ANÁLISIS DE LOS RESULTADOS DE LA EVALUACIÓN ALBAYZIN-TTS 2008

Iñaki Sainz

Aholab Signal Processing Laboratory.
Departamento de Electrónica y Comunicaciones.
Universidad del País Vasco

RESUMEN

En el presente artículo se describe el proceso de evaluación de la campaña Albayzin TTS 2008, que tiene como propósito comparar las distintas técnicas e implementaciones de los conversores de texto a voz participantes. Asimismo se dan a conocer los resultados de la evaluación subjetiva acompañados de un análisis estadístico de los datos.

1. INTRODUCCIÓN

La campaña de evaluación Albayzin TTS 2008 tiene como propósito principal comparar las distintas técnicas e implementaciones de los sistemas participantes partiendo de una base de datos de voz común. Para ello, sigue la línea trazada por la evaluación internacional "Blizzard Challenge".

Cada participante ha dispuesto de un periodo de 7 semanas para generar una voz a partir de la base de datos proporcionada [1]. Tras dicho periodo, se debían sintetizar múltiples textos de test que posteriormente han sido evaluados de forma subjetiva bajo los siguientes criterios de calidad: Parecido con la voz original, Naturalidad e Inteligibilidad.

En este artículo se dan a conocer los resultados de la evaluación subjetiva. En la sección 2 se presentan los participantes que tomaron parte en Albayzin TTS 2008. En el apartado 3 se describen las secciones que formaban el test. Los resultados se presentan en la sección 4. Finalmente, se presentan unas conclusiones sobre todo el proceso en la sección 5. En el apéndice I se muestra una tabla resumen que recoge las características principales de cada sistema participante. El apéndice II agrupa todas las gráficas y tablas de resultados.

2. PARTICIPANTES

La campaña de evaluación Albayzin TTS 2008 ha contado en esta su primera edición, con la participación de 8 sistemas:

- Barcelona Media Centre & Cereproc.
- Grup de Recerca en Processament Multimodal (Univ. Ramón Llull).
- Grupo de conversión texto-voz de Telefónica I+D (sistema 1).
- Grupo de conversión texto-voz de Telefónica I+D (sistema 2).
- Grupo de Tecnologías Multimedia (Univ. Vigo).
- Madrid-BSDM (UPM, Univ. Edinburgh, Univ. Alcalá).
- Signal Processing Laboratory, AHOLAB (Univ. País Vasco).
- Universitat Politècnica de Catalunya (UPC).

Como puede observarse, el grupo de conversión texto-voz de Telefónica I+D ha presentado dos sistemas (con distintos métodos para la generación de la onda). En el apéndice I puede consultarse una tabla resumen con las características principales de cada sistema.

3. EVALUACIÓN

Con el propósito de evaluar la calidad de cada uno de los sistemas participantes, se ha llevado a cabo una evaluación subjetiva. Se ha optado por utilizar un interfaz web para facilitar el acceso a los evaluadores. Cada sistema participante debía proporcionar un mínimo de 5 evaluadores; el resto han sido evaluadores voluntarios.

3.1. Frases de Test

Una vez hubo acabado el periodo disponible para la construcción de la voz en castellano, cada sistema participante recibió 350 frases de test. Se disponía de una semana para sintetizarlas y enviarlas de vuelta. Las frases fueron extraídas de diversos libros y periódicos buscando un balanceado fonético.

Para poder evaluar la inteligibilidad de los sistemas se incluyeron asimismo 25 frases SUS (Semantically Unpredictable Sentences) generadas utilizando distintas estructuras [2]:

- Determinante + nombre + verbo intransitivo + preposición + determinante + adjetivo + nombre.
- Determinante + adjetivo + nombre + verbo transitivo + determinante + nombre.
- Adverbio + verbo transitivo + determinante + nombre + conjunción + determinante + nombre.

- Adverbio interrogativo + determinante + nombre + verbo transitivo + determinante + adjetivo + nombre.
- Determinante + nombre + verbo transitivo + determinante + pronombre relativo + verbo intransitivo.

3.2. Diseño de la evaluación

La evaluación constaba de 4 secciones para evaluar la calidad de las voces sintéticas en 3 aspectos: parecido con la voz original, naturalidad e inteligibilidad. El test fue diseñado para poder ser completado en una única sesión (alrededor de 30 minutos). Se ha utilizado un reproductor embebido para acelerar el proceso y evitar que las señales se abriesen en distintas ventanas del navegador, pero ha surgido algún problema de compatibilidad.

3.2.1. Sección 1 (Parecido con voz original)

Cada evaluador debía escuchar tres señales con grabaciones de voz natural para familiarizarse con la voz original. Dichas señales provenían de uno de los subcorpus de la voz proporcionada: Frases balanceadas, Párrafos balanceados, Párrafos literarios. A continuación, el evaluador debía escuchar otra señal y valorarla en una escala MOS (Mean Opinion Score) respecto a su parecido con la voz original. La puntuación se daba en un rango de 1 (voces completamente diferentes) y 5 (exactamente la misma voz). Un total de 9 señales eran evaluadas en esta sección: 1 por cada sistema participante + 1 proveniente de una grabación original.

3.2.2. Sección 2 (Naturalidad, parte 1)

Un total de 18 señales (2 por cada sistema participante + 2 grabaciones originales) eran evaluadas respecto a su naturalidad. Se utilizó una escala entre 1 (la voz no es nada natural) y 5 (la voz es completamente natural). A la hora de evaluar, el oyente valora tanto la suavidad en las concatenaciones como la naturalidad de la prosodia, pero se le pedía puntuar la naturalidad de forma global.

3.2.3. Sección 3 (Naturalidad, parte 2)

Esta sección es análoga a la sección precedente (sección 2) salvo que no se incluyen grabaciones de voz natural y el número de señales a evaluar asciende a 24 (3 señales por sistema participante).

3.2.4. Sección 4 (Inteligibilidad)

Cada evaluador debía escuchar 16 señales (2 por cada sistema participante) y escribir lo que reconocía en una ventana de texto. Se avisaba de que las frases podían no tener sentido y se pedía limitar el número de escuchas por frase. Como medida de la inteligibilidad de cada sistema, se ha computado el WER (word error rate) durante el análisis de los resultados.

3.3. Grupos de oyentes

A lo largo del test cada evaluador valoraba un total de 67 señales (9+28+24+16). Para evitar que escuchase la misma señal sintetizada por distintos sistemas, se han creado grupos de oyentes. De esta forma, todos los evaluadores escuchaban las mismas 67 frases y en el mismo orden, pero sintetizadas por distintos sistemas (o voz natural). Se han creado grupos independientes para cada una de las 4 secciones de la evaluación. Para cada sección se han formado tantos grupos como participantes (más voz natural si fuera el caso). Así por ejemplo, en la sección 1 había 9 grupos de forma que cada evaluador puntuaba una frase de cada sistema.

3.4. Datos evaluadores

Un total de 103 evaluadores completaron el test de forma íntegra: 107 completaron la sección 1, 106 la sección 2, 104 la sección 3 y 103 la sección 4.

Tomando en cuenta únicamente los datos relativos a los evaluadores que finalizaron el test (103), se extrapola la siguiente información:

- *Entorno de escucha*:
  - Auriculares: 81
  - Altavoces: 22
- *Información evaluador*:
  - Experto en tecnologías del habla: 54
  - Inexperto: 49
  - Hablante nativo de castellano: 94
  - No nativo: 9
  - Hombre: 66
  - Mujer: 37

4. RESULTADOS

A la hora de computar los resultados sólo se tuvo en cuenta a los evaluadores que hubieran completado la sección a evaluar. Como se ha dicho anteriormente, el propósito de dicha evaluación era valorar la similitud con la voz original, la naturalidad y la inteligibilidad de las voces participantes. Dado que el apartado que ha tenido un mayor peso en la evaluación ha sido el que se refiere a la naturalidad (secciones 2 y 3, 42 señales de un total de 67), las gráficas y tablas han sido ordenadas en función del score medio obtenido por cada sistema en las secciones 2 y 3. Cada participante tiene una letra asignada que identifica su sistema (A-H), la letra I se refiere a la señal de voz natural.

4.1. Medidas

Debido a que las escalas MOS no garantizan que el intervalo entre puntuaciones se mantenga constante (por ejemplo la mejora entre un 2 y un 1 no es

necesariamente idéntica a la que aparece entre un 4 y un 3), no es estadísticamente significativo comparar medias entre sistemas. Por ello se recomienda comparar medianas; aún así, mantenemos el ordenamiento de los sistemas en función de sus medias.

4.1.1. Boxplot

Para cada una de las medidas MOS se ha generado un gráfica boxplot como la de la figura 1. El rectángulo representa el rango entre el primer y el tercer cuartil. La mediana (o segundo cuartil) se representa mediante una raya roja horizontal. Los valores fuera de rango (outliers) se representan mediante una "x" y el resto de valores se agrupan mediante una línea continua.

Junto al gráfico boxplot se adjunta una tabla como la de la tabla 1, con información relativa a: Mínimo, Primer cuartil (1C), Segundo cuartil (2C), Tercer Cuartil (3C), Máximo, Límite inferior de confianza para la mediana (LIM) y Límite superior de confianza para la mediana (LSM).

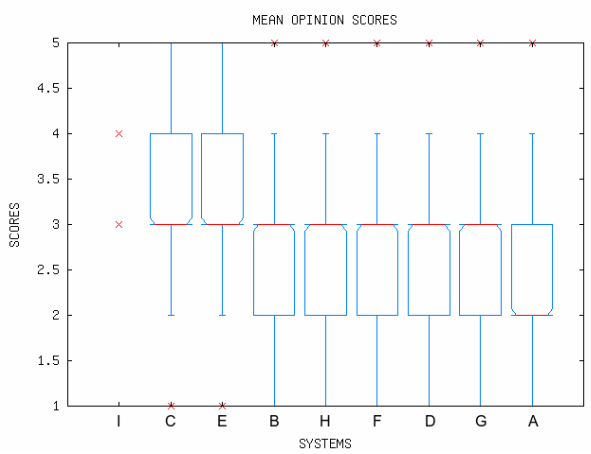


Figura 1. *Boxplot: MOS Naturalidad para todos los oyentes.*

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.06
E	1	3	3	4	5	3	3.06
B	1	2	3	3	5	2.93	3
H	1	2	3	3	5	2.93	3
F	1	2	3	3	5	2.93	3
D	1	2	3	3	5	2.93	3
G	1	2	3	3	5	2.93	3
A	1	2	2	3	5	2	2.06

Tabla 1. *Datos Boxplot: MOS Naturalidad para todos los oyentes.*

4.1.2. Wilcoxon

Para determinar si existían diferencias significativas entre los MOS de cada sistema, se han realizado tests de wilcoxon para datos apareados (pairwise Wilcoxon signed rank test). Como nivel de significancia se utilizó un valor de 0.05 con corrección de Bonferroni (dividiendo el nivel entre el número de hipótesis a evaluar). Los resultados se muestran en una tabla similar a la Tabla 3. Forma una matriz simétrica donde las diferencias estadísticamente significativas entre dos sistemas se reflejan con un 1, mientras que un 0 indica que no existen diferencias significativas.

4.2. Naturalidad

En la figura 1, tabla 1,2 y 3, se muestran los datos relativos al MOS de las secciones 2 y 3 para todos los oyentes. En el apéndice II pueden consultarse los resultados para los diferentes tipos de agrupaciones (escucha con auriculares, escucha con altavoces, expertos en tecnologías del habla, no-experto, hablantes nativos, no-nativos). La tabla 2 muestra información relativa a la Mediana, MAD (Median Absolute Deviation), Media, Desviación estándar y Número de señales evaluadas.

Sist	Median	MAD	Mean	SD	Samples
I	5	0.0	4.82	0.41	212
C	3	1.0	3.34	0.92	524
E	3	1.0	3.20	0.89	524
B	3	1.0	2.91	0.93	524
H	3	1.0	2.86	0.96	524
F	3	1.0	2.81	0.91	524
D	3	1.0	2.60	0.94	524
G	3	1.0	2.56	0.91	524
A	2	1.0	2.28	0.94	524

Tabla 2. *MOS Naturalidad para todos los oyentes.*

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	1	1	1	1	1	1
E	1	0		1	1	1	1	1	1
B	1	1	1		0	0	1	1	1
H	1	1	1	0		0	1	1	1
F	1	1	1	0	0		1	1	1
D	1	1	1	1	1	1		1	1
G	1	1	1	1	1	1	1		1
A	1	1	1	1	1	1	1	1	

Tabla 3. Wilcoxon test: MOS Naturalidad para todos los oyentes

4.3. Similitud con la voz original

En la figura 2, tablas 4,5 y 6, se muestran los datos relativos al MOS de la sección 1 para todos los oyentes. En el apéndice II pueden consultarse los resultados para diferentes tipos de agrupaciones.

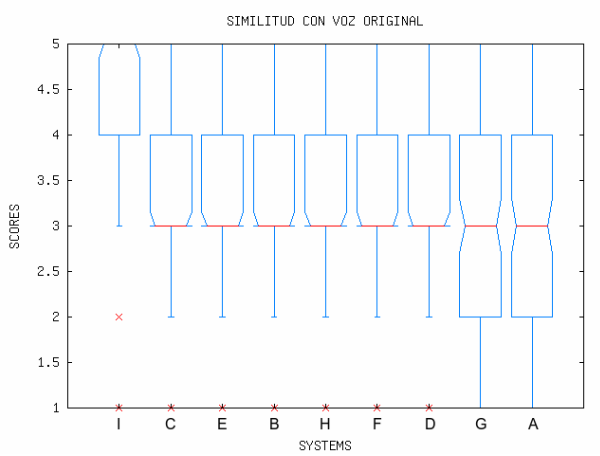


Figura 2. *Boxplot: MOS Similitud para todos los oyentes.*

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.06
E	1	3	3	4	5	3	3.06
B	1	2	3	3	5	2.93	3
H	1	2	3	3	5	2.93	3
F	1	2	3	3	5	2.93	3
D	1	2	3	3	5	2.93	3
G	1	2	3	3	5	2.93	3
A	1	2	2	3	5	2	2.06

Tabla 4. *Datos Boxplot: MOS Similitud para todos los oyentes.*

Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.11	1.17	107
C	3	1	3.25	0.84	107
E	3	1	3.35	0.94	107
B	3	1	3.36	0.93	107
H	3	1	3.29	0.91	107
F	3	1	3.23	0.81	107
D	3	1	3.11	0.89	107
G	3	1	3.07	0.92	107
A	3	1	2.96	0.98	107

Tabla 5. *MOS Similitud para todos los oyentes.*

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	0	0	0	0	0	0
E	1	0		0	0	0	0	0	1
B	1	0	0		0	0	0	1	1
H	1	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	0	0	0		0	0
G	1	0	0	1	0	0	0		0
A	1	0	1	1	0	0	0	0	

Tabla 6. *Wilcoxon test: MOS Similitud para todos los oyentes.*

4.4. Inteligibilidad

En la figura 3 se muestran los datos relativos al Word Error Rate (WER). En este caso se han utilizado únicamente las valoraciones de los oyentes nativos, ya que los no nativos (ver apéndice II) tenían una tasa de error muy superior debida seguramente a un menor dominio del idioma. En la tabla 7 se detallan los tipos de errores computados en el WER: Substitutions (S), Insertions (I), Deletions (D).

Antes de llevar a cabo la evaluación de la inteligibilidad, se realizaron una serie de correcciones en las respuestas de los evaluadores:

- Se prescindió de las tildes.
- Errores ortográficos: andubo, cavallo, i->y...
- Errores tipográficos claros: tranpsarencia,...

No se cree que dichas correcciones hayan tenido incidencia en las bajas tasas de error de todos los sistemas. Quizá el caso más peliagudo se refiera al hecho de no tener en cuenta las tildes ya que puede variar el significado de ciertas palabras empleadas en la evaluación: camino vs. caminó, intérprete vs. interpreté, etc. Pero la realidad es que buena parte de los evaluadores no utilizaron tildes o no lo hicieron de manera consistente a lo largo de sus respuestas.

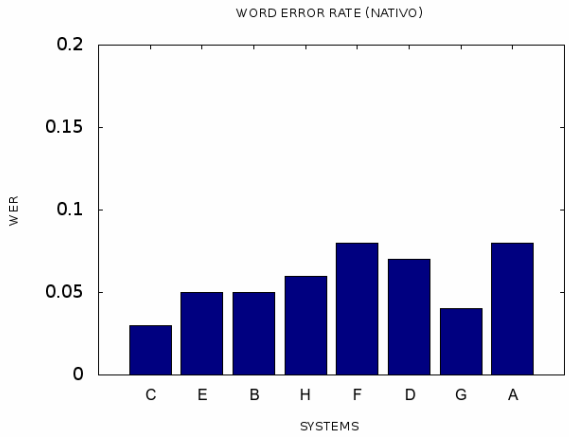


Figura 3. *WER para oyentes nativos*

5. CONCLUSIONES

Analizando los resultados relativos a la naturalidad pueden extrapolarse dos conclusiones principales:

1. Existe un amplio margen de mejora hasta alcanzar el nivel de la voz natural (Mediana 5 para la voz original, respecto al 3 de la mejor voz sintética).

2. Dos sistemas (C y E) obtienen unos resultados significativamente mejores que el resto de sistemas (consultar tabla 3.).

Sist	WER	Samples	Words	S	I	D
C	0.03	188	1233	29	5	9
E	0.05	188	1234	40	4	15
B	0.05	188	1233	47	13	1
H	0.06	188	1234	68	8	3
F	0.08	188	1233	62	29	10
D	0.07	188	1235	61	20	8
G	0.04	188	1234	30	3	12
A	0.08	188	1234	70	4	22

Tabla 7. *WER (substitutions, insertion, deletions) para oyentes nativos*

Curiosamente ambos sistemas (C y E) son los únicos que realizan algún tipo de modificación espectral de la señal: directamente mediante un modelo sinusoidal, o implícitamente utilizando modelos estadísticos para generar la excitación. Por lo tanto, se podría concluir que al menos para bases de datos de tamaño pequeño/mediano y un dominio de síntesis no restringido, este tipo de sistemas resultan más adecuados que los típicos de selección de unidades en los que las modificaciones de la señal natural se limitan a aspectos prosódicos (duración y pitch).

Cabría pensar que una mayor suavidad espectral de la voz sintética en las transiciones, traería consigo una considerable pérdida en la calidad de voz respecto a la natural. Pero la realidad es que ambos sistemas (C y E) obtienen buenas puntuaciones en la sección 1 referida a la similitud de la voz sintética respecto a la voz natural. Tampoco está claro hasta qué punto los ruidos segmentales o problemas relativos a la prosodia influyeron en la puntuación de los oyentes en dicha sección. Es posible que las respuestas del grupo de oyentes "expertos" sufran este problema en menor medida. En cualquier caso, en una próxima edición de Albayzin sería conveniente ser más explícito a la hora de explicar que en la sección 1 se pedía evaluar únicamente el parecido con la voz original, abstrayéndose en la medida de lo posibles de otros aspectos que se puntúan en las secciones de naturalidad.

En lo que se refiere a la sección de inteligibilidad y observando las bajas tasas de error, parece claro que este aspecto de la evaluación no resulta tan primordial como para otros idiomas (ver resultados de evaluaciones Blizzard Challenge para el inglés). Aunque lo ideal hubiera sido disponer de grabaciones de voz natural en esta sección para poder concluir si existe un salto de calidad comparable al de la naturalidad entre el mejor sistema y la voz natural.

5.1. Feedback oyentes

Una vez se había completado el test, los evaluadores tenían la posibilidad de realizar comentarios adicionales. Los más interesantes se resumen a continuación:

- *Sobre la inteligibilidad*: Algunos manifestaban que la prueba había sido demasiado sencilla debido al uso de palabras excesivamente largas. He de admitir que no se dispuso de tiempo suficiente para generar frases SUS fonéticamente balanceadas (se crearon a mano) y seguramente este hecho ha influido en la "sencillez" del test. Por último, hay quien comenta la posibilidad de incluir el número de escuchas realizadas, para poder ponderar el WER de cada frase.
- *Sobre el uso de altavoces*: Ha habido quien recomendaba que se usasen únicamente evaluaciones realizadas mediante auriculares. Si bien es cierto que las puntuaciones de ciertos sistemas varían prescindiendo de las evaluaciones realizadas con altavoces, se ha optado finalmente por utilizar todas las respuestas. El resto de resultados para auriculares/altavoces pueden consultarse en el apéndice II. También cabe comentar que antes de realizar el test se normalizó la potencia de las señales de todos los participantes, pero debido a que las señales de UPM tenían un rango dinámico grande, para evitar saturaciones (o utilizar un nivel medio muy bajo para todos los sistemas) sus señales quedaron a un nivel ligeramente inferior al resto de participantes.
- *Sobre el test*: Ha habido recomendaciones para utilizar párrafos más largos que permitan evaluar mejor la entonación. Otros se han quejado de que había frases más largas que otras. Se ha comentado la posibilidad de realizar comparaciones entre parejas de sistemas sobre una misma frase, evaluar la naturalidad en varios aspectos (prosodia, ruido segmental), o permitir la escucha de todas las señales de una sección antes de puntuar.

AGRADECIMIENTOS

Agradecer (una vez más) el esfuerzo realizado por todos los sistemas participantes inscritos en la campaña de evaluación Albayzin TTS 2008, y a todos los evaluadores voluntarios que hicieron esta campaña posible.

Esta campaña de evaluación y análisis de resultados se ha basado en el trabajo realizado por la organización del Blizzard Challenge durante sus múltiples ediciones.

El análisis estadístico ha sido realizado con el paquete Octave, y las gráficas con Gnuplot.

BIBLIOGRAFÍA

[1] Bonafonte, A., Moreno, A., "Documentation of the upc_esma spanish database", *TALP Research Center*, Universitat Politecnica de Catalunya, Barcelona; Spain , 2008.

[2] Grice, Martine., "Syntactic structures and lexicon requirements for semantically unpredictable sentences in a number of languages", *In SIOA, Vol.2, 19-22*. 1989.

APÉNDICE I

Sistema	Grupo			Sistema					Técnica		
	Personas	Horas/ persona	Nativos Castellano	Nombre Sistema	Años Desarrollo	Disponibilidad	Plataformas	Lenguaje	Memoria Ocupada	Unidad base	Tipo
BM / Cereproc	1 + (ayuda puntual de 2 personas)	40	Sí	Cerevoice	3 años	Comercial / Investigación	Linux, Windows, Mac	C (comercial) / Python (investigación)	21,4MB	Difonema	Concatenativo
Madrid-BSDM	6	40	Sí	HTS-2005 (Adaptado Castellano)	5 años	Festival y HTS (código abierto) / Straight (comercial, investigación)	Linux	C, C++, scheme, bash, perl, python	2,2 MB (voz) / 2.8GB (entrenamiento)	pentafonema	HMM
Telefónica I	3	85	Sí	-	18 años	Comercial	Windows, Solaris, Linux	C, C++	100MB	Difonema	Concatenativo
Telefónica II	3	40	Sí	-	18 años	Comercial	Windows, Solaris, Linux	C, C++	50MB	Difonema	Concatenativo
UPC	1	10	Sí	Ogmios	12 años	Investigación	Windows, Linux	C++	200MB	Semifonema con contexto	Concatenativo
UPV-EHU	1 + (ayuda puntual de 1 persona)	40	Sí	AhoTTS	13 años	Comercial / Investigación	Linux, Windows	C, C++	200MB	Semifonema con contexto	Concatenativo
URL	3	20	Sí	-	12 años	Investigación / Comercial (puntualmente)	Windows	C++	RAM: 200MB HD: 2GB (voz) 38GB (costes precalculados)	Difonema, Trifonema en algún caso	Concatenativo
UVigo	2	80	Sí	Cotovía	10 años	Investigación	Linux, Windows	C, C++	RAM:70MB HD: 300MB	Semifonema con contexto	Concatenativo

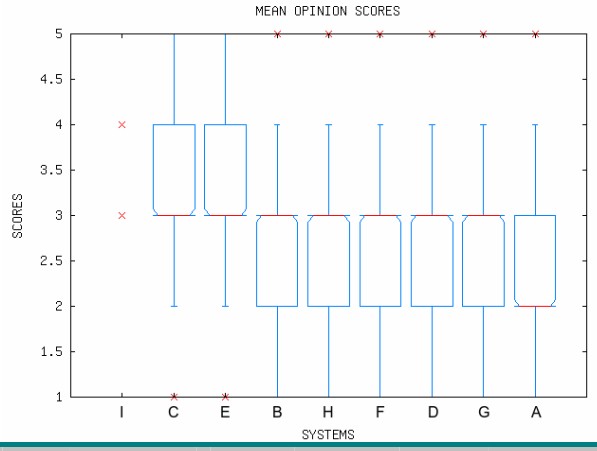
Sistema	Construcción Voz			Componentes			
	Resegmentado	Herramientas Segmentado	Corrección Manual	Componentes externos	Modelo Prosódico	Coste Objetivo	Coste Concatenación
BM / Cereproc	Sí	Cerevoice (VCK)	No	HTK	Muy simple, no adaptado al locutor	Acento, límites prosódicos, F0, duración	LSF, F0, Energía
Madrid-BSDM	Sí	HTS	Pequeñas Correcciones (30 minutos)	Festival, HTS	HMMs dependientes del contexto	No se usa	No se usa
Telefónica I	Sí	Sistema propio	Sí, 60 horas/persona	No	Duración: estadístico multiplicativo Entonación: selección por corpus	Identidad sonidos, contexto, prosodia	Identidad contexto, continuidad, prosodia
Telefónica II	ídem	ídem	ídem	ídem	ídem	ídem	ídem
UPC	Sí	Sistema propio RAMSES	No	RAMSES, Praat (f0)	Entonación basada en datos usando CART	prosodia, segmental (contextos), suprasegmental (palabra, acento, posición en frase)	F0, criterios fonológicos, distancia espectral
UPV-EHU	No	-	No	ESPS (pitch marking)	Duración: CART Entonación: selección por corpus	Info lingüística (ventana de 5 fonemas), Prosodia, acento, posición frase	F0, energía, Distancia espectral
URL	Sí	HTK	No	HTK, esps	No usa modelo prosódico	No usa coste objetivo	Diferencias pitch, energía y MFCC
UVigo	Sí	HTK	Sí, 70 horas/persona	Festival, Praat y HTK	Modelo entonativo basado en corpus	F0, duración, energía y envolvente espectral	F0, energía, y envolvente espectral (12 MFCC)

Sistema	Datos				Procesamiento de Señal				
	BD Externa	Poda	1ª Voz castellano	Comentarios	Representación espectral	Representación de fuente	Marcas pitch	Modificación	Representación espectral
BM / Cereproc	No	-	No	BD más pequeña.	LSF	-	VCK (utiliza epoch y getf0 de ESPS)	No	LSF
Madrid-BSDM	No	No	Sí	BD pequeña. ¿Financiar grabación BD mayor, entre grupos investigación?	Mel Cepstrum estimados con STRAIGHT	Modelo de excitación mixta	get_f0 (HTS), tempo	Modelo de producción tipo vocoder. Señal generada cada periodo de pitch (componente espectral, excitación, duraciones)	Mel Cepstrum estimados con STRAIGHT
Telefónica I	No	Se utilizó únicamente locuciones con segmentación revisada	No	Voz un poco monótona y BD pequeña.	Modelo sinusoidal	No directamente	No se precisa la localización explícita de onsets	Modificación selectiva de valores prosódicos (duración y F0), interpolación (amplitud, F0, espectro) en puntos de pegado	Modelo sinusoidal
Telefónica II	ídem	ídem	ídem	ídem	No	No	Información de onsets obtenida mediante análisis sinusoidal	Modificación selectiva de valores prosódicos (duración y F0), interpolación de la amplitud en puntos de pegado	No
UPC	No	Ninguna poda para la parte prosódica	No	-	No	No	Marcas generadas con Praat	Mínimas modificaciones de f0 y duración	No
UPV-EHU	No	Sí, score de la segmentación y distancia espectral al centroide	Sí	-	No	No	ESPS + corrección automática de pitch doubling/halving comparando con método de extracción de pitch propio	Modificaciones de duración, ligero suavizado de pitch en la concatenación	No
URL	No	Sí, outliers (marcas de segmentación y pitch)	No	-	MFCC para los costes	No	get_f0 + filtrado y postprocesado de las marcas	No	MFCC para los costes
UVigo	No	Sí, distancia de cada semifonema a su centroide	No	-	MFCC	No	Se empleó Praat	Modificación de F0 y duración cuando excede un umbral	MFCC

Sistema	Sistema Presentado		
	¿Mejor voz del sistema?	¿Qué parte del sistema destacarías?	¿Qué parte del sistema necesita ser mejorado?
BM / Cereproc	No, BD pequeña	Herramienta de creación de voces que automatiza el proceso	Tratamiento de la prosodia
Madrid-BSDM	De las generadas para castellano sí	Modelado acústico basado en HMM: mayor robustez y estabilidad que la selección de unidades para BD de este tamaño.	Analizador prosódico, morfo-sintáctico y el pausador
Telefónica I	No, base de datos con cobertura escasa, y no se ha dedicado tiempo suficiente a la construcción de la voz	Prosodia natural y suavidad de la señal sintética	¡Nos gustaría poder mejorar todos!
Telefónica II	No, base de datos con cobertura escasa, y no se ha dedicado tiempo suficiente a la construcción de la voz	Prosodia natural	El módulo de síntesis de la señal de voz
UPC	No, la voz se ha desarrollado automáticamente. Proceso de rechazo de unidades mal segmentadas ajustado para BD mayores	Modelo explícito de la prosodia	Concatenación de unidades, el tratamiento de los finales tanto a nivel de prosodia como de generación
UPV-EHU	No, debido principalmente a haber prescindido de la revisión manual del etiquetado	Selección de unidades	Sistema de generación de onda.
URL	No, se han obtenido mejores resultados en aplicaciones de dominio restringido	Rápida puesta a punto	Selección de unidades, procesado de la señal y modelado prosódico
UVigo	No, por la base de datos y el idioma	La prosodia, más rica y variable de lo habitual	Procesado de señal y la parte de revisión manual al generar la voz

APÉNDICE II:
NATURALIDAD (Sección 2 y 3)

- Naturalidad - Todos los Oyentes

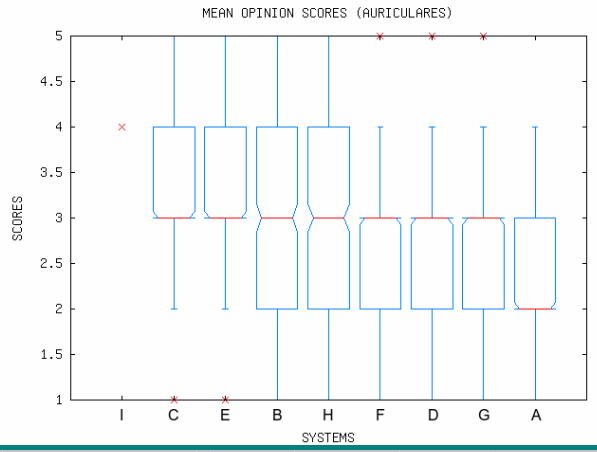


Sist	Median	MAD	Mean	SD	Samples
I	5	0.0	4.82	0.41	212
C	3	1.0	3.34	0.92	524
E	3	1.0	3.20	0.89	524
B	3	1.0	2.91	0.93	524
H	3	1.0	2.86	0.96	524
F	3	1.0	2.81	0.91	524
D	3	1.0	2.60	0.94	524
G	3	1.0	2.56	0.91	524
A	2	1.0	2.28	0.94	524

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.06
E	1	3	3	4	5	3	3.06
B	1	2	3	3	5	2.93	3
H	1	2	3	3	5	2.93	3
F	1	2	3	3	5	2.93	3
D	1	2	3	3	5	2.93	3
G	1	2	3	3	5	2.93	3
A	1	2	2	3	5	2	2.06

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	1	1	1	1	1	1
E	1	0		1	1	1	1	1	1
B	1	1	1		0	0	1	1	1
H	1	1	1	0		0	1	1	1
F	1	1	1	0	0		1	1	1
D	1	1	1	1	1	1		1	1
G	1	1	1	1	1	1	1		1
A	1	1	1	1	1	1	1	1	

- Naturalidad - Auriculares

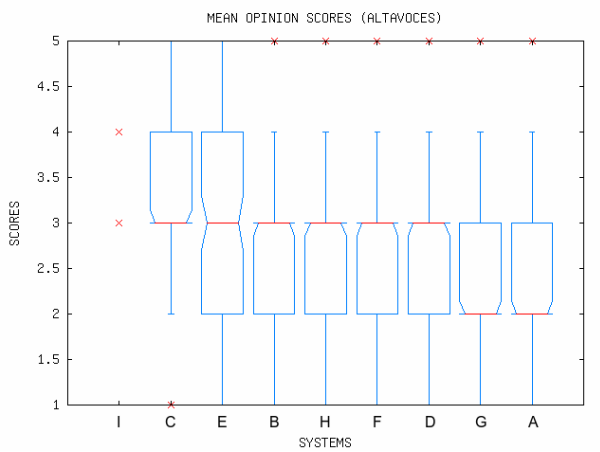


Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.83	0.38	162
C	3	1	3.3	0.93	405
E	3	1	3.24	0.89	405
B	3	1	2.93	0.93	405
H	3	1	2.9	0.96	405
F	3	1	2.81	0.89	405
D	3	1	2.57	0.92	405
G	3	1	2.6	0.91	405
A	2	1	2.24	0.92	405

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	4	5	5	5	5	5	5
C	1	3	3	4	5	3	3.07
E	1	3	3	4	5	3	3.07
B	1	2	3	4	5	2.84	3.15
H	1	2	3	4	5	2.84	3.15
F	1	2	3	3	5	2.92	3
D	1	2	3	3	5	2.92	3
G	1	2	3	3	5	2.92	3
A	1	2	2	3	4	2	2.07

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	1	1	1	1	1	1
E	1	0		1	1	1	1	1	1
B	1	1	1		0	0	1	1	1
H	1	1	1	0		0	1	1	1
F	1	1	1	0	0		1	0	1
D	1	1	1	1	1	1		0	1
G	1	1	1	1	1	0	0		1
A	1	1	1	1	1	1	1	1	

- Naturalidad - Altavoces

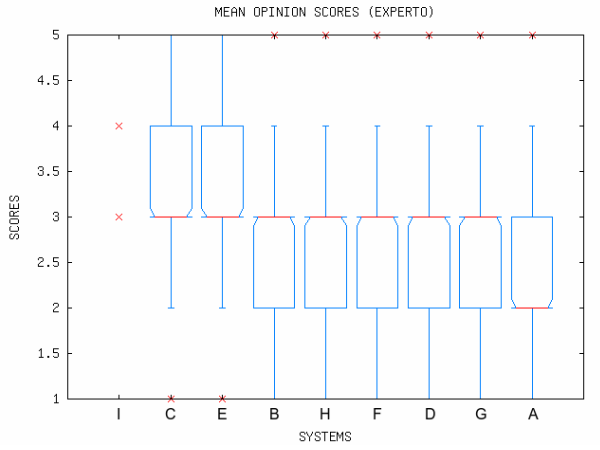


Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.14
E	1	2	3	4	5	2.7	3.29
B	1	2	3	3	5	2.85	3
H	1	2	3	3	5	2.85	3
F	1	2	3	3	5	2.85	3
D	1	2	3	3	5	2.85	3
G	1	2	2	3	5	2	2.14
A	1	2	2	3	5	2	2.14

Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.77	0.52	44
C	3	1	3.48	0.88	110
E	3	1	3.02	0.89	110
B	3	1	2.85	0.97	110
H	3	1	2.71	0.94	110
F	3	1	2.8	0.99	110
D	3	1	2.64	0.99	110
G	2	1	2.41	0.91	110
A	2	1	2.4	1.02	110

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		1	1	1	1	1	1	1
E	1	1		0	1	0	1	1	1
B	1	1	0		0	0	0	1	1
H	1	1	1	0		0	0	0	0
F	1	1	0	0	0		0	0	0
D	1	1	1	0	0	0		0	0
G	1	1	1	1	0	0	0		0
A	1	1	1	1	0	0	0	0	

- Naturalidad - Expertos

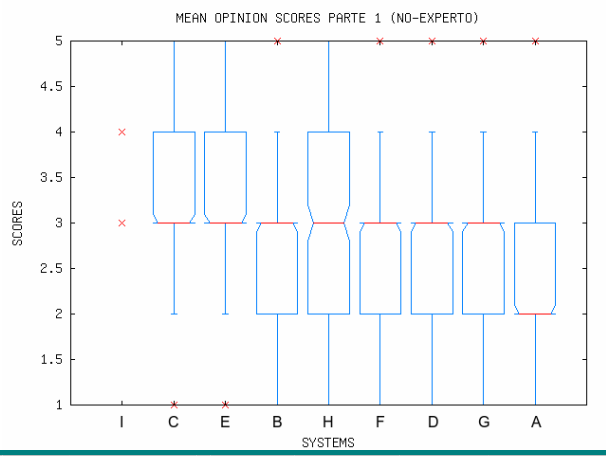


Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.09
E	1	3	3	4	5	3	3.09
B	1	2	3	3	5	2.9	3
H	1	2	3	3	5	2.9	3
F	1	2	3	3	5	2.9	3
D	1	2	3	3	5	2.9	3
G	1	2	3	3	5	2.9	3
A	1	2	2	3	5	2	2.09

Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.88	0.35	110
C	3	1	3.36	0.9	272
E	3	1	3.25	0.85	272
B	3	1	2.92	0.9	272
H	3	1	2.9	0.86	272
F	3	1	2.86	0.91	272
D	3	1	2.53	0.9	272
G	3	1	2.63	0.87	272
A	2	1	2.2	0.91	272

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	1	1	1	1	1	1
E	1	0		1	1	1	1	1	1
B	1	1	1		0	0	1	1	1
H	1	1	1	0		0	1	1	1
F	1	1	1	0	0		1	0	1
D	1	1	1	1	1	1		0	1
G	1	1	1	1	1	0	0		1
A	1	1	1	1	1	1	1	1	

- Naturalidad - No-Expertos

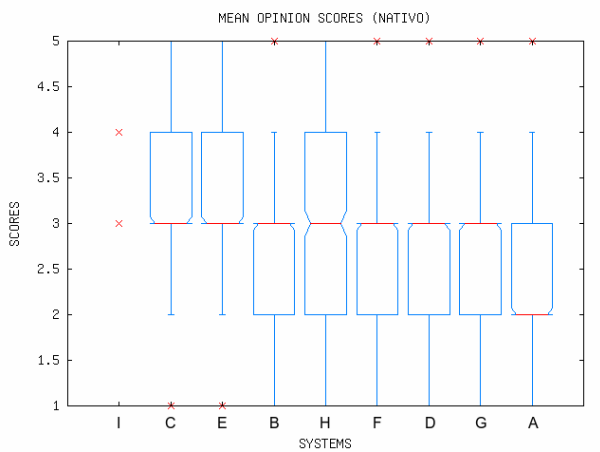


Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.75	0.45	102
C	3	1	3.33	0.94	252
E	3	1	3.15	0.92	252
B	3	1	2.89	0.96	252
H	3	1	2.82	1.06	252
F	3	1	2.75	0.91	252
D	3	1	2.66	0.98	252
G	3	1	2.49	0.94	252
A	2	1	2.37	0.96	252

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.09
E	1	3	3	4	5	3	3.09
B	1	2	3	3	5	2.9	3
H	1	2	3	4	5	2.8	3
F	1	2	3	3	5	2.9	3
D	1	2	3	3	5	2.9	3
G	1	2	3	3	5	2.9	3
A	1	2	2	3	5	2	2.09

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	1	1	1	1	1	1
E	1	0		1	1	1	1	1	1
B	1	1	1		0	0	1	1	1
H	1	1	1	0		0	0	1	1
F	1	1	1	0	0		0	0	1
D	1	1	1	1	0	0		0	1
G	1	1	1	1	1	0	0		0
A	1	1	1	1	1	1	1	0	

- Naturalidad - Nativos

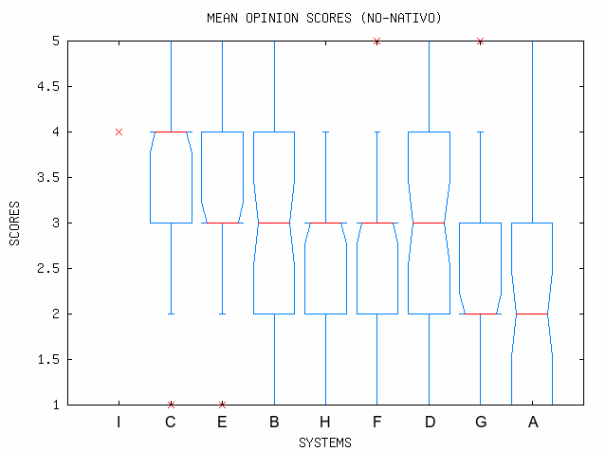


Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.81	0.42	194
C	3	1	3.33	0.91	479
E	3	1	3.2	0.88	479
B	3	1	2.91	0.91	479
H	3	1	2.88	0.96	479
F	3	1	2.8	0.9	479
D	3	1	2.57	0.93	479
G	3	1	2.57	0.89	479
A	2	1	2.29	0.92	479

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.07
E	1	3	3	4	5	3	3.07
B	1	2	3	3	5	2.92	3
H	1	2	3	4	5	2.85	3.14
F	1	2	3	3	5	2.92	3
D	1	2	3	3	5	2.92	3
G	1	2	3	3	5	2.92	3
A	1	2	2	3	5	2	2.07

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		1	1	1	1	1	1	1
E	1	1		1	1	1	1	1	1
B	1	1	1		0	0	1	1	1
H	1	1	1	0		0	1	1	1
F	1	1	1	0	0		1	1	1
D	1	1	1	1	1	1		1	1
G	1	1	1	1	1	1	1		1
A	1	1	1	1	1	1	1	1	

- Naturalidad - No-Nativos



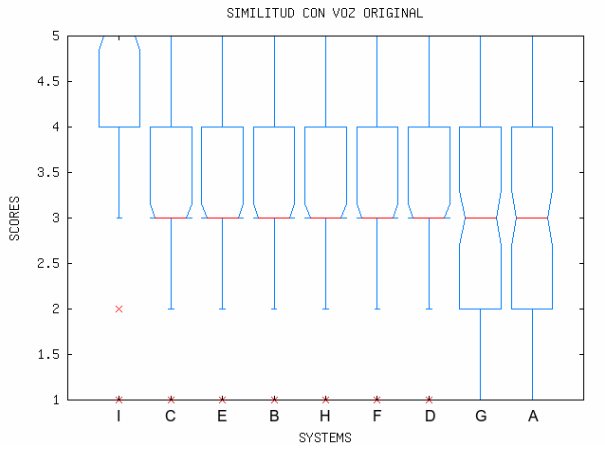
Sist	Min	1C	2C	3C	Max	LIM	LSM
I	4	5	5	5	5	5	5
C	1	3	4	4	5	3.76	4
E	1	3	3	4	5	3	3.23
B	1	2	3	4	5	2.53	3.46
H	1	2	3	3	4	2.76	3
F	1	2	3	3	5	2.76	3
D	1	2	3	4	5	2.53	3.46
G	1	2	2	3	5	2	2.23
A	1	1	2	3	5	1.53	2.46

Sist	Median	MAD	Mean	SD	Samples
I	5	0.0	4.89	0.32	18
C	4	1.0	3.49	1.04	45
E	3	1.0	3.24	0.98	45
B	3	1.0	2.82	1.09	45
H	3	1.0	2.64	0.98	45
F	3	1.0	2.84	1.02	45
D	3	1.0	2.91	0.95	45
G	2	1.0	2.49	1.06	45
A	2	1.0	2.18	1.07	45

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	1	1	1	1	1	1
E	1	0		0	1	0	0	1	1
B	1	1	0		0	0	0	0	1
H	1	1	1	0		0	0	0	0
F	1	1	0	0	0		0	0	0
D	1	1	0	0	0	0		0	0
G	1	1	1	0	0	0	0		0
A	1	1	1	1	0	0	0	0	

SIMILITUD (Sección 1)

- Similitud - Todos los oyentes

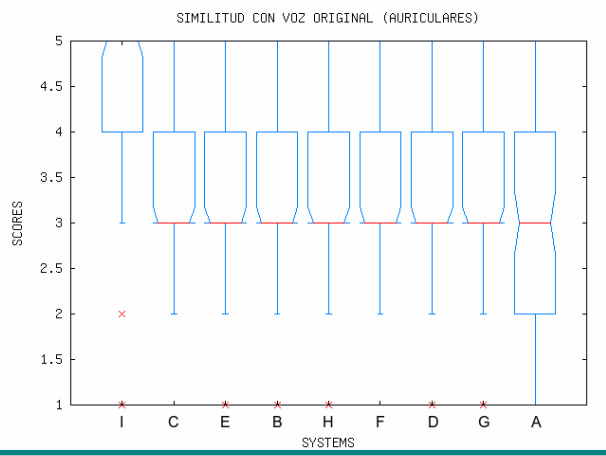


Sist	Min	1C	2C	3C	Max	LIM	LSM
I	3	5	5	5	5	5	5
C	1	3	3	4	5	3	3.06
E	1	3	3	4	5	3	3.06
B	1	2	3	3	5	2.93	3
H	1	2	3	3	5	2.93	3
F	1	2	3	3	5	2.93	3
D	1	2	3	3	5	2.93	3
G	1	2	3	3	5	2.93	3
A	1	2	2	3	5	2	2.06

Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.11	1.17	107
C	3	1	3.25	0.84	107
E	3	1	3.35	0.94	107
B	3	1	3.36	0.93	107
H	3	1	3.29	0.91	107
F	3	1	3.23	0.81	107
D	3	1	3.11	0.89	107
G	3	1	3.07	0.92	107
A	3	1	2.96	0.98	107

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	0	0	0	0	0	0
E	1	0		0	0	0	0	0	1
B	1	0	0		0	0	0	1	1
H	1	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	0	0	0		0	0
G	1	0	0	1	0	0	0		0
A	1	0	1	1	0	0	0	0	

- Similitud - Auriculares

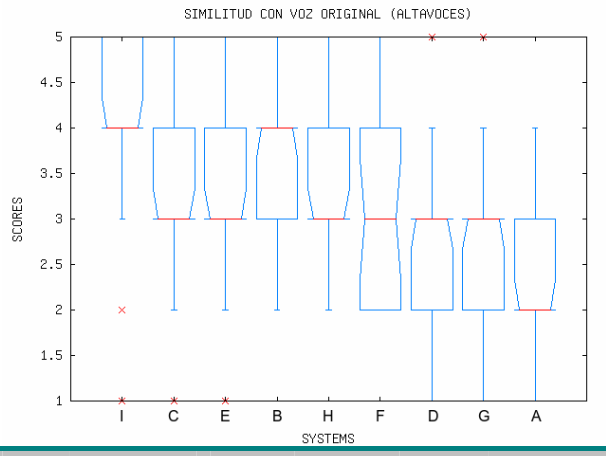


Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.09	1.19	81
C	3	1	3.27	0.79	81
E	3	1	3.35	0.92	81
B	3	1	3.36	0.97	81
H	3	1	3.3	0.91	81
F	3	1	3.32	0.72	81
D	3	1	3.19	0.9	81
G	3	1	3.14	0.9	81
A	3	1	3.07	0.97	81

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	1	4	5	5	5	4.82	5
C	2	3	3	4	5	3	3.17
E	1	3	3	4	5	3	3.17
B	1	3	3	4	5	3	3.17
H	1	3	3	4	5	3	3.17
F	2	3	3	4	5	3	3.17
D	1	3	3	4	5	3	3.17
G	1	3	3	4	5	3	3.17
A	1	2	3	4	5	2.65	3.34

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	0	0	0	0	0	0
E	1	0		0	0	0	0	0	0
B	1	0	0		0	0	0	0	0
H	1	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	0	0	0		0	0
G	1	0	0	0	0	0	0		0
A	1	0	0	0	0	0	0	0	

- Similitud - Altavoces

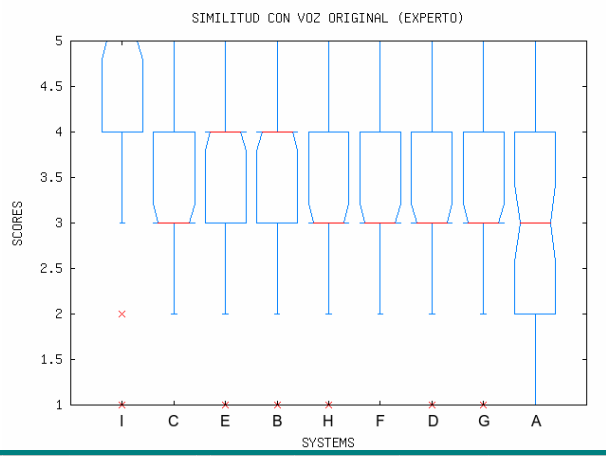


Sist	Median	MAD	Mean	SD	Samples
I	4	0.5	4.09	1.19	22
C	3	1	3.23	1.02	22
E	3	0.5	3.27	1.08	22
B	4	0.5	3.41	0.85	22
H	3	1	3.32	0.95	22
F	3	1	3.05	0.95	22
D	3	0.5	2.86	0.89	22
G	3	1	2.91	1.02	22
A	2	1	2.45	0.91	22

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	1	4	4	5	5	4	4.33
C	1	3	3	4	5	3	3.33
E	1	3	3	4	5	3	3.33
B	2	3	4	4	5	3.66	4
H	2	3	3	4	5	3	3.33
F	2	2	3	4	5	2.33	3.66
D	1	2	3	3	5	2.66	3
G	1	2	3	3	5	2.66	3
A	1	2	2	3	4	2	2.33

	I	C	E	B	H	F	D	G	A
I		1	0	0	0	1	1	1	1
C	1		0	0	0	0	0	0	0
E	0	0		0	0	0	0	0	1
B	0	0	0		0	0	1	0	1
H	0	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	1	0	0		0	0
G	1	0	0	0	0	0	0		0
A	1	0	1	1	0	0	0	0	

- Similitud - Expertos

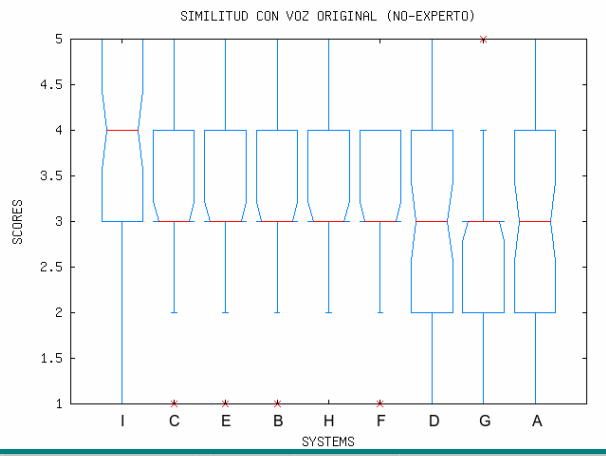


Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.22	1.18	55
C	3	1	3.23	0.81	55
E	4	1	3.6	0.89	55
B	4	1	3.49	0.98	55
H	3	1	3.38	0.93	55
F	3	1	3.29	0.88	55
D	3	1	3.25	0.89	55
G	3	1	3.27	0.93	55
A	3	1	3.11	0.96	55

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	1	4	5	5	5	4.78	5
C	2	3	3	4	5	3	3.20
E	1	3	4	4	5	3.78	4
B	1	3	4	4	5	3.78	4
H	1	3	3	4	5	3	3.21
F	2	3	3	4	5	3	3.21
D	1	3	3	4	5	3	3.21
G	1	3	3	4	5	3	3.21
A	1	2	3	4	5	2.57	3.42

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	0	0	0	0	0	0
E	1	0		0	0	0	0	0	1
B	1	0	0		0	0	0	0	0
H	1	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	0	0	0		0	0
G	1	0	0	0	0	0	0		0
A	1	0	1	0	0	0	0	0	

- Similitud - No-Expertos

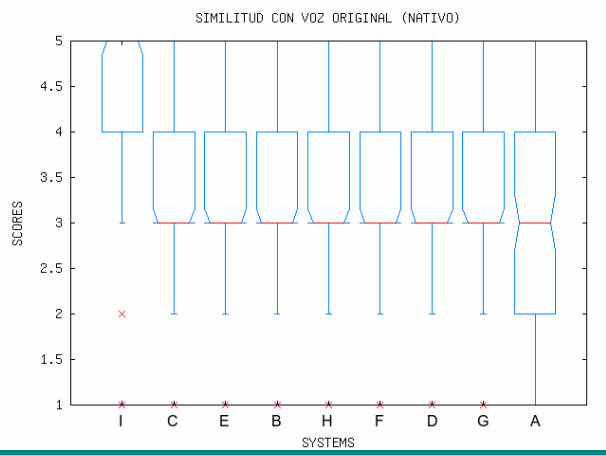


Sist	Median	MAD	Mean	SD	Samples
I	4	1	4	1.15	52
C	3	1	3.27	0.87	52
E	3	0.5	3.08	0.93	52
B	3	1	3.21	0.87	52
H	3	1	3.19	0.89	52
F	3	0.5	3.17	0.73	52
D	3	1	2.96	0.88	52
G	3	1	2.85	0.87	52
A	3	1	2.81	0.99	52

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	1	3	4	5	5	3.56	4.43
C	1	3	3	4	5	3	3.21
E	1	3	3	4	5	3	3.21
B	1	3	3	4	5	3	3.21
H	2	3	3	4	5	3	3.21
F	1	3	3	4	4	3	3.21
D	1	2	3	4	5	2.56	3.43
G	1	2	3	3	5	2.78	3
A	1	2	3	4	5	2.56	3.43

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	0	0	0	0	1	0
E	1	0		0	0	0	0	0	0
B	1	0	0		0	0	0	1	1
H	1	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	0	0	0		0	0
G	1	1	0	1	0	0	0		0
A	1	0	0	1	0	0	0	0	

- Similitud - Nativos

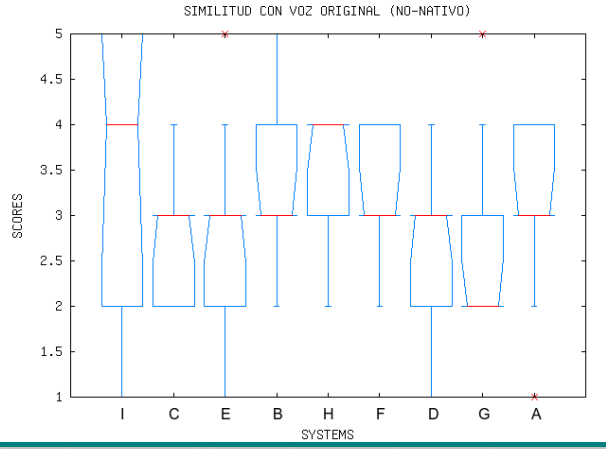


Sist	Median	MAD	Mean	SD	Samples
I	5	0	4.16	1.12	98
C	3	1	3.29	0.84	98
E	3	1	3.4	0.91	98
B	3	1	3.36	0.93	98
H	3	1	3.29	0.92	98
F	3	1	3.24	0.81	98
D	3	1	3.15	0.88	98
G	3	1	3.09	0.91	98
A	3	1	2.96	0.98	98

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	1	4	5	5	5	4.84	5
C	1	3	3	4	5	3	3.15
E	1	3	3	4	5	3	3.15
B	1	3	3	4	5	3	3.15
H	1	3	3	4	5	3	3.15
F	1	3	3	4	5	3	3.15
D	1	3	3	4	5	3	3.15
G	1	3	3	4	5	3	3.15
A	1	2	3	4	5	2.68	3.31

	I	C	E	B	H	F	D	G	A
I		1	1	1	1	1	1	1	1
C	1		0	0	0	0	0	0	0
E	1	0		0	0	0	0	0	1
B	1	0	0		0	0	0	1	1
H	1	0	0	0		0	0	0	0
F	1	0	0	0	0		0	0	0
D	1	0	0	0	0	0		0	0
G	1	0	0	1	0	0	0		0
A	1	0	1	1	0	0	0	0	

- Similitud - No-Nativos



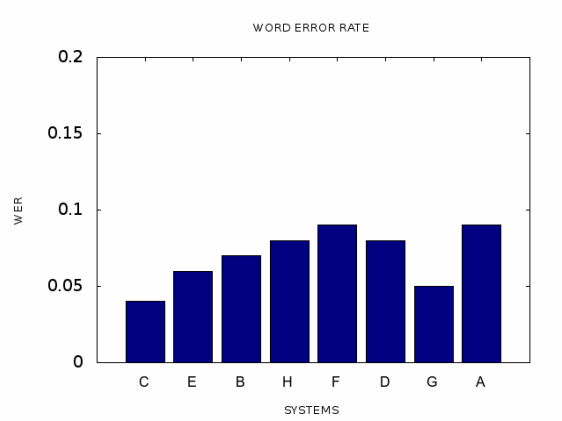
Sist	Median	MAD	Mean	SD	Samples
I	4	1	3.56	1.59	9
C	3	1	2.89	0.78	9
E	3	1	2.78	1.2	9
B	3	1	3.33	1	9
H	4	0	3.33	0.87	9
F	3	1	3.11	0.78	9
D	3	1	2.67	1	9
G	2	0	2.78	1.09	9
A	3	1	3	1	9

Sist	Min	1C	2C	3C	Max	LIM	LSM
I	1	2	4	5	5	2.43	5
C	2	2	3	3	4	2.47	3
E	1	2	3	3	5	2.47	3
B	2	3	3	4	5	3	3.52
H	2	3	4	4	4	3.47	4
F	2	3	3	4	4	3	3.52
D	1	2	3	3	4	2.47	3
G	2	2	2	3	5	2	2.52
A	1	3	3	4	4	3	3.52

	I	C	E	B	H	F	D	G	A
I		1	0	0	0	0	0	0	0
C	1		0	0	0	0	0	0	0
E	0	0		0	0	0	0	0	0
B	0	0	0		0	0	0	0	0
H	0	0	0	0		0	0	0	0
F	0	0	0	0	0		0	0	0
D	0	0	0	0	0	0		0	0
G	0	0	0	0	0	0	0		0
A	0	0	0	0	0	0	0	0	

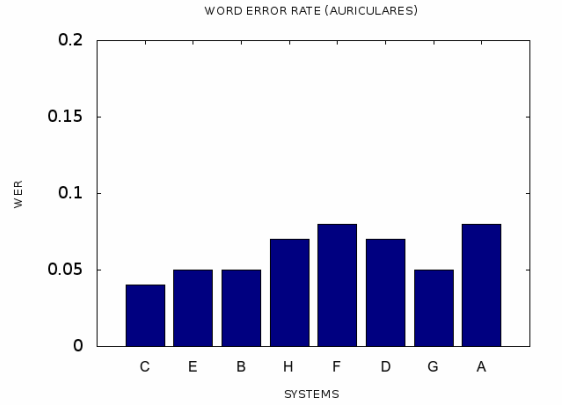
INTELIGIBILIDAD (Sección 4)

- Inteligibilidad - Todos los oyentes



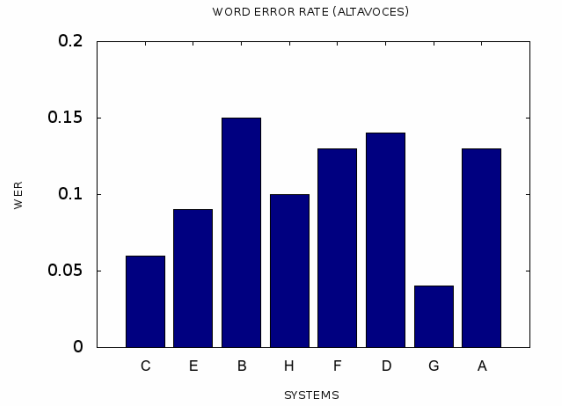
Sist	WER	Samples	Words	S	I	D
C	0.04	206	1352	37	5	16
E	0.06	206	1352	49	4	22
B	0.07	206	1352	66	17	15
H	0.08	206	1352	78	9	18
F	0.09	206	1352	72	30	25
D	0.08	206	1352	77	22	15
G	0.05	206	1352	40	6	20
A	0.09	206	1352	86	6	32

- Inteligibilidad - Auriculares



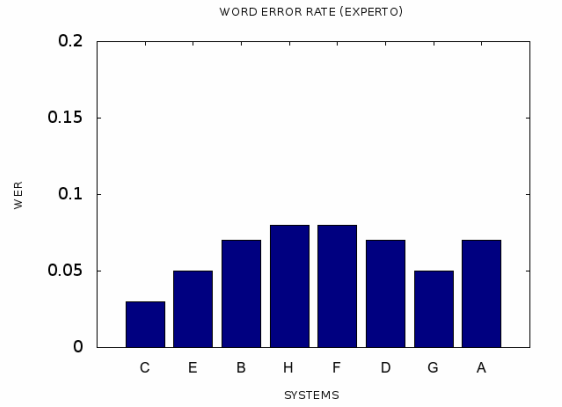
Sist	WER	Samples	Words	S	I	D
C	0.04	162	1062	30	1	9
E	0.05	162	1064	30	2	17
B	0.05	162	1064	41	7	6
H	0.07	162	1065	58	8	11
F	0.08	162	1062	50	24	15
D	0.07	162	1064	54	14	6
G	0.05	162	1060	31	6	16
A	0.08	162	1064	63	3	20

- Inteligibilidad - Altavoces



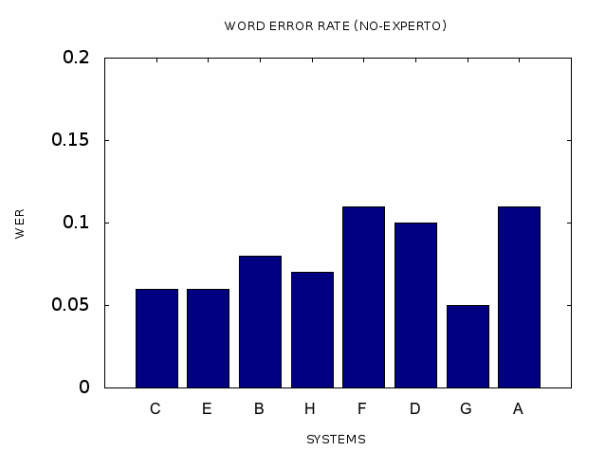
Sist	WER	Samples	Words	S	I	D
C	0.06	44	290	7	4	7
E	0.09	44	288	19	2	5
B	0.15	44	288	25	10	9
H	0.1	44	287	20	1	7
F	0.13	44	289	22	6	10
D	0.14	44	288	23	8	9
G	0.04	44	292	9	0	4
A	0.13	44	288	23	3	12

- Inteligibilidad - Expertos



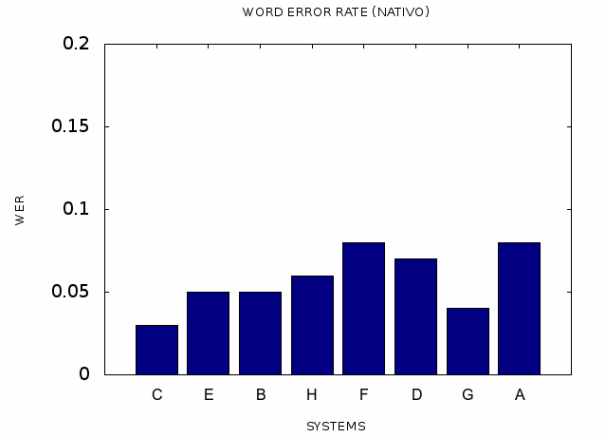
Sist	WER	Samples	Words	S	I	D
C	0.03	108	708	16	4	2
E	0.05	108	708	22	1	15
B	0.07	108	707	28	11	7
H	0.08	108	708	43	5	9
F	0.08	108	708	29	18	10
D	0.07	108	710	31	9	11
G	0.05	108	711	18	1	13
A	0.07	108	710	34	2	15

- Inteligibilidad - No-Expertos



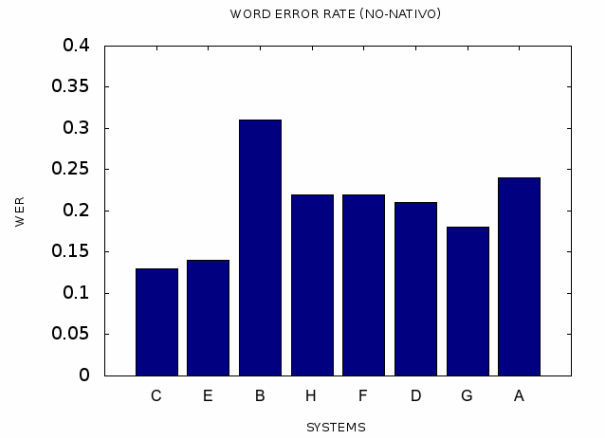
Sist	WER	Samples	Words	S	I	D
C	0.06	98	644	21	1	14
E	0.06	98	644	27	3	7
B	0.08	98	645	38	6	8
H	0.07	98	644	35	4	9
F	0.11	98	643	43	12	15
D	0.1	98	642	46	13	4
G	0.05	98	641	22	5	7
A	0.11	98	642	52	4	17

- Inteligibilidad - Nativos



Sist	WER	Samples	Words	S	I	D
C	0.03	188	1233	29	5	9
E	0.05	188	1234	40	4	15
B	0.05	188	1233	47	13	1
H	0.06	188	1234	68	8	3
F	0.08	188	1233	62	29	10
D	0.07	188	1235	61	20	8
G	0.04	188	1234	30	3	12
A	0.08	188	1234	70	4	22

- Inteligibilidad - No-Nativos



Sist	WER	Samples	Words	S	I	D
C	0.13	18	119	8	0	7
E	0.14	18	118	9	0	7
B	0.31	18	119	19	4	14
H	0.22	18	118	10	1	15
F	0.22	18	118	10	1	15
D	0.21	18	117	16	2	7
G	0.18	18	118	10	3	8
A	0.24	18	118	16	2	10