comprendidas.

Los resultados obtenidos son los siguientes:

Modelo utilizado	HNS	MBROLA
Palabras correctas	6139	6178
Palabras totales	6512	6508
Tasa de acierto	94'27%	94'92%

Tabla 21: Resultados de las pruebas de inteligibilidad de palabras en frases con sentido.

Según se observa en la tabla, en frases con sentido ambos sistemas consiguen una tasa de inteligibilidad de entorno al 94%.

3.5. Prueba de inteligibilidad de palabras en frases sin sentido

Teniendo en cuenta para esta prueba las mismas consideraciones que para la anterior, los resultados obtenidos son los siguientes:

Modelo utilizado	HNS	MBROLA
Palabras correctas	2396	2463
Palabras totales	2609	2662
Tasa de acierto	91'83%	92'52%

Tabla 22: Resultados de las pruebas de inteligibilidad de palabras en frases sin sentido.

A la vista de los datos de la tabla, las tasas de acierto son algo menores que en el caso anterior, pero de nuevo ambos modelos obtienen resultados muy parecidos (en torno al 92%).

3.6. Prueba de comprensión de textos

Los resultados obtenidos en la prueba de comprensión de textos son los siguientes:

Modelo utilizado	HNS	MBROLA
Respuestas correctas	122	116
Respuestas totales	175	145
Tasa de acierto	78'70%	80'00%

Tabla 23: Resultados de las pruebas de comprensión de textos.

En este caso se observa que de nuevo ambos modelos dan tasas de acierto muy parecidas, del orden del 80%. Se observa que el modelo MBROLA responde algo mejor.

En la Figura 8 se detallan los resultados por texto.

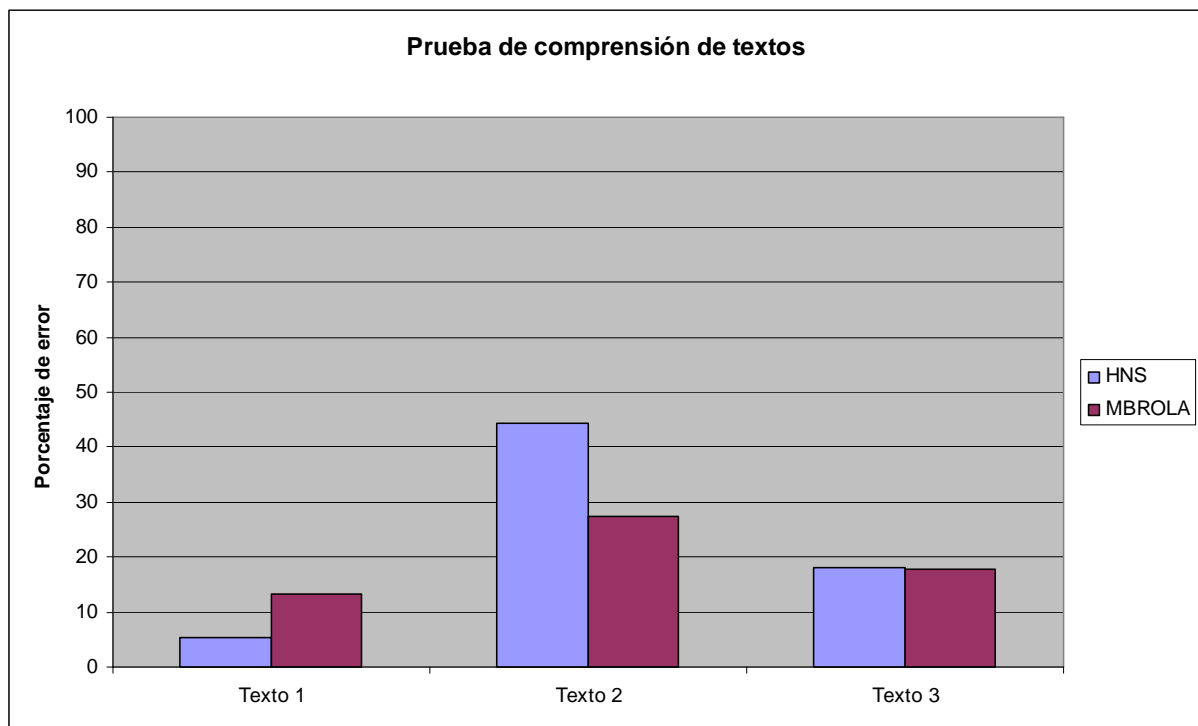


Figura 8: Comparación entre HNS y MBROLA para la prueba de comprensión de textos.

Según se ve, para ambos modelos se han obtenido peores resultados en el texto 2, lo que hace pensar que la dificultad intrínseca para comprender el texto era mayor en este caso. Para comprobarlo, acudimos a la prueba que se realizó en (Hernández, 1999), en la que se incluía una prueba de comprensión de estos mismos textos, en la cual los estímulos habían sido generados con una voz natural, de manera que utilizando como referencia los resultados obtenidos con esa voz natural se puede evaluar la dificultad intrínseca de comprensión de los textos.

Si tenemos en cuenta la voz natural como referencia, vemos que el modelo HNS ha respondido mejor que el MBROLA en el primer texto, peor en el segundo, y sólo ligeramente peor, pero muy parecido en el tercero. En cualquier caso, el modelo HNS se ha mantenido siempre entre el 71 y el 102 por ciento, consiguiendo resultados más homogéneos que el MBROLA, cuyos resultados oscilan entre el 88 y el 133%.

	Texto 1	Texto 2	Texto 3
Índice de aciertos con estímulos de voz natural	97'73%	54'81%	85'56%
Índice de acierto HNS absoluto	94'54%	55'56%	81'82%
Índice de acierto HNS relativo a la voz natural	96'73%	101'37% ¹¹	95'63%
Índice de acierto MBROLA absoluto	86'67%	72'73%	82'22%
Índice de acierto MBROLA relativo a la voz natural	88'68%	132'69% ¹¹	96'10%

Tabla 24: Resultados de la prueba de comprensión de textos con referencia de la prueba con estímulos de voz natural.

¹¹ El porcentaje de acierto mayor del 100% implica que se han obtenido mejores resultados con voz sintética que con la voz natural.

3.7. Prueba de Calidad Global

Esta prueba consiste en dos partes. En la primera, los evaluadores dan su opinión ante diferentes cuestiones referidas a un estímulo de texto, que en algunos casos será MBROLA y en otros HNS, evaluándolas del 1 al 5. Los resultados de esta primera parte se ilustran en la Figura 9.

En el formulario aparecen algunas opciones que se refieren a una valoración positiva contrapuesta a una negativa (como por ejemplo desagradable –agradable) y otras que simplemente evalúa una faceta de la voz entre dos contrapuestas (grave – aguda).

En las primeras cuatro cuestiones, que se refieren a la calidad suprasegmental, el modelo HNS recibe una mejor valoración que el MBROLA, dado que en las preguntas ‘ayuda a comprender’ y ‘hace bien las pausas’ obtiene una valoración de 3 puntos, es decir, el valor intermedio entre el sí y el no, mientras que el modelo MBROLA, a juicio de los evaluadores, ayuda menos a comprender, y hace algo peor las pausas. En las dos siguientes preguntas, la valoración de los evaluadores es negativa para ambos modelos: no acentúa bien (en esto el modelo MBROLA obtiene una puntuación levemente mejor, pero es el peor aspecto en ambos) y no pronuncia bien (también con una leve ventaja para el modelo MBROLA en este aspecto).

Las siguientes siete cuestiones se refieren al estilo de lectura. Entre ellas, las cuatro primeras, que no son valorativas sino meramente cualitativas, han presentado diferencias pequeñas entre ambos modelos, y los evaluadores han definido la voz como más lenta que rápida, más formal que informal y más culta que coloquial. En las tres restantes observamos que el habla generada con el modelo MBROLA resulta para los usuarios ligeramente fácil de entender, mientras que el habla generada con el modelo HNS resulta más difícil de comprender, mientras que ambos modelos resultan artificiales y aburridos.

Las últimas valoraciones se refieren a la calidad de la voz. En las valoraciones que siguen, las más negativas hablan de una voz metálica y monótona, en ambos modelos. Las mayores diferencias entre los modelos son las siguientes:

La voz sintetizada mediante el modelo HNS resulta más sugestiva y agradable al oído de los evaluadores que la generada con el modelo HNS.

A pesar de que los estímulos de ambos modelos han sido generados utilizando los mismos valores para la entonación, incluyendo la misma frecuencia fundamental, la voz generada con el modelo MBROLA aparece a los evaluadores como más aguda que la del modelo HNS.

La segunda parte de esta prueba de calidad global corresponde a las opiniones generales de los evaluadores sobre los sistemas de síntesis, no

refiriéndose a ningún modelo de síntesis en particular. Los resultados se ven en la Figura 10.

Como se observa, el sistema de evaluación parece inaceptable y poco adecuado a los evaluadores, si bien lo han encontrado levemente satisfactorio y eficaz.

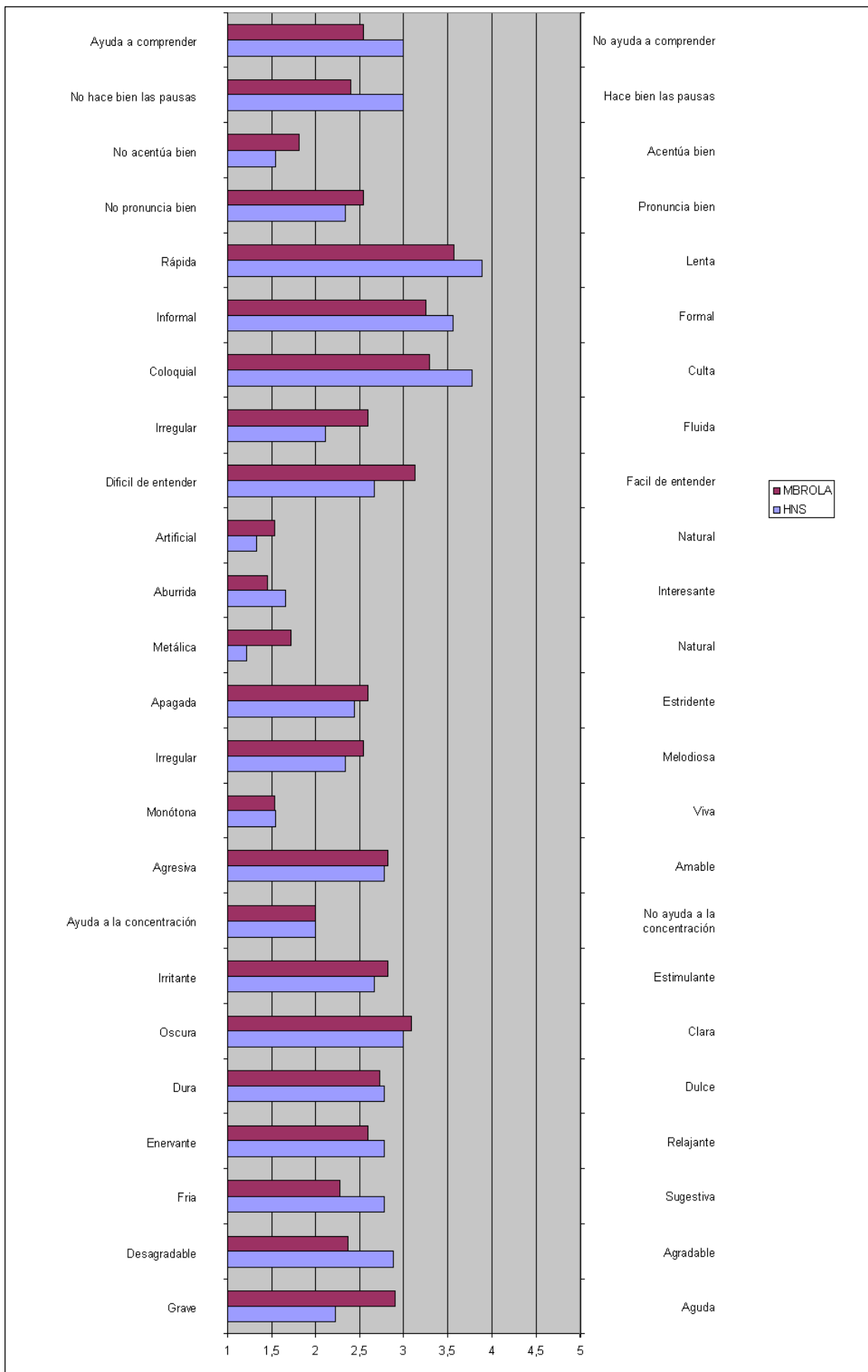
Por otro lado, la síntesis de habla parece más adecuada para sistemas telefónicos de información, sea información general o noticias en particular, que para la lectura de textos o para enseñar euskera a extranjeros. Como conclusión final, la tendencia o no a usar un sistema de síntesis de voz, es positiva.

Como último aspecto dentro de la prueba de calidad global, se recogen los comentarios de los evaluadores. Algunos comentan aspectos de la evaluación: “Ebaluazioa luzeegia iruditu zait” (la evaluación me ha parecido demasiado larga), pero los comentarios más interesantes se refieren a la propia voz sintética. Se reproducen a continuación los más representativos, traducidos a castellano:

- La voz me ha parecido un poco monótona, por lo demás es bastante comprensible. Cuando fuera necesaria agilidad no la utilizaría, pero por lo demás, no es un mal sistema.
- El sistema puede ser adecuado para dar información corta. Para estudiar el idioma o dar información larga no me parece adecuado.
- Los acentos siguen el modelo de castellano, parece que muchas palabras se acentúan como si fueran llanas, en lugar de colocar el acento en la segunda sílaba.
- La voz es tal vez demasiado metálica y monótona
- Desde el punto de vista de la entonación, no me parece bien: artificial, en mi opinión.
- Se nota bastante que hay dos tipos de voz, y que una de ellas se ha trabajado mucho más que la otra, porque se oye, entiende, y sigue mejor.
- En lo que corresponde a la voz, me parece agradable.

3.8. Comparaciones y otras pruebas

Los datos que aquí se han compilado no pasan de ser meros números si no hay alguna referencia para compararlos. Es por ello necesario comparar los resultados obtenidos, con otros que puedan servir de referencia. Una prueba que puede hacerse es la comparación de los resultados con la voz real, que queda como trabajo futuro. Las comparaciones que aquí se plantean son, por un lado con las pruebas realizadas por evaluadores acostumbrados a la voz sintética, y por otro, con los resultados obtenidos por otros autores.



1. Figura 9: Comparación entre HNS y MBROLA para la prueba de calidad global.

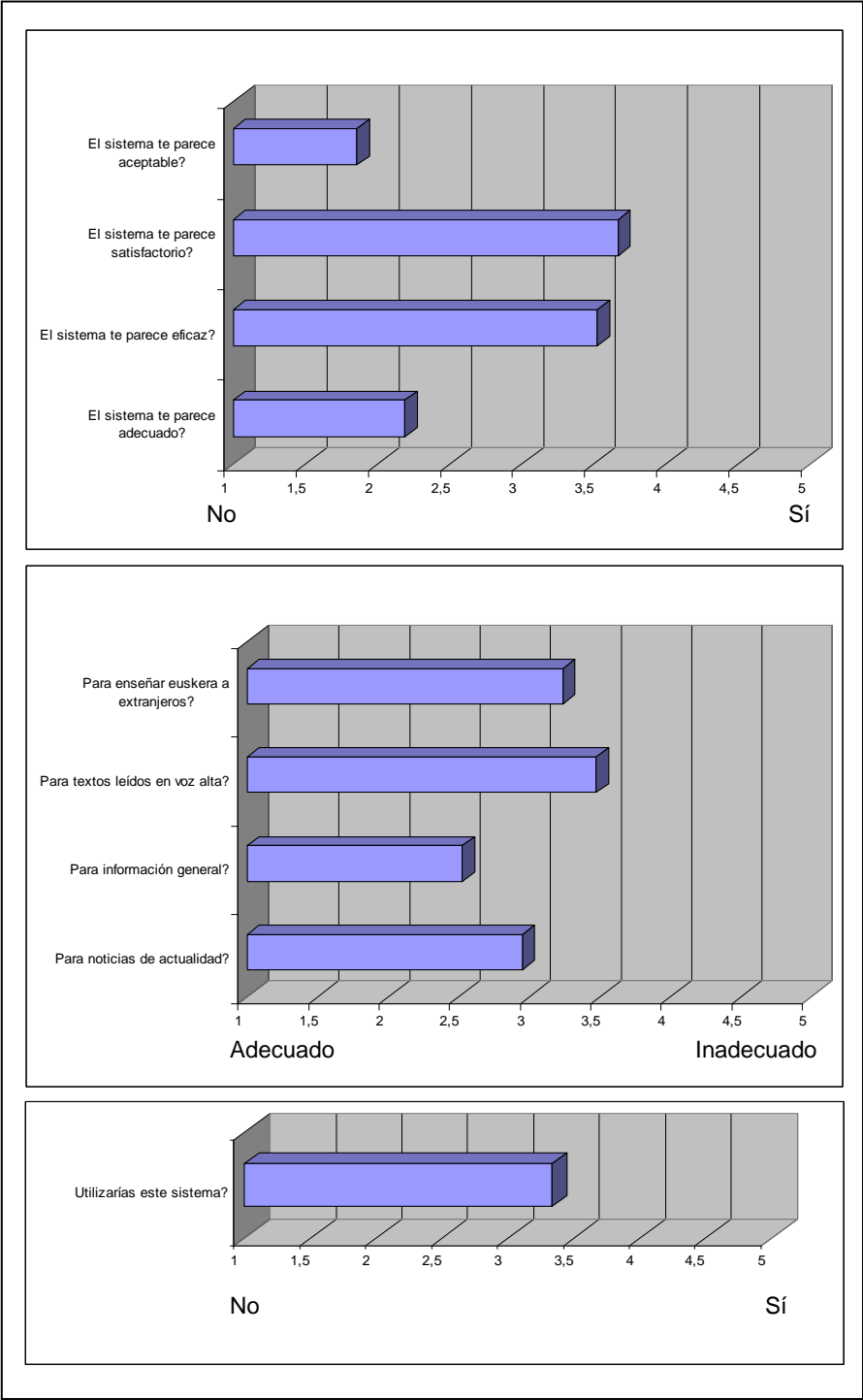


Figura 10: Resultados de la prueba de calidad global independiente del modelo.

3.9. Evaluando el efecto aprendizaje

De cara a comprobar la robustez del sistema de evaluación se realizó un primer juego de encuestas utilizando como evaluadores a sujetos que sí tenían experiencia previa en la utilización de habla sintética. Estas evaluaciones pueden utilizarse para evaluar la influencia de esa preparación previa sobre el resultado de las pruebas de inteligibilidad.

Se trata de una prueba de evaluación más restringida y por tanto menos representativa que la general (solamente la han llevado a cabo 6 evaluadores, todos ellos con experiencia previa en voz sintética), y que por tanto usaremos solamente a efectos de comparación entre los resultados generales.

En las siguientes tablas se exponen los índices de acierto de los distintos sistemas para todas las pruebas de evaluación realizadas.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
89,59%	92,25%	91,63%	92,50%

Tabla 25: Comparación entre los resultados de las pruebas de evaluación de consonantes en posición intermedia de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
83,33%	84,76%	78,58%	76,85%

Tabla 26: Comparación entre los resultados de las pruebas de evaluación de consonantes en posición inicial de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
74,22%	72,00%	76,41%	86,81%

Tabla 27: Comparación entre los resultados de las pruebas de evaluación de consonantes en posición final de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
77,12%	86,75%	69,06%	81,69%

Tabla 28: Comparación entre los resultados de las pruebas de evaluación de grupos consonánticos de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
97,60%	100,00%	96,06%	97,37%

Tabla 29: Comparación entre los resultados de las pruebas de evaluación de combinaciones de vocales de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
94,93%	93,87%	94,27%	94,27%

Tabla 30: Comparación entre los resultados de las pruebas de comprensión de palabras en frases con sentido de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
92,52%	92,83%	91,84%	93,03%

Tabla 31: Comparación entre los resultados de las pruebas de comprensión de palabras en frases sin sentido de evaluadores con experiencia en habla sintética y sin ella.

MBROLA Sin aprendizaje	MBROLA Con aprendizaje	HNS Sin aprendizaje	HNS Con aprendizaje
80'00%	90'00%	78,71%	90'00%

Tabla 32: Comparación entre los resultados de las pruebas de comprensión de textos de evaluadores con experiencia en habla sintética y sin ella.

Tal y como se observa en las tablas anteriores, en prácticamente todos los experimentos se han obtenido mejores resultados para el caso de evaluadores con experiencia previa en voz sintetizada. Aunque en algunos casos no es especialmente relevante, en otros (por ejemplo, evaluación de grupos consonánticos o comprensión de textos) las diferencias pueden llegar a cambios en los niveles de aciertos de más del 10%.

Por tanto los resultados de las evaluaciones realizadas confirman la necesidad de realizar las evaluaciones de sintetizadores de voz utilizando evaluadores que no hayan tenido experiencia previa en voz sintética.

3.9.1. Comparación con los resultados de otros sintetizadores.

Otros autores han utilizado pruebas equivalentes a las que aquí se han comentado para evaluar la calidad de la voz de sus sintetizadores.

Evaluación del euskera

La referencia más parecida de que disponemos para poder comparar estos resultados son los obtenidos en la prueba en la que se basa ésta, recogidos en (Hernández, 1999). La comparación entre los resultados obtenidos entonces y los recopilados en la nueva prueba de evaluación se resumen en las siguientes tablas, en las cuales se dan los índices de error del sistema, en tanto por uno, para cada una de las pruebas:

Alófono	HNS	MBROLA	Voz Natural	Prueba Original
m	0,00	0,98	0,00	0,00
p	0,02	0,02	0,01	0,07
B	0,02	0,11	0,00	0,21
f	0,12	0,12	0,09	0,02
T	0,31	0,59	0,33	0,24
t	0,16	0,16	0,00	0,00
D	0'09	0,24	0,01	0,00
n	0,14	0,14	0,01	0,02
rr	0'10	0'02	0,00	0,00
l	0'10	0'02	0,00	0,07
s	0,02	0,07	0,02	0,16
s'	0,00	0,09	0,01	0,00
J	0,00	0,00	0,00	0,05
tS	0,02	0,00	0,00	0,00
jj	0,00	0,00	0,00	0,05
S	0,02	0,02	0,01	0,00
c	0,07	0,05	0,00	0,00
k	0,012	0,07	0,01	0,02
G	0,14	0,10	0,00	0,00
L	0,00	0,35	0,00	0,00
r	0,19	0,05	0,00	0,01
ts	0,05	0,10	0,00	0,00
ts'	0,00	0,05	0,00	0,00
x	0,23	0,15	0,03	0,30
Media	0,08	0,11	0,02	0,05

Tabla 33: Comparación entre los resultados actuales y los de (Hernández, 1999) en la prueba de evaluación de consonantes en posición intermedia

	HNS	MBROLA	Voz Natural	Prueba Original
m	0,00	0,02	0,01	0,01
p	0,52	0,14	0,24	0,50
b	0,08	0,06	0,01	0,03
f	0,24	0,00	0,26	0,05
T	0,37	0,58	0,25	0,33
t	0,40	0,00	0,00	0,00
d	0,29	0,34	0,00	0,10
n	0,09	0,20	0,00	0,00
rr	0,14	0,19	0,00	0,00
l	0,05	0,00	0,00	0,00
s	0,09	0,05	0,00	0,02
s`	0,53	0,03	0,06	0,00
J	0,07	0,02	0,00	0,02
tS	0,90	0,23	0,00	0,02
S	0,29	0,19	0,02	0,12
c	0,14	0,24	0,10	0,02
gj	0,13	0,12	0,07	0,04
k	0,44	0,26	0,01	0,20
g	0,14	0,19	0,00	0,02
L	0,09	0,19	0,00	0,00
Media	0,25	0,15	0,5	0,07

Tabla 34: Comparación entre los resultados actuales y los de (Hernández, 1999) en la prueba de evaluación de consonantes en posición inicial

	HNS	MBROLA	Voz Natural	Prueba Original
B	0,40	0,12	0,03	0,25
p	0,14	0,20	0,02	0,51
f	0,32	0,25	0,00	0,02
t	0,19	0,9	0,00	0,00
D	0,78	0,81	0,06	0,29
n	0,06	0,00	0,02	0,03
rr	0,33	0,09	0,00	0,00
l	0,09	0,00	0,00	0,00
s	0,05	0,00	0,00	0,00
s`	0,14	0,00	0,00	0,14
ts	0,02	0,30	0,00	0,18
ts`	0,12	0,28	0,00	0,93
tS	0,19	0,24	0,00	0,00
S	0,05	0,05	0,00	0,02
k	0,70	0,68	0,00	0,02
G	0,28	0,38	0,09	0,44
Media	0,24	0,22	0,1	0,18

Tabla 35: Comparación entre los resultados actuales y los de (Hernández, 1999) en la prueba de evaluación de consonantes en posición final

Como se aprecia en las tres tablas anteriores, los resultados obtenidos en la prueba original fueron mejores que los actuales en las pruebas de identificación de consonantes. Especialmente los resultados obtenidos con voz natural son mejores, como cabía esperar.

	HNS	MBROLA	Voz Natural	Prueba Original
br	0,29	0,09	0,00	0,42
bl	0,00	0,00	0,00	0,00
dr	0,27	0,46	0,12	0,12
fr	0,00	0,14	0,00	0,52
fl	0,00	0,00	0,00	0,00
gr	0,03	0,13	0,00	0,22
gl	0,19	0,33	0,00	0,22
kr	0,27	0,30	0,00	0,18
kl	0,28	0,09	0,05	0,07
pr	0,50	0,35	0,02	0,04
pl	1,00	0,55	0,00	0,05
tr	0,73	0,21	0,00	0,08
tl	0,70	0,18	0,10	0,04
Media	0,33	0,22	0,02	0,18

Tabla 36: Comparación entre los resultados actuales y los de (Hernández, 1999) en la prueba de evaluación de grupos consonánticos

También en la prueba de evaluación de grupos consonánticos original se han obtenido mejores resultados que en las actuales, tanto con el modelo MBROLA como con el modelo HNS.

	HNS	MBROLA	Voz Natural	Prueba Original
ae	0,15	0,18	0,00	0,00
ai	0,00	0,00	0,00	0,00
ao	0,45	0,10	0,00	0,05
au	0,00	0,00	0,00	0,00
ei	0,00	0,00	0,00	0,00
eo	0,09	0,00	0,00	0,05
eu	0,00	0,00	0,00	0,00
ia	0,00	0,00	0,00	0,11
ii	0,00	0,00	0,00	0,05
oi	0,00	0,05	0,00	0,00
ui	0,10	0,00	0,00	0,00
Media	0,07	0,03	0,00	0,02

Tabla 37: Comparación entre los resultados actuales y los de (Hernández, 1999) en la prueba de evaluación de combinaciones de vocales

Se observa que la prueba de evaluación de combinaciones de vocales obtiene un resultado perfecto para el caso de voz natural, mientras que en el caso de voz sintética el resultado es similar al obtenido en la prueba actual de MBROLA, resultando la de HNS con más equivocaciones.

	HNS	MBROLA	Voz Natural	Prueba Original
Palabras en frases con sentido	0,06	0,05	0,01	0,04
Palabras en frases sin sentido	0,08	0,07	0,02	0,08

Tabla 38: Comparación entre los resultados actuales y los de (Hernández, 1999) en las pruebas de evaluación de palabras en frases.

Los resultados de la prueba de evaluación de palabras en frases son similares en ambos sistemas, siendo siempre mejores en el caso de la voz natural.

Como resumen de esta comparación, se puede concluir que los resultados obtenidos en la prueba actual han sido, como cabía esperar, peores que los correspondientes a la voz natural. También han resultado, en general, peores que los de la voz sintética utilizada en la prueba original.

Evaluación en español

La prueba realizada en (Aguilar et al., 1994) para el idioma español es muy similar a la que en este documento se ha descrito, lo cual permite comparar los resultados de manera precisa, tal como se ve en la Tabla 39, en la cual se representan porcentajes de acierto para las diferentes pruebas.

	Euskera	Euskera	Español	Español
Consonantes en posición intermedia	91'63%	89'59%	86'45%	84'95%
Consonantes en posición inicial	78'58%	83'33%	93'2%	90'6%
Consonantes en posición final	76'41%	74'22%	79'7%	79'3%
Grupos consonánticos	69'06%	77'12%	84%	59'66%
Combinaciones de vocales	97'60%	96'06%	79'4%	79'28%
Palabras en frases con sentido	94'27%	94'93%	96'65%	94'69%
Palabras en frases sin sentido	91'83%	92'52%	87'3%	84'2%
Comprensión de textos	78'70%	80'00%	73'88%	72'22%

Tabla 39: Comparación con la evaluación del sintetizador en español de Telefónica: tasa de acierto de ambos sistemas.

En esta comparación vemos como, dentro de las voces sintéticas del conversor de texto en habla de telefónica en español, la masculina tiene mayor tasa de aciertos que la femenina. Dado que en el caso de la evaluación objeto de este trabajo las voces generadas con ambos modelos son masculinas, la comparación de resultados es más interesante será con la masculina de español. Ésta se comporta mejor que las utilizadas para los modelos HNS y MBROLA de euskera en las pruebas de evaluación de consonantes en posición inicial y grupos consonánticos, mientras que sucede al contrario en todas las demás pruebas, exceptuando la evaluación de palabras en frases con sentido, en que los resultados obtenidos son muy similares.

Otros trabajos

Por último, comentar trabajos que otros autores han realizado con pruebas de inteligibilidad de elementos segmentales basadas en DRT y MRT, para una voz sintetizada en idioma árabe egipcio (Mayfield Tomokiyo et al., 2003).

La síntesis se realiza en base a un modelo cepstral LLC de selección de unidades, diseñado para ser ejecutado en una PDA. Las pruebas realizadas son dos, una basada en MRT y DRT, basada, al igual que las que en este trabajo se han realizado para la evaluación de elementos segmentales, en presentación a los evaluadores estímulos monosilábicos, de manera que ellos puedan elegir entre varias opciones, diferentes solo en un sonido. En esta prueba han conseguido tasas de reconocimiento correcto de entre el 72% y el 78%.

Los resultados de evaluación de elementos segmentales de la prueba desarrollada en este trabajo están entre el 97'60% y el 69'06%.

Por otro lado, los mismos autores han realizado también una prueba de evaluación de comprensión de palabras en frases. En este caso las frases provenían de una base de datos desarrollada para integrar el motor de síntesis en un sistema de traducción automática inglés árabe, por tanto todas las frases tenían sentido. Las frases se presentaban a los evaluadores (hablantes nativos de árabe egipcio), quienes sólo debían marcar aquellas palabras que no les sonaran 'bien', lo cual es un enfoque diferente al que se ha dado a la prueba de evaluación de palabras en frases con sentido en este trabajo. Los resultados que han obtenido han sido del 84'7% en la tasa de palabras aceptadas como correctas. Los resultados obtenidos en la prueba de evaluación de palabras en frases con sentido en este trabajo han sido del 94'27% para MBROLA y 94'93% para HNS.

4. Conclusiones

Se ha desarrollado, adaptando diseños ya existentes (Hernández, 1999 y Martín, 2004) un sistema de evaluación que permite realizar múltiples pruebas de cara a conocer la calidad subjetiva de la voz en un sistema de síntesis, incidiendo especialmente en el ámbito de la inteligibilidad.

Se ha utilizado dicho sistema para evaluar los motores HNS y MBROLA del Conversor de Texto en Habla AhoTTS, obteniendo en las distintas pruebas las tasas de acierto que a continuación se resumen:

	Euskera HNS	Euskera MBROLA	Español voz masculina	Español voz femenina
Consonantes en posición intermedia	91'63%	89'59%	86'45%	84'95%
Consonantes en posición inicial	78'58%	83'33%	93'2%	90'6%
Consonantes en posición final	76'41%	74'22%	79'7%	79'3%
Grupos consonánticos	69'06%	77'12%	84%	59'66%
Combinaciones de vocales	97'60%	96'06%	79'4%	79'28%
Palabras en frases con sentido	94'27%	94'93%	96'65%	94'69%
Palabras en frases sin sentido	91'83%	92'52%	87'3%	84'2%
Comprensión de textos	78'70%	80'00%	73'88%	72'22%

Tabla 40: Resumen de los resultados obtenidos.

Los resultados obtenidos son muy similares para los dos modelos, dándose la diferencia más significativa en la evaluación de grupos consonánticos, en la que el modelo MBROLA resulta mejor valorado.

En base a las pruebas realizadas, se han localizado también ciertos puntos susceptibles de mejora en los sistemas de síntesis, particularmente en lo que se refiere a alófonos, en diferentes posiciones, en los cuales se cometen errores comunes en una mayoría de evaluadores:

Para ambos modelos de síntesis, el sonido [T], fricativo sordo interdental, es frecuentemente confundido con el [f], fricativo sordo labiodental, tanto en posición inicial como en posición intermedia. En el modelo HNS en particular, el sonido africado sordo prepalatal [tS] ha sido mayoritariamente evaluado incorrectamente, confundida con otras fricativas alveolares sordas, [s] y [s`].

En posición final, el sonido aproximante sonoro dental [D] ha sido mayoritariamente reconocido como [rr], en ambos modelos. También la plosiva sorda velar [k] necesita ser más trabajada en ambos modelos.

En la prueba de inteligibilidad de grupos consonánticos es de destacar que en el modelo HNS necesitan mejorarse los grupos que comienzan por [p] y [t], dado que han obtenido índices muy bajos de inteligibilidad, llegando incluso a cero en el caso del grupo pl. Éste último grupo, junto con el dr, son los que peores resultados han obtenido también utilizando el modelo MBROLA.

En la prueba de inteligibilidad de combinaciones de vocales, lo más reseñable es la confusión habitual de la combinación ao por au en el modelo HNS.

Para conocer con exactitud la bondad de los resultados hay que tener en cuenta que en la calidad de la voz no se puede medir de manera absoluta: la comunicación de la propia voz humana no está exenta de errores, fallos de dicción... Se han estado utilizando en lo posible los resultados de las pruebas de evaluación con voz real realizados en (Hernández, 1999), si bien esta referencia es necesariamente inexacta por tratarse de distintas épocas, distintos evaluadores, y haber también matices diferentes en la metodología de realización de la prueba (el más importante, que la prueba objeto de este trabajo se realiza a través de Internet). Por tanto, queda como trabajo futuro el diseñar un sistema que realice todas las pruebas de la batería de evaluación utilizando como estímulo voz humana, de manera que se pueda conocer cual es el objetivo a cumplir. La peculiaridad que deberá cumplir ese sistema es que, cuando estamos evaluando únicamente la calidad del sistema de generación de habla, no podemos permitir que las variaciones en la prosodia introduzcan diferencias que quedan fuera de nuestro alcance. Es necesario por tanto que tanto las grabaciones de voz real como estímulos de voz sintetizada utilicen la misma prosodia.

5. AGRADECIMIENTOS

Los autores desean agradecer especialmente a los evaluadores que han dedicado altruistamente su tiempo para que este trabajo haya podido plasmarse en unos resultados. Tiempo que en ocasiones ha sido especialmente largo, o que ha implicado solucionar infinidad de problemas técnicos.

Leire Romero	Alfonso (nkterri)
Igor Fernández	Imanol Madariaga
Jon del Portillo	Alfredo Hurtado de Saratxo
Iratxe Landa	Belen Olaetxea
Ibon Sarasola	Borja Uriarte
Jaime Jiménez	Larraitz Aramburu
Jon Arrue	Iker Apalategi
Juan Luis Ordiales	Esti Amorrortu
Pablo Angueira	Eduardo Zabala
Eneko Arróspide	Eleder Aurtenexea
Xabier Zalbide	Alfonso Gallastegi
Xabier Ibarrondo	Iker Luengo
Leire Romero	Gaizka Durana

Agradecimiento también para Aitziber Villalba, cuya ayuda con la configuración del sistema PHP-MySQL ha sido inestimable.

6. Bibliografía

- Gaminde, I. (1995). *Bizkaieraren azentu-moldeez*. Bilbao: Labayru Ikastegia.
- Mitxelena, K. (1977). Euskal hizkera eta euskal neurkera, *Euskera* 22: 721-733.
- Hualde, J. I.; Mugarza, P.; Zuazo, K. (2002). The Accentual System of Mallabia Basque. *Euskalingua* 1: 15-25.
- Aholaben webgunea (2002). http://bips.bi.ehu.es/ahoweb/eu_index.html
- Aguilar, L.; Fernandez, J.M.; Garrido, J.M.; Llisterri, J. (1994). Diseño de pruebas para la evaluación del habla sintetizada en español y su aplicación a un sistema de conversión de texto a habla, *Actas del X congreso de la SEPLN*. Córdoba.
- Egan, J.P. (1948). Articulation Testing Methods. *Laryngoscope*, Vol. 58, pp. 955-991.
- Fairbanks, G. (1958). Test of Phonemic Differentiation: The Rhyme Test. *Journal of the Acoustical Society of America*, 30,7, pp. 596-600.
- Gilmore, W. J. (2001). *Getting Started with MySQL* http://dev.mysql.com/tech-resources/articles/mysql_intro.html
- Hernández, I., Navas, E., Murugarren, J.L., Etxebarria, B. (2001). Description of the AhoTTS Conversion System for the Basque Language *4th ISCA Tutorial and Research Workshop on Speech Synthesis*
- Hernández, I. (1999). Informe final del Proyecto Telefonica I+D '99
- Martín, I., Navas, E., Hernández, I., Sánchez, J. (2004). Sistemas de Evaluación Subjetiva vía Internet para Aplicaciones de Tecnologías del Habla. *XIX Simposium Nacional de la Unión Científica Internacional de Radio*. Gandía.
- Mayfield Tomokiyo, L., Black, A. W., Lenzo, K. A. (2003). Arabic in my Hand: Small-footprint Synthesis of Egyptian Arabic *8th European Conference on Speech Communication and Technology*
- Navas, E., Hernaez, I., Sanchez, J. (2002). Basque Intonation Modelling for Text to Speech Conversion. *Proceedings of the International Conference on Spoken Language Processing*. Denver, Colorado. USA.
- Nye, P. W., Gaitenby, J. (1974). The Intelligibility of Synthetic Monosyllable Words in Short, Syntactically Normal Sentences. *Haskins Laboratories Status Report on Speech Research SR-37/38*, pp. 169-190.
- Pisoni, D. B. (1987). *From test to Speech. The MITalk System: Some measures of intelligibility and comprehension* (pp. 151-171) Cambridge University Press.
- Voiers, W.D. (1984). Evaluating Processed Speech using the Diagnostic Rhyme Test. *Speech Technology* 1,4 pp. 30-39
- Aholab webgunea: Basque SAMPA (2004) <http://bips.bi.ehu.es/ahoweb/files/sampa.html>
- The phpMyAdmin Project <http://www.phpmyadmin.net>