

# Estudio del foco: las prominencias acentuales, el modelado acústico y la detección automática

Jorge Gurlekian

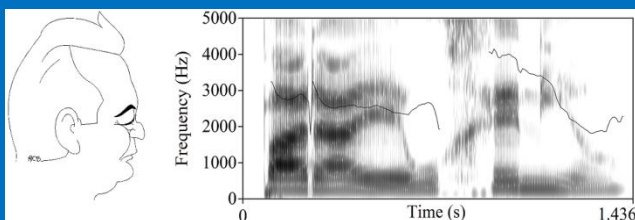
Humberto Torres

Diego Evin

Hansjörg Mixdorff

Christian Cossio-Mercado

Mercedes Güemes



Fernández Planas, A. Ma. (ed.) (2016): *53 reflexiones sobre aspectos de la fonética y otros temas de lingüística*, Barcelona, págs. 209-219.

ISBN: 978-84-608-9830-6.



# Estudio del foco: las prominencias acentuales, el modelado acústico y la detección automática

Jorge Gurlekian  
LIS CONICET Buenos Aires  
[jag@fmed.uba.ar](mailto:jag@fmed.uba.ar)

Humberto Torres  
LIS CONICET Buenos Aires  
[hmtorres@conicet.gov.ar](mailto:hmtorres@conicet.gov.ar)

Diego Evin  
LIS CONICET Buenos Aires  
[diegoevin@gmail.com](mailto:diegoevin@gmail.com)

Hansjörg Mixdorff  
Beuth University Berlin  
[mixdorff@bht-berlin.de](mailto:mixdorff@bht-berlin.de)

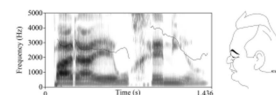
Christian Cossio-Mercado  
LIS CONICET Buenos Aires  
[cgcossio@gmail.com](mailto:cgcossio@gmail.com)

Mercedes Güemes  
LIS CONICET Buenos Aires  
[mercedesguemes@gmail.com](mailto:mercedesguemes@gmail.com)

*Para Eugenio Martínez Celdrán, incansable precursor de la fonética experimental de la lengua española, que con sus iniciativas nos ha permitido compartir conocimientos y realizaciones.*

## 1. INTRODUCCIÓN

Durante la producción de habla, la información de la estructura entonativa de una oración se refleja en el énfasis relativo dado a sus palabras de contenido y en las características tonales de los segmentos finales de cada frase. Dado que las sílabas con acento léxico pueden presentar un aumento en la frecuencia fundamental, energía y duración y, además, una mayor estabilidad espectral, si estas sílabas son enfatizadas en el discurso, algunos o todos estos parámetros pueden ser modificados. Por ejemplo, los acentos tonales altos se integrarán a los rasgos acústicos presentes en la sílaba, aumentando o disminuyendo el tono fundamental. El correlato perceptual de este proceso de énfasis e integración es la prominencia silábica (Fry, 1958). Junto con los aspectos acústicos de los acentos lexicales y tonales coexisten otros factores a nivel segmental y supra-segmental que afectan la percepción de prominencia, tales como la información lingüística y paralingüística. Los aspectos pragmáticos asociados a la información nueva que se transmite, también pueden influenciar los juicios de prominencia. Debido a que la función focal se define como la



selección de un segmento de la oración como el de mayor jerarquía pragmática, por su relevancia en el procesamiento informativo del destinatario (Dik, 1997), su realización necesita de recursos léxico-gramaticales y fónicos. Dado que en español las estructuras sintácticas son flexibles, resulta importante establecer el papel del factor acústico para generar la percepción del foco. A pesar de que la estructura de la información se pueda jerarquizar principalmente a partir del orden de palabras, es la prosodia la que determina la interpretación final frente a estructuras sintácticamente similares. Además, el foco sintáctico se manifiesta en la superficie con una mayor prominencia prosódica y así, permite distinguir entre estructuras marcadas de las no marcadas (Zubizarreta, 1998). La diferencia entre los parámetros prosódicos (entonación, duración, sonoridad, pausa), los sintácticos y los informativos a la hora de marcar el foco radica en que los prosódicos siempre estarán presentes y el foco se da por la naturaleza de su manifestación. El valor del foco prosódico difiere en el tipo de discurso. Como el habla espontánea no está muy organizada ni focalizada, la prosodia cumple un rol fundamental para estructurar el plano sintáctico, es decir, organizar el discurso en unidades menores. A nivel pragmático, la prosodia comunica estados de ánimo e intereses elocutivos. En el nivel de la cláusula se parte de la idea de que cualquier oración es respuesta de una pregunta supuesta, y el foco marcaría el constituyente que motiva la respuesta a esa pregunta implícita. Dentro de este dominio, Ladd (1983) define el foco como una tendencia a acentuar o desacentuar porciones de una oración relacionadas con el significado; así, los autores indican que las oraciones tendrían ítems marcados con los rasgos [-foco] o [+foco] según la intención del hablante. El alcance del foco puede variar: puede abarcar toda la oración (foco neutral o normal) o estar sobre un constituyente como el predicado (foco amplio) o sobre un solo ítem léxico (foco estrecho) (Ladd, 83). El foco puede transmitir varios tipos de intereses comunicativos, agregar información nueva, corregir información precedente, seleccionar un elemento entre otros ofrecidos por el contexto o contrastar información supuesta por el oyente.

Nuestra propuesta consiste en realizar un estudio perceptual del foco y de los niveles de prominencia para vincular las realizaciones de frases con las distintas modalidades y para asociar la posición del foco con el grado de prominencia percibida. Luego, se obtendrán los parámetros del modelo de entonación para esas oraciones y se presentará el grado de asociación con las prominencias percibidas y con el foco. Finalmente, se presentarán los resultados obtenidos con un sistema automático para la detección de prominencias. Este sistema de clasificación utilizará la información acústica tonal con la de los acentos léxicos. Esta búsqueda apunta a integrar el conocimiento de alto nivel en los sistemas de reconocimiento y comprensión automática de lenguaje (Nöth *et ál*, 2001), dado que estos todavía no explotan o hacen un uso limitado de la información prosódica.

## 2. METODOLOGÍA

### 2.1. Oraciones y Sujetos

Se emplearon nueve oraciones declarativas y nueve interrogativas, mediante la combinación en el sujeto y predicado de tres palabras de contenido trisilábicas con acento léxico. El texto de las oraciones es el empleado para el proyecto AMPER Argentina



(Atlas Multimedia de la Prosodia del Espacio Románico), en sus modalidades declarativa e interrogativa absolutas (Gurlekian y Toledo, 2008), y consiste en oraciones del tipo sintáctico SVO (Sujeto-Verbo-Objeto), concretamente SN (sintagma nominal) + SV (sintagma verbal) + SPrep (sintagma preposicional). El SN y el SPrep contienen una palabra de contenido, con variantes paroxítonas, proparoxítonas y oxítonas. El SV tiene una palabra de contenido paroxítona. Diez sujetos oriundos de distintas provincias argentinas (Chaco, Corrientes, Entre Ríos y Buenos Aires), universitarios y sin entrenamiento (5 femeninos y 5 masculinos), participaron en la grabación en una sala no tratada acústicamente con un micrófono dinámico. Los hablantes emitieron tres repeticiones para cada una de las nueve oraciones a una velocidad de habla semi-rápida. Los hablantes fueron instruidos sobre el objetivo de la prueba con la indicación del tipo de foco que debía generar. Los tipos solicitados fueron: sin foco o foco neutro, foco estrecho en la palabra de contenido del sujeto (triángulo, guitarra y saxofón y foco estrecho en la palabra de contenido del predicado (pánico, medida, obsesión). Las emisiones generadas para el estudio perceptual, acústico y de reconocimiento automático fueron 540 (9 oraciones x 3 focos x 2 modalidades x 10 hablantes).

### 3. EVALUACION PERCEPTUAL

El objetivo de la evaluación perceptual es determinar si los oyentes rescatan tanto la información de modalidad como la del foco y obtener los niveles de prominencia prosódica percibida de cada sílaba en las dos modalidades: declarativa e interrogativa. El entrenamiento se realizó mediante un programa interactivo. Para la prueba perceptual la instrucción fue: *Usted puede oprimir la tecla de reproducción las veces que lo desee. Luego debe decidir el tipo de modalidad de la oración entre las opciones «declarativa», «interrogativa» o «ambigua». Luego debe indicar la pregunta o respuesta que mejor se corresponde con la afirmación o la interrogación escuchada: Por último debe posicionar los cursores para indicar el grado de prominencia de cada sílaba.* La tarea se desarrolló en 5 sesiones, con aproximadamente 110 estímulos presentados en forma aleatoria. Siete oyentes participaron en la prueba, todos hablantes nativos de Buenos Aires. La tarea demanda 135 minutos en total (15 seg por oración). El uso de una matriz de cursores deslizables para evaluar la prominencia fue inspirado por Erickson *et ál.* (2001).

#### 3.1. Resultados de la prueba perceptual

##### 3.1.1. Evaluación de la Modalidad

La tabla 1 presenta la matriz de confusión de la modalidad reconocida, indicándose los porcentajes de identificación con relación a la producción pretendida de los locutores.

Modalidad Pretendida	Modalidad Percibida		
	Declarativa	Interrogativa	Ambigua
Declarativa	99,59%	0,29%	0,12%
Interrogativa	1,24%	97,04%	1,72%

Tabla 1. Porcentajes de identificación de la modalidad de la oración.



### 3.1.2. Evaluación del foco

La evaluación del foco percibido se realizó en forma indirecta al analizar la selección por parte de los oyentes de las preguntas asociadas a las oraciones declarativas e interrogativas. Las respuestas se muestran en la tabla 2 donde se indican los porcentajes de asignación de foco en relación a la producción pretendida de los locutores. Los valores corresponden al promedio de identificación cuando se elige el voto ganador por mayoría para cada oración.

Foco Pretendido	%Foco Percibido por la mayoría Afirmativas				%Foco Percibido por la mayoría Interrogativas			
	Neutro	Sujeto	Objeto	Ambiguo	Neutro	Sujeto	Objeto	Ambiguo
Neutro	<b>13.48</b>	1.12	76.40	8.99	<b>1.14</b>	40.91	45.45	12.50
Sujeto	2.25	<b>92.13</b>	3.37	2.25	0.00	<b>94.38</b>	3.37	2.25
Objeto	0.00	0.00	<b>97.65</b>	2.35	0.00	22.47	<b>75.28</b>	2.25

Tabla 2. Porcentajes de identificación de las condiciones de foco para las modalidades afirmativa e interrogativa obtenida por la elección de las preguntas.

En los casos de foco estrecho la ubicación fue identificada de acuerdo a lo esperado. Sin embargo, las oraciones con foco neutro no fueron bien producidas y/o reconocidas, siendo identificadas mayoritariamente como con foco en el objeto (76.40% y 45.45%, para oraciones afirmativas e interrogativas, respectivamente).

### 3.1.3 Evaluación de las prominencias

Los sujetos asignaron niveles de prominencia a todas las sílabas mediante las escalas móviles, que fueron luego mapeadas a una escala entera de -5 a +5. En la Figura 1 se presentan los valores de prominencia normalizados por palabra mediante Z- Scores, para toda la oración, de acuerdo a las asignaciones de foco percibidas por los oyentes.

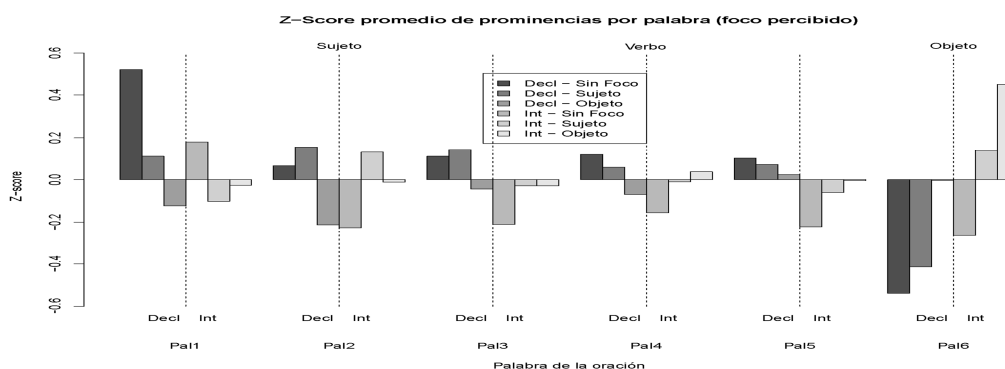


Figura 1. Prominencias máximas normalizadas por z-scores para las tres palabras de contenido y las tres de función. Las líneas de punto separan declarativas de interrogativas.



### 3.1.4. Relación entre Prominencia y Foco percibido

Nuestra hipótesis inicial considera que la asignación de foco está vinculada con la máxima prominencia percibida en toda la frase. Esta relación no se verificó. En segundo lugar se trató de vincular el foco con las diferencias entre las prominencias del sujeto y el objeto. Nuevamente las asociaciones no mostraron ser confiables y significativas. Finalmente, se observaron las prominencias locales en el sujeto y en el objeto por separado. En esta situación, la valoración de la prominencia máxima en el predicado mostró ser un indicador adecuado para la marcación de foco en todas sus variantes, tanto para las oraciones afirmativas como para las interrogativas, como se observa en los gráficos de caja de la Figura 2. El análisis estadístico permite ver que la prominencia máxima en el objeto –no necesariamente la máxima de la oración– permite distinguir los tres grupos focales, tanto en las oraciones declarativas como en las interrogativas (test de Suma de rangos Wilcoxon con  $p < 0.001$ ). Esta medida de prominencia se evaluó también como parámetro para distinguir los casos de foco en oraciones declarativas de las interrogativas, donde se observa un incremento en la percepción de prominencias para las interrogativas. La prominencia máxima en el predicado mostró diferencias significativas para distinguir todos los casos (test de Wilcoxon,  $p < 0.001$ ), salvo para los casos de foco en el sujeto en declarativas vs. no foco de las interrogativas.

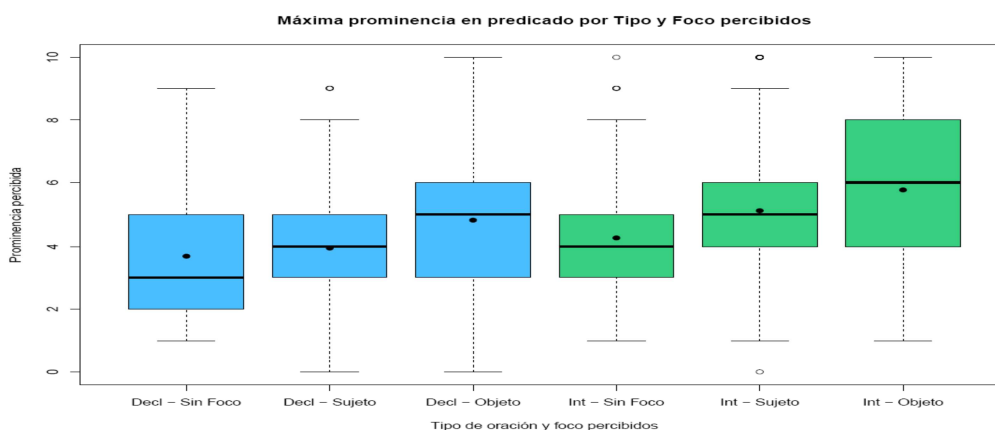
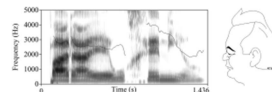


Figura 2. Gráficos de caja para la máxima prominencia percibida en el objeto en las diferentes condiciones de foco para oraciones afirmativas e interrogativas.

## 4. ANÁLISIS ACÚSTICO

En esta sección se muestran los resultados del análisis acústico de las oraciones emitidas para observar el comportamiento del contorno de frecuencia fundamental. Para realizar esta tarea en forma semi-automática, dado el gran número de oraciones, se emplea el modelo de entonación propuesto por Fujisaki et al. (1994) para el español de Buenos Aires. Este modelo determina parámetros asociados al contorno de entonación original mediante la generación de los comandos de frase y los comandos de acento tonal. De los comandos de acento tonal se estiman la amplitud de los picos de F0 (Aa) y su duración



(T2-T1). Para aplicar el modelo, las 540 emisiones fueron alineadas automáticamente con los fonemas y alófonos resultantes de la conversión del texto empleando el alineador forzado LingWaves, y posteriormente se segmentaron en sílabas. Las alineaciones y etiquetado fueron controlados y corregidos manualmente utilizando el sistema de análisis acústico ANAGRAF del LIS. Así, a partir de las transcripciones y los contornos de F0, se estimaron los modelos para las oraciones del corpus, según el método propuesto en (Torres y Gurlekian, 2016), que vincula los comandos de acento y de frase con la estructura lingüística de la oración. La eficacia del modelo se mide al comparar el contorno de F0 estimado con el original, y el error cuadrático promedio para el contorno de F0 estimado fue de .89 semitonos (ST), con un desvío estándar de .28 ST.

#### 4.1. Foco vs. Parámetros del Modelo

Además de la amplitud Aa y duración del comando de acento tonal (T2-T1), se define la *energía* de un comando de acento tonal como el producto  $Aa \cdot (T2-T1)$ . La energía ha mostrado ser el mejor indicador frente a la amplitud y la duración por separado. Se analizó la energía del comando de acento con el foco producido y percibido como se resume en la tabla 3, donde se observa el porcentaje de alineaciones del pico de F0 con la sílaba que posee acento léxico.

% de Alineaciones del pico de F0 en cada sílaba con acento lexical								
Foco Producido	Total de evaluadores				Voto de la mayoría			
	fon/ta/trjan		sjon/su/pa		fon/ta/trjan		sjon/su/pa	
	D	I	D	I	D	I	D	I
sin Foco	40	52	62	45	50	52	75	45
en el Sujeto	90	72	75	55	<b>97</b>	72	74	55
en el Objeto	46	52	82	58	49	52	<b>84</b>	58

Tabla 3. Alineación del pico máximo de F0 respecto a la sílaba léxica para declarativas e interrogativas.

#### 4.2 Prominencias vs Parámetros Acústicos

Para analizar la correlación entre la prominencia y los parámetros acústicos obtenidos de los comandos de Fujisaki, se procedió a contar cuantas veces coincidía la sílaba más prominente con el comando de mayor energía. Dada la variabilidad inter-evaluador, se considera la sílaba más prominente según el voto de la mayoría de los evaluadores. Los resultados se presentan disgregados por tipo y modo de oración. Podemos observar que la energía del comando es el parámetro que mejor indica la presencia de una sílaba prominente.

Los resultados muestran un porcentaje de correlación elevado del comando de acento con la sílaba más prominente de un 82% en promedio. La tabla 4 muestra las correlaciones r de Pearson entre la prominencia versus los rasgos acústicos de la sílaba F0, Duración, Intensidad y los parámetros de los comandos que genera el modelo: Amplitud, Duración y Energía.





Rasgo	Sílabas			Comando		
	F0	DUR	INT	Aa	T2-T1	Energía
Prominencia	.32	.55	.31	.56	.44	.50
F0		.24	.37	.35	.25	.38
DUR			.17	.45	.33	.47
INT				.22	.13	.20
Aa					.62	.90
T2-T1						.71

Tabla 4. Matriz de correlaciones entre la prominencia perceptual y los rasgos prosódicos de las sílabas. Se muestran los valores *r* de Pearson. Significación  $p < 0.001$ .

## 5. DETECCIÓN AUTOMÁTICA DE SÍLABAS PROMINENTES

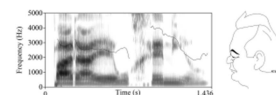
En esta sección, se implementó un detector automático de prominencia a nivel de sílabas empleando rasgos prosódicos obtenidos de las emisiones acústicas. A partir de las puntuaciones de prominencia obtenidas mediante la evaluación perceptual, se definió para cada oyente las sílabas prominentes como aquellas en las que se observó un pico en las puntuaciones informadas. Posteriormente, se aplicó un proceso de votación simple para obtener las etiquetas de prominencia de consenso, que conformaron las salidas deseadas del detector de prominencias propuesto. Como observaciones o parámetros de entrada para este detector se utilizaron para describir cada sílaba los parámetros prosódicos propuestos en (Rosenberg, 2009): duración en segundos, valores mínimos, máximos y desvío estándar de F0, intensidad y énfasis espectral (Heldner, 2001). Para el caso de F0 se obtuvieron dichos valores tanto en Hz como en escala Erb y de semitonos. Para todos los casos se trabajó con los valores normalizados de todos estos parámetros empleando z-score a nivel de locutor. Una vez obtenidos los vectores de observaciones con sus correspondientes salidas deseadas, se construyeron los conjuntos de entrenamiento y evaluación siguiendo la metodología de validación cruzada de diez particiones, dejando en cada partición todos los datos de un hablante como conjunto de evaluación. Se utilizó el programa Weka (Hall, 2009) para construir y evaluar los resultados obtenidos empleando como clasificadores los algoritmos J48 (árboles de clasificación), regresión logística, y SMO (máquinas de vectores soporte con kernel lineal). Las tasas de reconocimiento de prominencias silábicas promediadas para cada partición y para cada algoritmo de clasificación se presentan en la tabla 5.

	J48 %	SMO %	RL %
Promedio	88.24	90.60	90.78
Desvío Estándar	3.57	3.11	3.43

Tabla 5. Tasas de detección de prominencias silábicas promedio para tres clasificadores.

## 6. RESULTADOS Y DISCUSIÓN

Los resultados sobre la percepción de modalidad muestran porcentajes de identificación altos para las oraciones afirmativas (99%) e interrogativas (97%) indicando que las pistas



acústicas emitidas son bien reconocidas. La modalidad de interrogación muestra más errores probablemente debido a la producción interrogativa con un acento bitonal de frase+juntura de tipo retroflejo, típico del español de Buenos Aires. Este efecto puede dar la sensación de afirmación (1.24%) o ambigüedad (1.72%) en unas pocas emisiones. La percepción del foco (tabla 2) muestra una clara tendencia a ubicar a este en el objeto. En particular, las oraciones de foco neutro fueron percibidas como con foco en el objeto en un alto porcentaje (76.40%). Aquí podemos especular que los oyentes al recibir las oraciones declarativas de foco neutro, aun sin indicadores acústicos específicos, presuponen que la información de carácter pragmático se completa habitualmente en el predicado. En particular, las oraciones interrogativas de foco neutro reciben, además, rasgos acústicos que favorecen la percepción del foco en el predicado. En las estructuras sujeto-verbo-objeto, prototípicas en lenguas como el español, se asume que el predicado es más informativo que el sujeto, es decir se reconoce como información no presupuesta o no conocida. Si la información prosódica no indica lo contrario, esa es la estructura no marcada y por ello, el oyente reconocerá al predicado como la información nueva que está focalizada temáticamente (Zubizarreta, 1999). En los contornos de F0 se verifica una característica adicional, no contemplada en el análisis del modelo, dada por los alineamientos tonales. En declarativas e interrogativas de foco neutro, los acentos tonales se desplazan sistemáticamente de las sílabas acentuadas lexicalmente en el sujeto y objeto a la siguiente sílaba. En las declarativas con foco en el sujeto y en el objeto, el acento tonal muestra una clara alineación con la sílaba acentuada correspondiente, aun cuando el comando no posea la máxima energía. La información acústica de foco estrecho en interrogativas no sigue el mismo patrón. En este caso los acentos tonales se desplazan en un número importante a la sílaba siguiente como ocurre para el foco neutro. Estos resultados permiten afirmar que la detección de foco basada únicamente en F0 o en los comandos del modelo de entonación no son suficientes como indicadores acústicos. La alineación de los máximos de F0 son medidas complementarias para la determinación del foco en el sujeto y el objeto en oraciones declarativas, pero no para las interrogativas, que presentan la misma desalineación que las oraciones de foco neutro.

La percepción de las prominencias indica una marcación de niveles decrecientes para las sílabas con foco en el objeto para declarativas, en el sujeto para interrogativas, en el objeto para interrogativas y para el sujeto en declarativas respectivamente. Las prominencias percibidas en cada sílaba no están asociadas directamente con la palabra que se indicó con foco. Sin embargo, el valor de prominencia máxima observada en el objeto resulta ser un buen discriminador de la marcación de foco en todas las condiciones. Puede especularse como hipótesis futura que la información de foco se encuentra –al igual que la modalidad– indicada por la prominencia al final de la frase. Cuando se emplea la información acústica completa de las vocales acentuadas lexical y tonalmente, la detección de las prominencias es satisfactoria como se indica en la tabla 5.

## 7. CONCLUSIONES

La indicación de foco no está asociada directamente con el nivel de prominencia percibida en el segmento focalizado. El valor de prominencia máxima en el objeto resulta ser el indicador que discrimina con significancia estadística la marcación de foco en todas las condiciones. Los máximos de F0 y los comandos del modelo de entonación se asocian



mejor con los niveles de prominencia. Los parámetros acústicos directos como los máximos de F0, la energía en dB, la duración y el espectro vocálico se complementan para dar la información acústica de prominencia, como se verifica al emplear estos datos en el sistema de clasificación automático. Se concluye en este trabajo que el foco estaría marcado por el nivel de prominencia en el objeto y que los niveles de prominencia se asocian con los componentes acústicos con una precisión cercana al 90%.

## AGRADECIMIENTOS

Este trabajo recibió el apoyo financiero del Fondo para la Ciencia y Tecnología de Argentina FONCYT para el proyecto bilateral con Alemania patrocinado por la BMBF.

## 8. REFERENCIAS BIBLIOGRÁFICAS

- DIK S. (1997): «The theory of functional grammar I. The Structure of the Clause», *Functional Grammar Series*, 20, Berlin, Mouton de Gruyter.
- FERNÁNDEZ PLANAS A. Ma. (2005): «Aspectos generales acerca del proyecto Internacional AMPER en España», *Estudios de fonética experimental*, XIV, pp.13-27.
- FRY D. B. (1958): «Experiments in the perception of stress», *Language and Speech*, 1, pp. 126-152.
- FUJISAKI, H.; K. NAKAMURA; M. GUIRAO y J. A. GURLEKIAN (1994): «Computational modelling of accent and intonation in declarative Sentences of Spanish», *Journal of the Acoustical Society of America*, 95, 5, p. 2949.
- GURLEKIAN J. A. y G. TOLEDO (2008): «Datos preliminares del AMPER Argentina: Las oraciones declarativas e interrogativas absolutas sin expansión», en A.Pamies, M. C. Amorós y J.M. Pazos (eds.): *Language Design, Special Issue: Experimental Prosody*, 2, pp. 213-220.
- HALL M.; E. FRANK; G. HOLMES; B. PFAHRINGER; P. REUTEMANN y I. H. WITTEN (2009): «The Weka data mining software: an update», *ACM SIGKDD explorations newsletter*, 11(1), pp. 10-18.
- HELDNER M. (2001): «Spectral emphasis as an additional source of information in accent detection», en M. Bacchiani, J. Hirschberg, D. Litman y M. Ostendoirf (eds.): *Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*. ISCA, Red Bank, pp 57-60.
- LADD, D. R. (1983): «Even, Focus and normal Stress», *Journal of Semantics*, 2, pp. 157-170.
- MIXDORFF H.; C. COSSIO-MERCADO; A. HÖNEMANN; J. A. GURLEKIAN; D. EVIN y H. TORRES (2015): «Acoustic Correlates of Perceived Syllable Prominence in German», en S. Moller, H. Ney, B. Mobius, E. Noth y S. Steidl (eds.): *Proceedings of Interspeech'15*, Dresden, ISCA, pp. 51–55.
- NÖTH E.; A. BATLINER, A. KIEBLING, R. KOMPE y H. NIEMANN (2000): «The Use of Prosody in the Linguistic Components of a Speech Understanding System», *IEEE Transactions on Speech and Audio Processing*, 8, 5, pp. 519-532.
- ROSENBERG A. (2009): *Automatic Detection and Classification of Prosodic Events*, tesis doctoral, Columbia University.
- TOLEDO G. (2008): «Fonología entonativa: los acentos tonales finales de frase entonativa intermedia (ipT\*) frente al tono de frontera (H-) en discursos y texto leídos del español de Buenos Aires», *Language Design*, 9, 2, pp. 129-136.
- TORRES H. y J. GURLEKIAN (2016): «Novel Estimation Method for the Superpositional Intonation Model», *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24, 1, pp. 151-160.
- VILAPLANA E. y P. PRIETO VIVES (2008): «La notación prosódica del español: una revisión del sp\_ToBI», *Estudios de fonética experimental*, XVII, pp. 263-283.
- ZUBIZARRETA ML. (1999): «Las funciones informativas: tema y foco», en I. Bosque y V. Demonte (eds.): *Gramática descriptiva de la lengua española*, Madrid, Espasa Calpe, pp. 4215-4244.