

Marrero, V., López Bascuas, L. E., & Martín Fernández, J. L. (2015). El ritmo lingüístico en la caracterización del locutor. Percepción y aplicaciones judiciales. Perspectivas actuales en el análisis fónico del habla. *Tradición y avances en la fonética experimental. Normas. Revista de Estudios Lingüísticos Hispánicos (Monografías científicas)*, 7, 335-346.

http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE_CIFE14_ritmo.pdf

El ritmo lingüístico en la caracterización del locutor. Percepción y aplicaciones judiciales

MARRERO AGUIAR, VICTORIA; LÓPEZ BASCUAS, LUIS ENRIQUE; MARTÍN FERNÁNDEZ, JOSÉ LUIS

1 Resumen¹

Presentamos en este trabajo los primeros resultados de un análisis sobre la percepción de las diferencias rítmicas y sus variaciones inter- e intralocutor. Hemos analizado, en función de las ocho métricas del ritmo más conocidas (%V, ΔV , ΔC , VarcoC, VarcoV, rPVI, nPVI-V y nPVI-C; Ramus, Nespors & Mehler, 1999; Low, Grabe & Nolan, 2000; White & Mattys, 2007, entre otros) tres sesiones de lectura de un mismo texto en 12 locutores de la base de datos AHUMADA (Ortega-García, González-Rodríguez & Marrero-Aguiar, 2000); se han seleccionado contrastes máximos entre hablantes y entre emisiones de un mismo hablante, controlando estrictamente el contenido segmental, las variaciones dialectales y la tasa de habla. Los estímulos han sido manipulados para privarlos de cualquier clave distinta a la estrictamente temporal (deslexicalización y asimilación de frecuencias e intensidades). Han sido presentados en una prueba de discriminación (paradigma 2IAX) a 10 jueces. Según nuestros resultados, las diferencias en métricas básicas como %V permitirían diferenciar emisiones de diferentes locutores, y no presentan variabilidad perceptible entre las distintas emisiones de un mismo locutor.

¹ Este trabajo se ha desarrollado en el marco del proyecto Estudio acústico y perceptivo de la variación prosódica inter e intralocutor en español- VILE-P (FFI2010-21690-Co2-02) del Ministerio de Ciencia e Innovación.

2 Introducción

El estudio sobre las características individuales de la voz y sus aplicaciones forenses tradicionalmente no ha abordado el análisis de rasgos del dominio temporal, entre los que se cuenta el ritmo lingüístico, en comparación con la atención recibida por los pertenecientes al dominio de las frecuencias; entre los primeros, además, los elementos más asiduamente mencionados en la bibliografía han sido los relacionados con la tasa de habla y la proporción entre tiempo de habla y silencio (Battaner *et al.*, 2003). Sin embargo, muy recientemente (Leeman, Kolly & Dellwo, 2014) se ha empezado a considerar también el papel del ritmo en la caracterización individual del locutor; el ritmo lingüístico podría ofrecer la misma estabilidad intraindividual que otras conductas motoras, como los movimientos de los dedos o las piernas, y por lo tanto convertirse en un rasgo de gran interés en el ámbito judicial (Eriksson & Wretling, 1997, 1046).

2.1 El concepto de ritmo lingüístico

En este trabajo consideramos el ritmo lingüístico como la percepción de una sucesión regular de eventos con una estructura reconocible (en términos de duración o prominencia). Aunque esa regularidad física subyacente pueda ser más o menos mensurable, la percepción del ritmo no se relaciona directa o automáticamente con ningún evento acústico, sino que es una construcción mental resultado de una interpretación subjetiva de la señal: “es la mente humana la que percibe que ciertas claves físicas constituyen patrones rítmicos” (Auer, Couper-Kuhlen & Müller, 1999: 23).

En lingüística, tradicionalmente se ha identificado la ritmicidad lingüística con la isocronía; en principio, no necesariamente deben ir unidos ambos conceptos (en música existen ritmos anisócronos, los sincopados), pero lo cierto es que la investigación sobre ritmo ha estado condicionada por la búsqueda de la isocronía (cf. 2.2). En ocasiones se observa en la bibliografía otro paralelismo que, este sí, es indudablemente erróneo: el ritmo es un fenómeno diferente al de la tasa de habla o velocidad de elocución; no obstante, la segunda afecta al primero (Dellwo & Wagner, 2003), por lo que la velocidad de elocución ha de ser una variable controlada en las investigaciones sobre ritmo.

2.2 Los estudios sobre el ritmo lingüístico

Los estudios sobre el ritmo en las lenguas han seguido, durante la segunda mitad del siglo XX, la estela de la metáfora propuesta por Llyod James, en 1940, al diferenciar el ritmo tipo *morse*, propio del inglés y otras lenguas germánicas, del ritmo tipo *ametralladora* que observaba en las lenguas románicas; en cada grupo habría elementos diferentes que aparecen a intervalos regulares: el acento en el primero y las sílabas en el segundo. Pike, en 1945, reformula una propuesta (ampliada por muchos autores posteriores, como Abercrombie (1967), que la generalizó a las lenguas del mundo, o Ladefoged) en la que se establecen, pues, dos grandes categorías para el ritmo lingüístico: las lenguas de ritmo silábico (*syllable-timed*), caracterizadas por mostrar sílabas isócronas (típicas, se indicaba, de las lenguas románicas); y las de ritmo acentual (*stress-timed*), en las que la isocronía se da solo entre sílabas tónicas, de modo que la unidad rítmica es el pie; se consideró una características de las lenguas

germánicas. Posteriormente se propuso una tercera categoría, la de las lenguas de ritmo moraico, en las que la regularidad de intervalos se da entre moras (así ocurriría en japonés y tamil). Paralelamente se observaba una relación entre la complejidad de la estructura silábica y el tipo de ritmo: las reglas fonotácticas de las lenguas de ritmo acentual típicamente permiten más complejidad en ataques y codas silábicas; además tienen las vocales tónicas significativamente más largas que las átonas, típicamente reducidas; las lenguas de ritmo silábico tienden más a la estructura CV y tienen preponderancia de sílabas abiertas; la percepción del ritmo, por lo tanto, podría emerger de contrastes en la estructura silábica como mecanismo general.

A medida que aumentaban los estudios sobre emisiones reales en diversas lenguas, empezaron a surgir críticas respecto a esta visión; así, Dauer, por ejemplo, en 1983, observó que ni las sílabas mantenían una duración estable en lenguas de supuesto ritmo silábico, ni los intervalos interacentuales en las de ritmo acentual. También Borzone de Manrique y Signorini (1983) encontraron que en el español, considerado de ritmo silábico, la duración de las sílabas no era constante, y en cambio los intervalos interacentuales sí tendían a presentar una duración similar. Tampoco Wenk & Wiolland (1982) encontraron sílabas isócronas en francés, por lo que propusieron la existencia de unidades rítmicas más largas (la frase fonológica). Las moras, por último, tienen en japonés diferentes duraciones (Otake *et al.*, 1993). En definitiva, aunque se demostró que la hipótesis de la isocronía no encuentra soporte acústico, la idea de una tipología rítmica para las lenguas ha mantenido su vigencia.

A final de siglo, la psicolingüística evolutiva también hizo suya la propuesta dicotómica, y le dio un nuevo soporte: Mehler, Dupoux, Nazzi & Dehaene-Lambertz (1996) se basaron en ella para explicar cómo los niños pueden aprender aspectos de la fonología de la lengua materna, puesto que el ritmo les permite segmentar el continuo fónico, atendiendo a diferentes indicios, en un tipo de lengua u otro.

Pero el trabajo que marcó una nueva etapa en los estudios sobre ritmo lingüístico fue el de Ramus, Nespors & Mehler (1999): en él se defiende que "a simple segmentation of speech into consonants and vowels can:

- Account for the standard stress- / syllable-timing dichotomy and investigate the possibility of other types of rhythm;
- account for language discrimination behaviors observed in infants;
- clarify how rhythm might be extracted from the speech signal" (Ramus, Nespors & Mehler, 1999: 270-271).

Las fórmulas propuestas son las siguientes:

- %V: proporción de intervalos vocálicos sobre la duración total de un texto; se relaciona con la reducción vocálica en sílabas átonas y la complejidad de ataques y codas silábicas; por eso es más baja en las lenguas germánicas (cuyas vocales átonas se reducen, y cuya estructura silábica contiene más consonantes) y más alta en las romances (donde las vocales átonas apenas se acortan y cuyas sílabas son más sencillas).
- ΔV : desviación estándar de la duración de los intervalos vocálicos en un texto; se ve afectada por la existencia de oposición fonológica entre vocales largas y breves y por los fenómenos de alargamiento vocálico contextual.
- ΔC : desviación estándar de la duración de los intervalos consonánticos en un texto.

Una nueva variable aparece cuando Ramus (2002) introduce una tasa de normalización de la velocidad de habla: VarcoC, la desviación estándar de la duración de cada intervalo consonántico dividida entre la media de la duración

¹ Barry, Adreeva, Russo, Dimitrova & Kostadinova (2003) demostraron que ΔV y ΔC son inversamente proporcionales a la tasa de habla.

de todos ellos. Dellwo (2006) encontró que VarcoC resultó mejor que ΔC para distinguir inglés y alemán del francés, en todas las tasas de habla. El correlato de VarcoC para los intervalos vocálicos es VarcoV.

Low, Grabe & Nolan (2000) desarrollan otro argumento en este ámbito: el elemento esencial para capturar el ritmo es su naturaleza secuencial; para poder cuantificarla proponen usar el índice normalizado de variabilidad emparejada (*Normalised Pairwise Variability Index*) PVI, dividiendo la diferencia entre pares de intervalos vocálicos entre la suma de los intervalos. Luego Grabe & Low (2002) propusieron “a non-rate normalised PVI” para los intervalos consonánticos, porque la normalización podría enmascarar una variación rítmicamente relevante en la estructura de ataques y codas silábicos.

En los años siguientes aparecen nuevas fórmulas con matices destinados a captar la complejidad del ritmo lingüístico. Entre ellas, el *Compensation and Control Index* (CCI), de Bertinetto & Bertini (2008), definido como “un índice fonológicamente motivado”: toma como base el rPVI y lo divide entre el número de segmentos fonológicos del texto (de la definición de estos últimos surge el carácter fonológico de la propuesta). Otra de las fórmulas recientes es %VO (Dellwo, Fourcin & Abberton, 2007), el porcentaje de segmentos sonoros de un texto, que considera prioritario el rasgo de sonoridad sobre el carácter consonántico o vocálico de un sonido.

En conclusión, aunque no podemos descartar la presencia de otras variables importantes en la codificación del ritmo en las lenguas (Prieto *et al.* [2012] resaltan el papel de la prominencia de frase, y las fronteras prosódicas), las diferencias de duración han demostrado jugar un papel importante, y las llamadas *métricas* del ritmo se han convertido en un instrumento generalizado para analizarlas. Considerando que los rasgos temporales son más resistentes que los frecuenciales ante las distorsiones generadas por ruido ambiente o por el filtrado telefónico (condiciones muy frecuentes en las grabaciones que se suelen presentar en el ámbito judicial), su aplicabilidad en el entorno de la llamada *fonética forense* constituirá el objetivo de los dos apartados siguientes.

2.3 ¿Es el ritmo un rasgo individual, caracterizador del hablante?

Existen numerosas fuentes de variabilidad rítmica. La mayor parte de la bibliografía se ha centrado en la que tiene su origen en la diversidad tipológica de las lenguas, pero también contamos con trabajos que demuestran las diferencias de ritmo entre diversas variedades geolectales de algunas de ellas (principalmente el inglés; para dialectos del español cf. O'Rourke 2008; Toledo, 2010). Se ha mencionado ya la influencia de la tasa de habla en la variabilidad del ritmo (Dellwo & Wagner, 2003); factores estilísticos también provocan fluctuaciones (por ejemplo, Payne *et al.*, 2010 encuentran diferencias rítmicas entre el habla entre adultos y la dirigida a niños); e incluso el contenido de la muestra: Widget *et al.* (2010) encuentran que las desviaciones entre diferentes locutores son menores que las causadas por las distintas oraciones del texto analizado.

Otro tipo de variabilidad más relevante, en el contexto de este trabajo, es el que tiene su origen en las diferencias entre el habla de distintos locutores (variabilidad interlocutor) y las emisiones de una misma persona (variabilidad intralocutor). Yoon (2010) encuentra que las PVIs son las métricas más estables entre varios hablantes (por lo tanto, las que menos interés revisten para

aplicaciones judiciales, donde lo que se busca son parámetros que caractericen al individuo). Mairano (2011: 106) encuentra, por el contrario, gran variabilidad en PVI entre locutores, mientras que el CCI resultaría más estable. En el trabajo de Loukina *et al.* (2011) los locutores tenían la posibilidad de repetir la lectura de los textos que se les presentaban: un 15 % de ellos fue repetido, y al comparar ambas emisiones se observó que la repetición “clearly did not affect the values of rhyhtm metrics” (Loukina *et al.*, 2011: 3262). Este resultado apunta, por tanto, a una baja variabilidad intralocutor en el ritmo lingüístico, lo cual es el objetivo de los rasgos válidos para la caracterización individual con fines judiciales.

2.4 Estudios de ritmo con finalidad forense

La lógica subyacente a la aplicabilidad del ritmo lingüístico en el ámbito forense está perfectamente expuesta en la siguiente cita de Dellwo, Leeman & Kolly (2012: 1584):

Humans have highly individual ways in which they move their legs when walking and that identification of individuals based on these movements is well possible. [...] We argue that speech is similar to walking in that it is a highly complex brain operated series of muscle movements which may be carried out to a considerable degree in individual ways [...] There is increasing evidence revealing that rhythm measures like %V, deltaV or the PVI may vary significantly between speakers.

En Leeman, Kolly y Dellwo (2014) se analizaron las *métricas* del ritmo anteriormente mencionadas, además de los intervalos entre picos silábicos (*syllable-peak-to-syllable-peak interval durations*, una medida basada en el concepto de *centros perceptivos*, Fowler 1979), en tres tipos de corpus, grabados con los mismos 16 sujetos (8 hombres y 8 mujeres, aunque no se encontraron efectos de esta diferencia): habla espontánea por micrófono, por teléfono y lectura por micrófono. Según sus resultados, %V y %VO permitieron distinguir entre locutores, tanto en habla espontánea como en lectura, sin presentar diferencias entre ambos estilos ni entre diferentes emisiones del mismo locutor, tal como requiere la comparación judicial de voces (Leeman, Kolly & Dellwo, 2014: 65).

Entre los temas que requieren, según los autores, investigación adicional, se menciona la necesidad de llevar a cabo estudios perceptivos: si los rasgos que caracterizan el ritmo lingüístico de una persona no resultan claramente perceptibles tampoco podrán ser fácilmente disimulados. Con el fin de contribuir al avance de conocimientos en este sentido hemos desarrollado este trabajo

3 Objetivos e hipótesis

El objetivo de este estudio es valorar la aplicabilidad de la caracterización rítmica del locutor en el ámbito judicial seleccionando las métricas más sensibles a la variación interlocutor y menos sensibles a la variación intralocutor, mediante un experimento diseñado con el mayor control posible de la variabilidad rítmica, excepto la dependiente del locutor; con ese fin se ha controlado la variabilidad lingüística y geolectal (todas las muestras pertenecen a la variedad centro-peninsular del español), el estilo de habla (lectura), el contenido segmental (idéntico para todos los locutores, y además deslexicalizado), la velocidad de habla (solo se consideran muestras con la

misma duración), la intensidad y la frecuencia fundamental (manipuladas para igualarlas).

Nuestras hipótesis principales son dos:

- Los oyentes serán capaces de discriminar dos emisiones basándose exclusivamente en diferencias de ritmo entre ellas.
- Los estímulos que se diferencian por duración vocálica serán los que obtengan mejores tasas de discriminación entre locutores y los menos afectados por la variación intralocutor.

4 Metodología

a) Los parámetros analizados corresponden a ocho de las *métricas* del ritmo más utilizadas en la bibliografía:

- %V: proporción de intervalos vocálicos sobre la duración total de la secuencia
- ΔC : desviación estándar de las duraciones de todos los intervalos consonánticos en la secuencia
- ΔV : desviación estándar de las duraciones de todos los intervalos vocálicos en la secuencia
- VarcoC: cambios relativos en ΔC considerando la tasa de habla
- VarcoV: cambios relativos en ΔV considerando la tasa de habla
- rPVI-C (*Pairwise variability index, non-rate normalised*: diferencias de duración entre un intervalo consonántico y el siguiente, dividido por el número de intervalos de la secuencia)
- nPVI-C: similar a rPVI pero considerando la tasa de habla
- nPVI-V: similar a nPVI-C pero considerando intervalos vocálicos y no consonánticos

Estímulos. Se han extraído de la base de datos AHUMADA (Ortega, González & Marrero, 2000), especialmente diseñada con fines forenses, datos de un texto equilibrado fonéticamente, leído por 12 locutores masculinos y grabados en tres sesiones de cada uno (por tanto, 36 emisiones de cada estructura silábica, 216 realizaciones). Aunque disponemos de secuencias de habla espontánea etiquetadas, para esta primera etapa se ha optado por la tarea de lectura de texto, que permitía controlar de forma precisa la variable contenido segmental. Se han seleccionado dos secuencias, una representante de la estructura silábica más simple, CV (“te toca la cara”), y la otra con mayor complejidad estructural, con predominio de CVC (“buscando los campos”). Se han controlado la aparición de rasgos geolectales en las secuencias y la tasa de habla, para evitar que las diferencias percibidas puedan residir en una mayor o menor velocidad de elocución (diferencias menores a las mínimas perceptibles, según Martín Fernández, López Bascuas & Marrero Aguiar, 2014). Los pares seleccionados fueron manipulados mediante un *script* para Praat elaborado Lahoz (2012) y resintetizados mediante Mbrola (Dutoit *et al.*, 1996), con el fin de convertir todas las consonantes en /s/ y todas las vocales en /a/; en principio, este método iguala las diferencias en F0 e intensidad, pero como medida de control adicional, utilizamos una aplicación de Praat para trasladar el perfil tonal de un miembro de la pareja al otro, y garantizar así que el f_0 no intervenía como variable. De los datos acústicos (extraídos durante una etapa previa del proyecto VILE-P), se han seleccionado las parejas que ofrecían mayor contraste en cada una de las métricas entre dos locutores diferentes y entre dos sesiones de un mismo locutor. Ha habido ocasiones en que el mayor contraste en una pareja coincidía entre diferentes métricas; así ocurrió, por ejemplo, con ΔC , VarcoC, rPvi y nPviC: en los cuatro casos, fue el locutor núm. 13, en sus

grabaciones 1 y 3, el que presentó los mayores contrastes. En la tabla 1 pueden verse todos los casos en que se dio esta circunstancia.

Tarea: finalmente la prueba contenía 20 parejas de estímulos (a, b) de sílabas tipo *sasasa*, en cuatro agrupaciones ($a-a, a-b, b-a, b-b$), que se presentaron repetidas ocho veces en orden aleatorio (pero en dos bloques separados, uno para la secuencia CV y otro para CVC), y por tanto constaba de 640 *ensayos*. Los sujetos debían indicar si se trataba de estímulos iguales o diferentes (tarea de discriminación con elección forzada). De acuerdo con la Teoría de Detección de Señales (Green & Swets, 1966), cuando los estímulos presentados son iguales, y el sujeto responde que lo son, nos encontramos ante un *rechazo correcto*; cuando en esta condición el sujeto responde que se trata de estímulos distintos, aparece una *falsa alarma*; por el contrario, cuando los estímulos son diferentes, se obtiene un *acierto* cuando el sujeto responde que lo son, y un *error* cuando los considera estímulos iguales.

Sujetos: 10 estudiantes universitarios, cuatro hombres y seis mujeres, sin problemas auditivos conocidos, sin conocimientos musicales avanzados y monolingües. Los sujetos no tenían experiencia previa en la tarea, pero contaron con una pequeña fase de entrenamiento. La prueba duró unos 45-50 minutos, y los sujetos recibieron una pequeña compensación económica por su participación.

5 Resultados

Se han obtenido 6368 respuestas (3184 corresponden a parejas iguales y 3184 a parejas distintas). La tasa global de identificación correcta (sumando aciertos y rechazos correctos) fue del 69 %; cuando los estímulos eran iguales, los rechazos correctos alcanzaron el 84 % de las presentaciones, pero cuando eran diferentes, solo un 54 % de las veces se identificaron como tales.

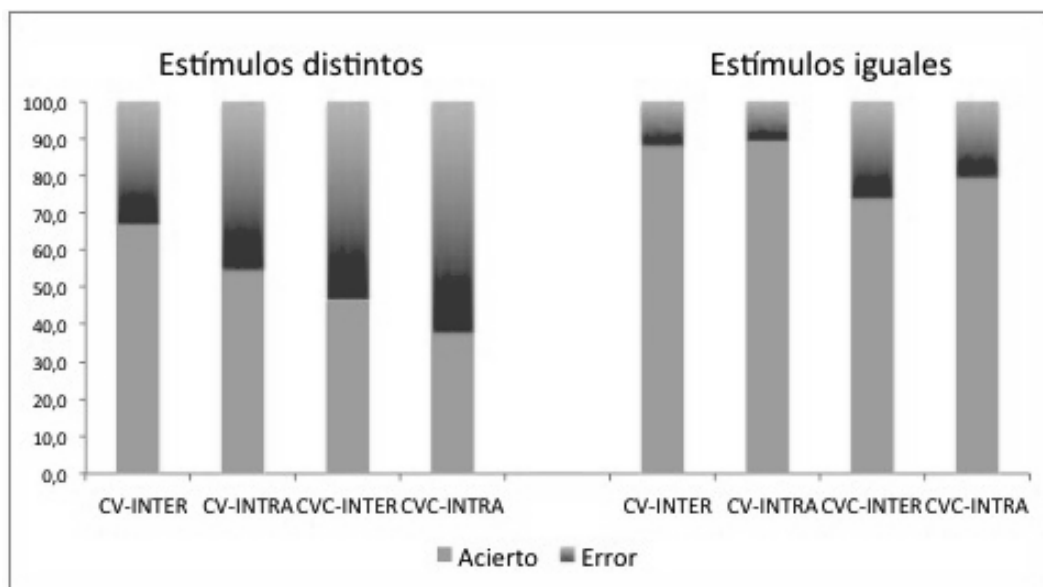


Figura 1. Resultados de estímulos CV y CVC en parejas iguales y diferentes de un mismo locutor (INTRA) y dos locutores distintos (INTER)

Como se comprueba en la figura 1, los resultados con secuencias CV son siempre mejores que con CVC, tanto en estímulos iguales como diferentes, y tanto en parejas de estímulos procedentes del mismo locutor como de distintos locutores; esta estructura presenta los resultados esperados: tasas de discriminación cercanas o superiores al 70 % para estímulos distintos procedentes de distintos locutores, y cercanas al 50 % (acierto por azar) para las que proceden del mismo locutor. En cambio, los estímulos obtenidos con

secuencias CVC presentan tasas de acierto inferiores al 50 % en todos los casos, e incluso una mejor discriminación de las diferencias intralocutor que las interlocutor en estímulos iguales. La prueba T de Student indica que cuando los estímulos son iguales las diferencias inter- e intralocutor no son significativas ni con estímulos CV ($p = 0.27$) ni con estímulos CVC ($p = 0.08$); por el contrario, ante estímulos distintos los resultados en emisiones de un mismo locutor y de dos hablantes diferentes resultaron estadísticamente significativas, tanto en estímulos CV ($p = 0.033$) como CVC ($p = 0.027$).

Para determinar cuáles son las *métricas* más adecuadas en el ámbito judicial nos centraremos solo en los resultados con estímulos diferentes, considerando tanto los aciertos en parejas de locutores distintos (INTER) como en diferentes emisiones del mismo hablante (INTRA).

CV			
INTERLOCUTOR	%	INTRALOCUTOR	%
%V	93,75	%V	26,25
ΔC	41,25	$\Delta C = \text{VarcoC} = rPvi = nPviC$	66,25
ΔV	84,38	ΔV	55,13
rPvi	51,88		
nPvi	39,10	nPvi	58,13
VarcoV	75,00	VarcoV	66,88
VarcoC=nPviC	81,88		
CVC			
%V	61,25	%V	45,00
$\Delta C = \text{VarcoC}$	59,38	$\Delta C = \text{VarcoC}$	39,74
$\Delta V = \text{VarcoV} = nPvi = nPviC$	36,54	$\Delta V = \text{VarcoV} = nPvi$	39,38
rPvi	29,38	$rPvi = nPviC$	26,25

Tabla 1. Porcentajes de acierto y medida de sensibilidad (d') en estímulos distintos según las diferentes métricas utilizadas

Las pruebas estadísticas de significatividad de las diferencias (ANOVA) en estímulos CV interlocutor indican que no hay diferencias significativas entre %V, ΔV , $\text{VarcoC} = nPviC$ y VarcoV ; ni entre rPVI, ΔC y nPVI, aunque sí entre ambos grupos ($p < 0.001$). En estímulos CV intralocutor solo %V se diferencia significativamente de todas las demás *métricas*. ($p < 0.001$). En cuanto a los estímulos CVC, en emisiones de locutores diferentes se observan diferencias significativas entre %V- $\Delta C = \text{VarcoC}$ y las demás métricas; en emisiones de un mismo locutor no hay diferencias en ninguna de las métricas.

Las diferencias en %V son las que mejor responden a las hipótesis planteadas: resultan claramente perceptibles cuando corresponden a diferencias entre locutores y muy difíciles de discriminar cuando proceden del mismo locutor. El resto de las parejas presentan una perceptibilidad baja en estímulos CVC, aunque algunas (ΔV , VarcoV y $\text{VarcoC} = nPVI - C$) se distinguieron bien en estímulos CV de diferentes locutores. Las demás métricas presentan una baja discriminabilidad tanto cuando los locutores son diferentes como cuando es el mismo en diferentes emisiones.

Por lo tanto, nuestros resultados conducen a afirmar que las diferencias de ritmo en condiciones de máximo control de los estímulos son percibidas de

forma consistente por sujetos hispanohablantes cuando la estructura silábica es simple (CV), con tasas de discriminación correcta superiores al 80 % en métricas como %V, ΔV o VarcoC-nPVI-C, pero fue difícil con estructuras más complejas (en estímulos CVC esas tasas no subieron del 66 %). En general, las diferencias entre locutores fueron mejor percibidas que los contrastes intralocutor, aunque la diferencia entre las dos estructuras silábicas condiciona este resultado.

6 Conclusiones

Si retomamos las hipótesis planteadas al comienzo, podemos concluir que algunas de las diferencias de ritmo estudiadas han resultado claramente perceptibles contando única y exclusivamente con información temporal en la señal (se confirma, por tanto, nuestra primera hipótesis). En concreto, %V, una medida básica de duración vocálica, ha presentado el comportamiento esperado: muy alta discriminabilidad entre locutores diferentes y muy baja entre distintas emisiones de un mismo locutor (se confirmaría así también nuestra segunda hipótesis).

Algunos de los demás resultados obtenidos eran esperables: las métricas formuladas para compensar las diferencias en tasa de habla (VarcoC/V y nPviC/V) no resultan eficaces con nuestra muestra, cuya tasa de habla había sido controlada previamente.

No obstante, algunos resultados han sido inesperados. Esencialmente, el comportamiento de los estímulos contruidos con predominio de secuencias CVC, contrario al previsto: una discriminación inferior al 50 % (acierto por azar) cuando los estímulos eran diferentes; y mayor discriminación para las diferencias intralocutor que para las interlocutor en algunas condiciones. Creemos que la explicación puede residir en el tratamiento experimental que han sufrido los segmentos consonánticos sonoros de la secuencia elegida ("buscando los campos"). Con la manipulación descrita en la metodología (la habitual en estudios de percepción del ritmo), se han convertido en [s], igual que las consonantes sordas. Se pierde así el rasgo de sonoridad, que según se ha demostrado en investigaciones anteriores (Leeman, Kolly & Dellwo, 2014; Dellwo, Leeman & Kolly, 2012) puede jugar un papel muy relevante en la codificación del ritmo lingüístico.

Por lo tanto, entre los próximos pasos de nuestra investigación se incluirá %VO entre las métricas analizadas, y se modificará la manipulación de los estímulos para su deslexicalización, con el fin de mantener el rasgo sonoro de las consonantes que lo presenten. Además, se corroborará en los datos acústicos correspondientes al texto completo de los 24 locutores analizados en VILE-P que la variabilidad rítmica en %V y %VO es baja entre distintas emisiones de un mismo hablante y alta entre emisiones de hablantes diferentes, y se realizará un nuevo experimento con mayor número de sujetos.

7 Bibliografía

- Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time: The rhythm and tempo of spoken interaction*. New York: Oxford University Press.
- Battaner, E., Gil, J., Marrero, V., Llisterri, J., Carbó, C., Machuca, MJ, de-la Mota & Ríos, A. (2003). VILE: Estudio acústico de la variación inter- e intralocutor en español. En *SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense* Barcelona, SEAF, 59-70.
- Bertinetto, P. M., & Bertini, C. (2008). On modeling the rhythm of natural languages. In *Proceedings of the Fourth International Conference on Speech Prosody*. Recuperado el 12 del 11 de 2014, de <http://linguistica.sns.it/QLL/QLL0>

- 7_08/Bertinetto_Bertini_campinas.PDF
- Borzone de Manrique, A.M. y Signorini, A. (1983): Segmental duration and rhythm in Spanish. *Journal of Phonetics*, 11 (2), 117-128.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*, 11, 51-62.
- Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate. *15th ICPHS, Barcelona*. Recuperado el 11 del 12 de 2014 de <http://pub.uni-bielefeld.de/luur/download?func=downloadFile&recordId=1785384&fileId=2352642>
- Dellwo, V., Fourcin, A. & E. Abberton (2007). Rhythmical classification based on voice parameters, En: *International Conference of Phonetic Sciences (ICPhS)*, Saarbrücken, 1129-1132.
- Dellwo, V., Leemann, A., & Kolly, M. J. (2012). Speaker idiosyncratic rhythmic features in the speech signal. In *INTERSPEECH 2012*. Recuperado el 12 del 11 de 2014, de http://20.21019352.unknown.qala.com.sg/archive/archive_papers/interspeech_2012/i12_1584.pdf
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 3, pp. 1393-1396). IEEE.
- Dutoit, T.; Pagel, V.; Pierret, N.; Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings of the Fourth International Conference on Spoken Language (ICSLP'96)*. Philadelphia. IEEE: 1393-1396.
- Eriksson, A., & P. Wretling (1997): How Flexible is the Human Voice? A Case Study of Mimicry. *Proceedings of the European Conference on Speech Technology, Eurospeech 1997*, 1043-1046.
- Fowler, A. (1979). Perceptual centers in speech production and perception, *Perception and Psychophysics* 25, 375-388.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, 7 (515-546).
- Green, D. M., & Swets, J. A. (1966): *Signal detection theory and psychophysics*. Oxford, Robert E. Krieger. Reimpreso en 1974.
- Lahoz, J. M.^a (2012). Syllable, accent, rhythm: typological and methodological considerations for teaching Spanish as a foreign language. *Revista Internacional de Lenguas Extranjeras*, (1), 129-150. Recuperado el 12 del 11 de 2014, de <http://www.publicacionsurv.cat/ojs/revistes/index.php/rile/article/viewFile/13/15>
- Lahoz, J. M.^a (2012). La enseñanza de la entonación, el ritmo y el tempo. En J. Gil Fernández, (ed.). *Aproximación a la enseñanza de la pronunciación en el aula de español*. Madrid. Edinumen, 93-132.
- Leemann, A., Kolly, M. J., & Dellwo, V. (2014). Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic science international*, 238, 59-67.
- Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and speech*, 43(4), 377-401.
- Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *The Journal of the Acoustical Society of America*, 129 (5), 3258-3270.
- Low, L. E., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and Speech*, 43(4), 377-401.
- Mairano, P. (2011). *Rhythm typology: acoustic and perceptive studies* (Doctoral dissertation, Università degli studi di Torino). Recuperado el 12 del 11 de 2014, de http://halshs.archives-ouvertes.fr/docs/00/65/42/61/PDF/MairanoPaolo_These.pdf
- Martín Fernández, J. L., López Bascuas, L. E., & Marrero Aguiar, V. (2014). Percepción del tempo en estímulos lingüísticos de diferente complejidad. VI CIFE, Valencia (comunicación).
- O'Rourke, E. 2008. Correlating speech rhythm in Spanish: Evidence from

- two Peruvian dialects. *Selected Proceedings of the 10th Hispanic Linguistics Symposium*. J. Bruhn de Garavito y E. Valenzuela. Somerville (eds.). Massachusetts: Cascadilla Proceedings Project. 276-287
- Ortega-Garcia, J., Gonzalez-Rodriguez, J., & Marrero-Aguiar, V. (2000). AHUMADA: A large speech corpus in Spanish for speaker characterization and identification. *Speech communication*, 31(2), 255-264.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258-278.
- Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. D. M. (2010). A cross-linguistic study of prosodic lengthening in child-directed speech. Prieto *et al.*, 2012
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. *Linguistic controversies*, 73-79.
- Toledo, G. (2010). Métricas rítmicas en tres dialectos Amper-España. *Estudios Filológicos* 45, 93-110.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501-522.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm?. *The Journal of the Acoustical Society of America*, 127(3), 1559-1569.
- Wretling, P., & Eriksson, A. (1998). Is articulatory timing speaker specific? Evidence from imitated voices. En *Proc. FONETIK* (vol. 98, 48-52). Recuperado el 12 del 11 de 2014, de <http://www.ling.gu.se/~anders/papers/imit98.pdf>
- Yoon, T. J. (2010). Capturing inter-speaker invariance using statistical measures of rhythm. In *Proceedings of speech prosody*. Recuperado el 12 del 11 de 2014, de http://20.21019352.unknown.qala.com.sg/archive/sp2010/papers/sp10_201.pdf