

# **EVOLUCIÓN DE LOS FORMANTES DEL HABLA. RESULTADOS OBTENIDOS UTILIZANDO ALGORITMOS ESPECÍFICOS DE TRATAMIENTO DE SEÑAL**

[1] Jesús Bobadilla, [2] Pedro Gómez y [1] Jesús Bernal

[1]

Departamento de Informática Aplicada  
Escuela Universitaria de Informática  
Ctra. de Valencia Km. 7, 28031 Madrid  
Tel: +34.91.3367862, Fax: +34.91.3367527  
e-mail: [jbobi@eui.upm.es](mailto:jbobi@eui.upm.es), [jbernal@eui.upm.es](mailto:jbernal@eui.upm.es)

[2]

Departamento de Arquitectura y Tecnología de Sistemas Informáticos  
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n,  
Boadilla del Monte, 28660 Madrid  
Tel: +34.91.3367384, Fax: +34.91.3367412  
e-mail: [pedro@pino.datsi.fi.upm.es](mailto:pedro@pino.datsi.fi.upm.es)

## **RESUMEN**

Este artículo aporta una gran cantidad de sonogramas que muestran la evolución de los formantes del habla en porciones de sonido VCV. Los sonogramas han sido escogidos como representativos entre un gran número de pruebas realizadas con cinco hablantes de diferentes sexos. Los algoritmos de tratamiento de la señal que se han utilizado para obtener los sonogramas se basan en desarrollos originales centrados en el método de Predicción Lineal (LPC).

## **ABSTRACT**

This article exhibits a large quantity of spectral images showing the speech formant evolution in VCV patterns. The spectral images have large been chosen as representative among a number of experiments accomplished with five speakers (male and female). The signal processing algorithms used to obtain the spectral images are based on original programs centered in Linear Prediction (LPC) methods.

## 1. INTRODUCCIÓN

Entre la información más representativa que permite la comprensión de un espectrograma de voz esta la posición y evolución de los formantes existentes en los sonidos sonoros [PET52], [KAT95], sin embargo, una vez obtenidos, es necesario conocer sus peculiaridades, que varían apreciablemente según quien sea el hablante, los distintos contextos que presentan los sonidos, la entonación empleada en las frases, etc. [TOK93]

Aunque las consonantes en sí mismas presentan características importantes que contribuyen a su identificación, resulta fundamental determinar las transiciones vocálicas que las rodean. En este artículo se realiza un estudio de la evolución de los formantes en las vocales coarticuladas con algunos sonidos consonánticos del castellano.

Es importante recordar que diversas publicaciones [MAS75], [REP78] señalan la importancia de la vocal posterior a la consonante como elemento de identificación consonántica, por lo que en los resultados que a continuación se presentan conviene fijar una mayor atención en la vocal posterior que en la anterior en los grupos /VCV/.

En [QUI93] se realiza un breve repaso de las evoluciones más significativas de los formantes ante algunos sonidos del español. La imprecisión de los resultados conseguidos hasta el momento para su utilización en aplicaciones informáticas, nos da una idea de la complejidad que presenta la investigación en este aspecto de la fonética acústica.

Los sonogramas que se presentan en este artículo han sido obtenidos mediante aplicaciones informáticas propias que utilizan algoritmos originales de tratamiento de señal ideados a partir del método de predicción lineal (LPC) [RAB93], [RAN95], [ROW92], [PAR86].

## **2. METODOLOGÍA**

El estudio que se muestra ha sido realizado con cinco hablantes de diferentes sexos y edades, utilizando la siguiente metodología:

1. Realizar grabaciones de secuencias /VCV/ tales como 'apa', 'epe', 'ipi', 'opo', 'upu', 'aβa', 'eβe', etc.
2. Obtener los espectros básicos y mejorados utilizando algoritmos propios desarrollados a partir del método de Predicción Lineal (LPC).
3. Repetir el proceso con las grabaciones en las que no aparece una evolución de formantes bien definida.
4. Estudiar y comparar las evoluciones más comunes que se presentan en los formantes, mostrando espectros representativos de cada uno de los casos estudiados.

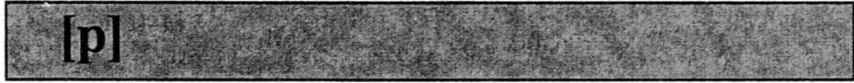
En este trabajo se incluye parte de los espectros más representativos obtenidos utilizando el proceso detallado. Los espectros han sido clasificados atendiendo al modo de articulación de los sonidos consonánticos. En cada caso se detalla la cuantía de las subidas o bajadas de las transiciones vocálicas medidas en Hercios. Puesto que el tono de la voz varía según el sexo, en cada caso se indica como H/M (Hombre/Mujer) esta característica, que influye en la altura frecuencial de los formantes.

La tabla siguiente puede ser utilizada para identificar los diferentes sonidos que mostraremos a lo largo de este artículo. Como se puede observar, la notación que se ha seguido es la definida por el AFI.

	Bilabial		Labiodental		Dental		Interdental	
	sordas	sonor.	sordas	sonor.	sordas	sonor	sordas	sonor
Oclusivas	[p]	[b]			[t]	[d]		
Fricativas		[β]	[f]				[θ]	[ð]
Africadas								
Nasales		[m]		[ɱ]		[ɳ]		[ɲ]
Laterales						[l̪]		[l̟]
Vibr. simple								
Vibr. doble								

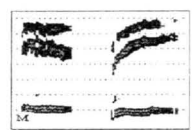
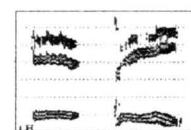
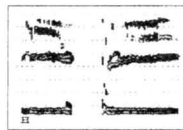
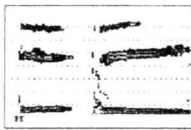
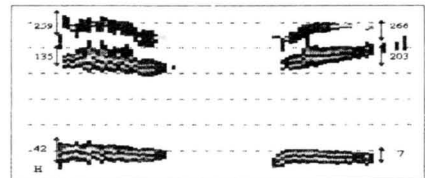
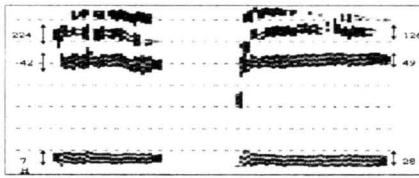
	Alveolar		Palatal		Velar	
	sordas	sonor	sord.	sonor	sord	sonor
Oclusivas					[k]	[g]
Fricativas	[s]			[ç]	[x]	[χ]
Africadas			[tʃ]	[dʒ]		
Nasales		[n]		[ɲ][ɲ,]		[ŋ]
Laterales		[l]		[ʎ][l,]		
Vibr. simple		[r]				
Vibr. doble		[r̄]				

### 3. RESULTADOS



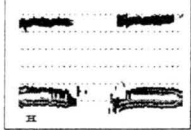
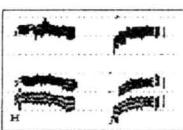
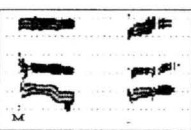
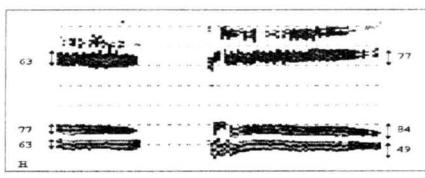
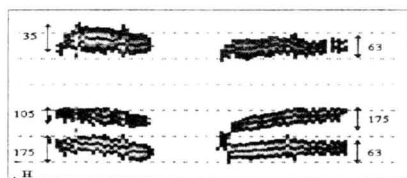
ipi

epe

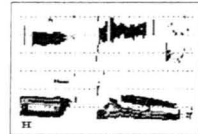
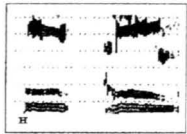
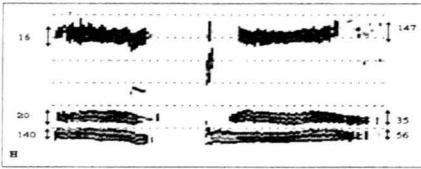


apa

opo

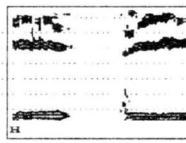
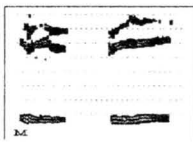
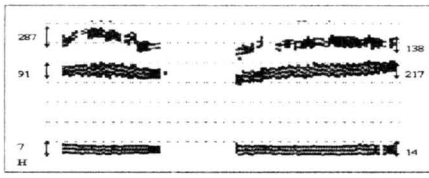


**upu**

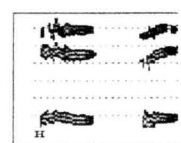
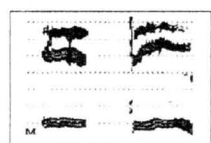
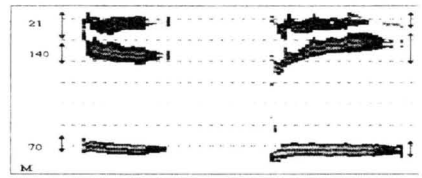


**[t]**

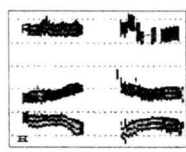
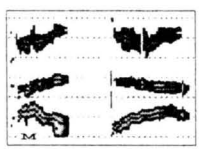
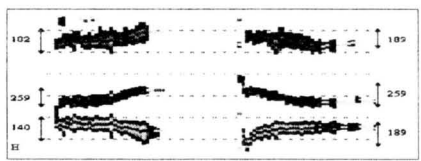
**iti**



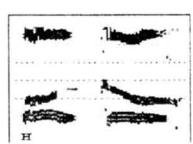
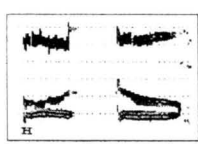
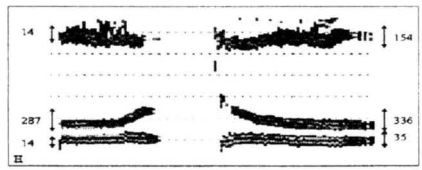
**ete**



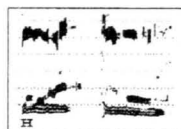
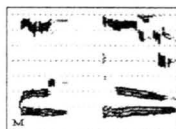
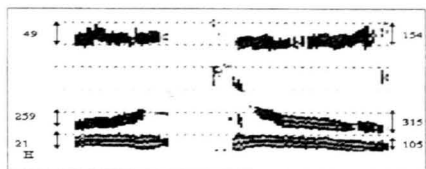
**ata**



**oto**



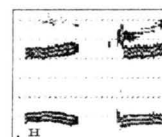
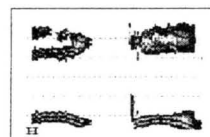
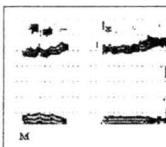
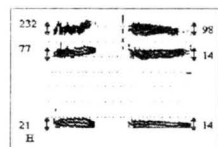
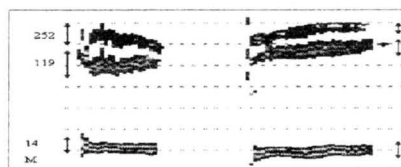
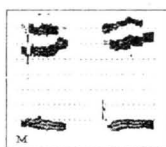
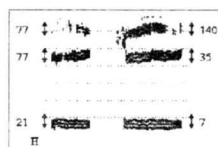
**utu**



**[k]**

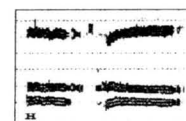
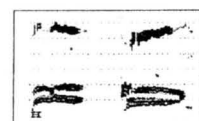
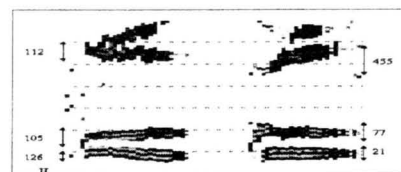
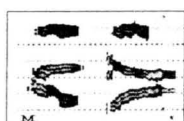
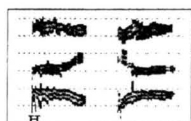
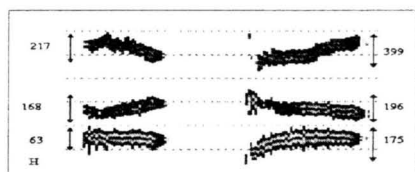
**iki**

**eke**



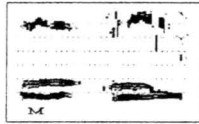
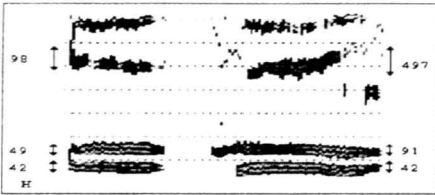
**aka**

**oko**



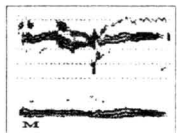
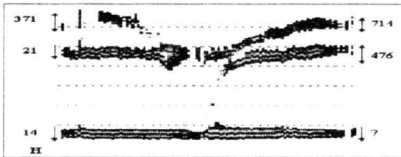


**uku**

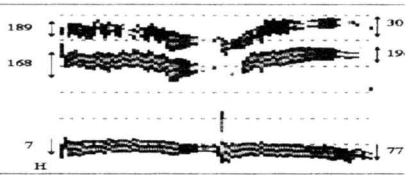


**[β]**

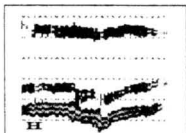
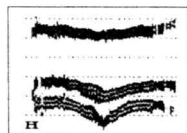
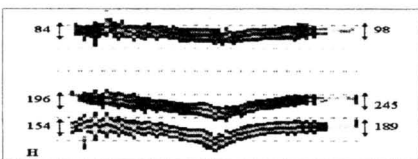
**ɪβi**



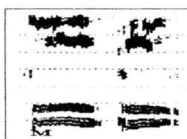
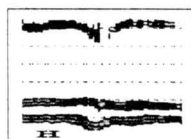
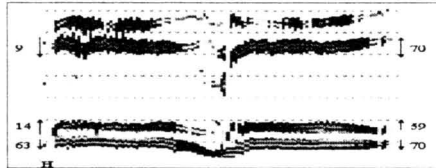
**eβe**



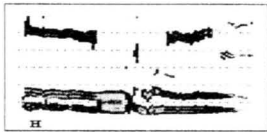
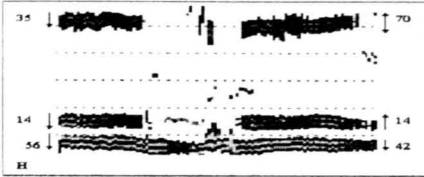
**aβa**



**oβo**

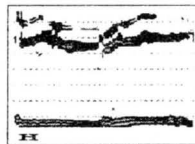
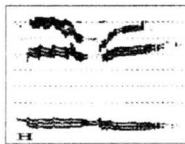
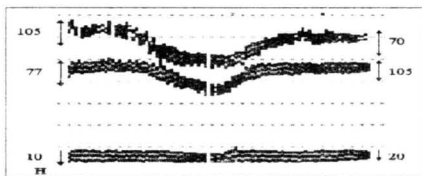


**uβu**

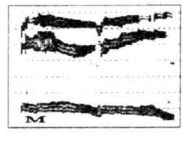
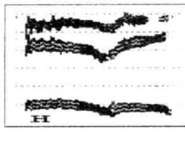
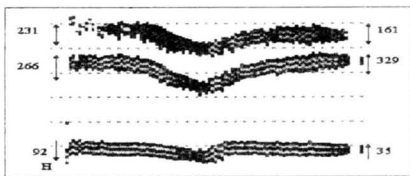


**[ð]**

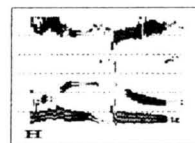
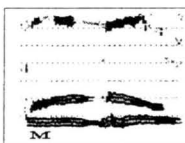
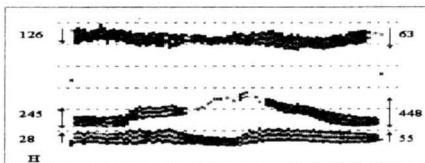
**i ð i**



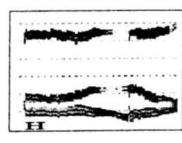
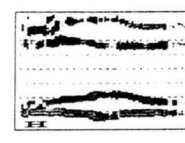
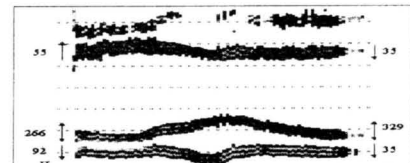
**e ð e**



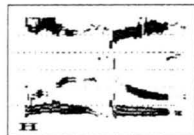
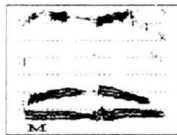
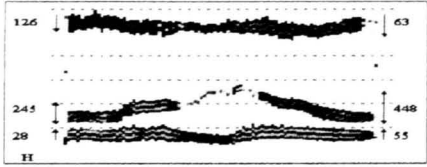
**a ð a**



**o ð o**

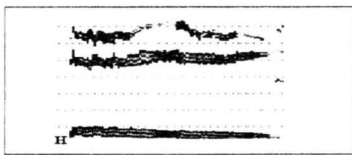
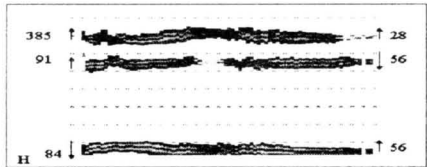


u ð u

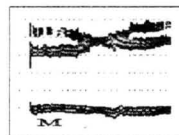
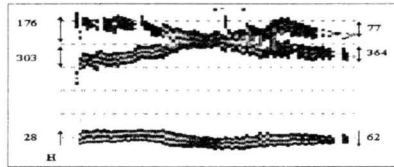


[ɣ]

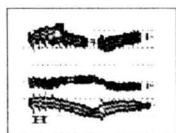
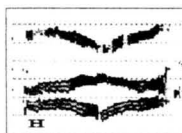
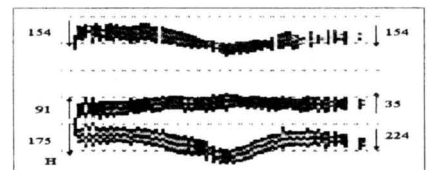
iyi



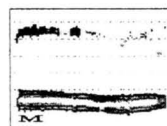
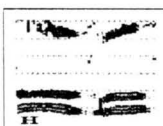
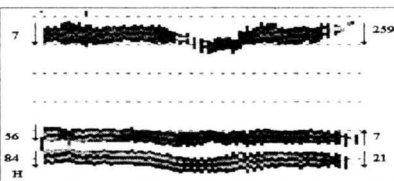
eye



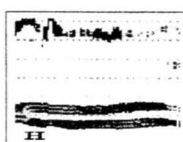
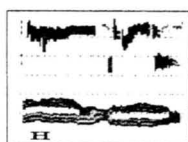
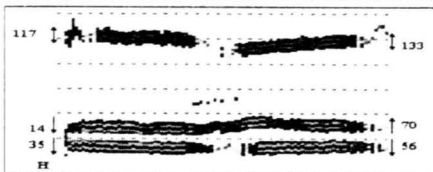
aya



oyo

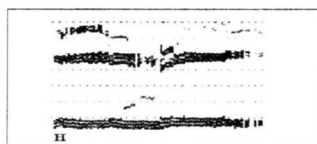
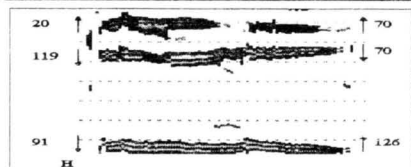


**uyu**

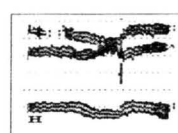
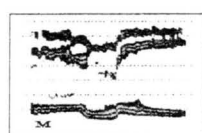
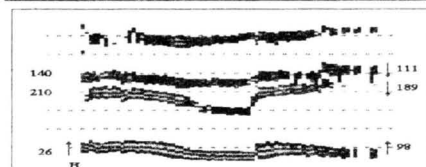


**[m]**

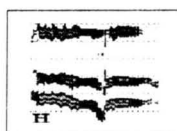
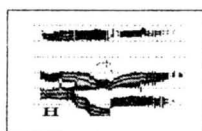
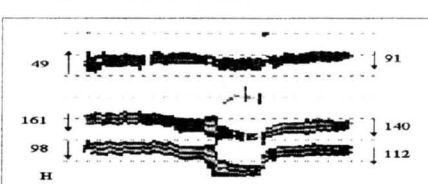
**imi**



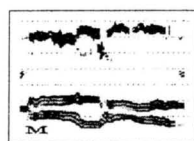
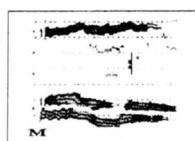
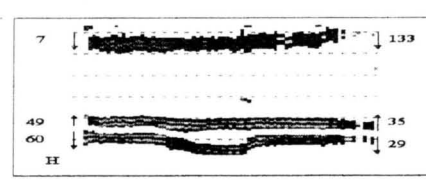
**eme**



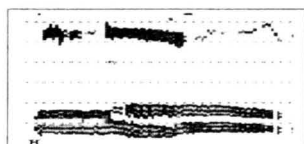
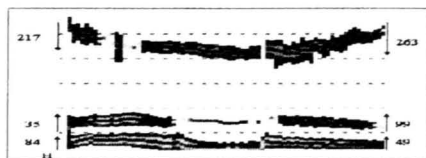
**ama**



**omo**

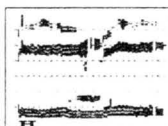
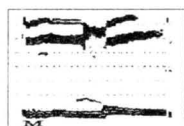
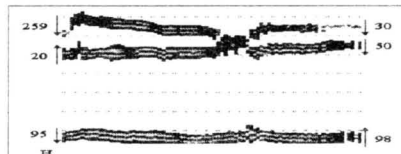


**umu**

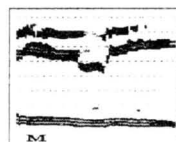
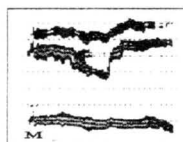
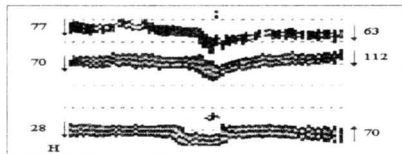


**[n]**

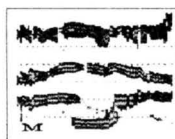
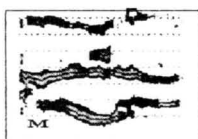
**ini**



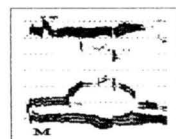
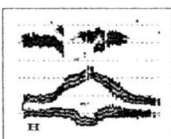
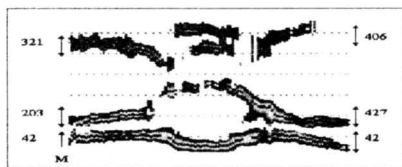
**ene**



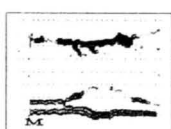
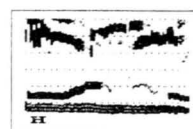
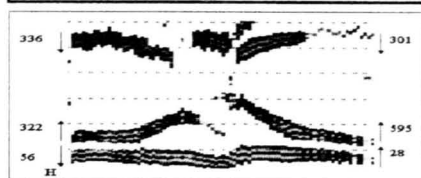
**ana**



**ono**

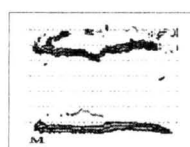
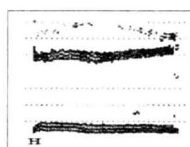
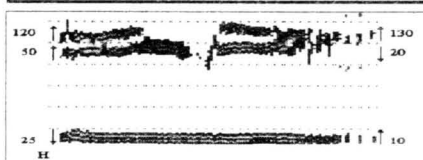


**unu**

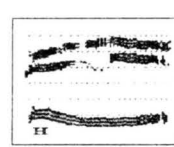
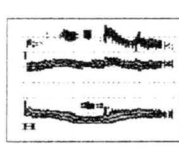
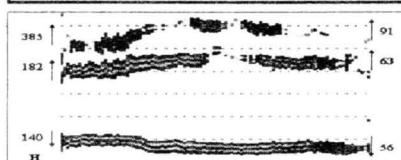


**[ŋ]**

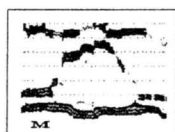
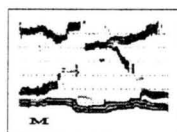
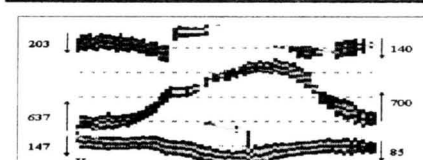
**i ŋ i**



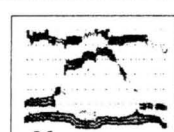
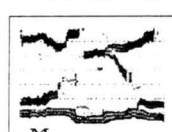
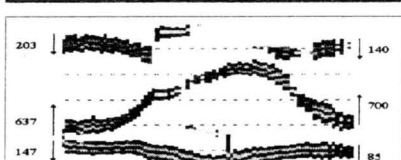
**e ŋ e**



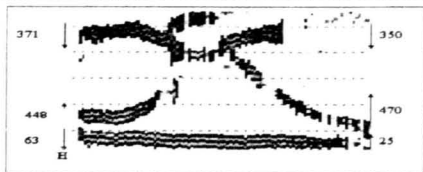
**a ŋ a**



**o ŋ o**

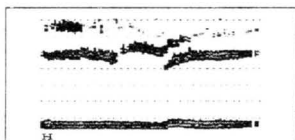
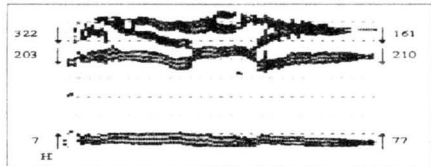


**u η u**

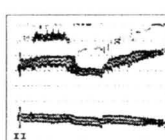
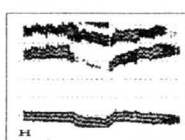
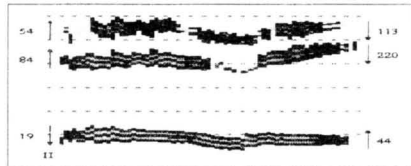


**[I]**

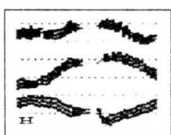
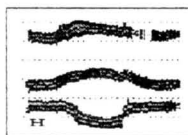
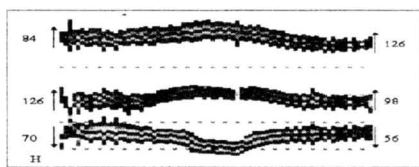
**ili**



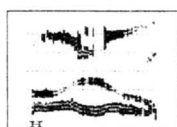
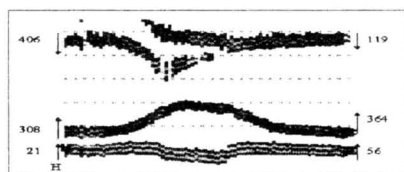
**ele**



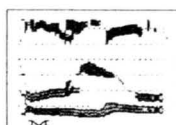
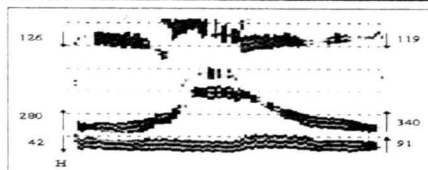
**ala**



**olo**

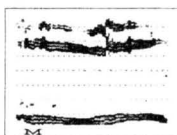
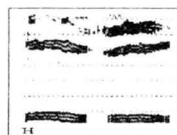
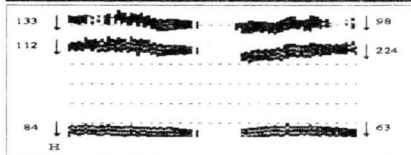


**ulu**

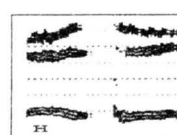


**[λ]**

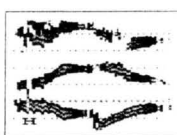
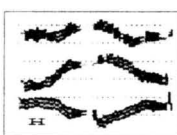
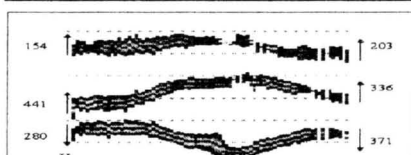
**i λ i**



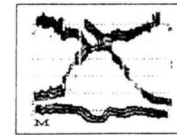
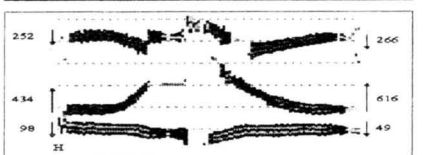
**e λ e**



**a λ a**

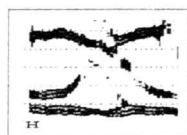
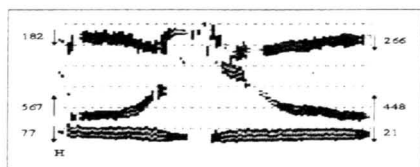


**o λ o**



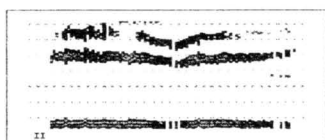
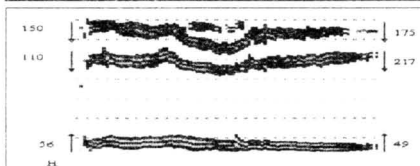


u ʌ u

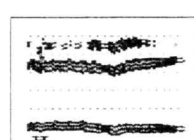
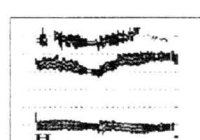
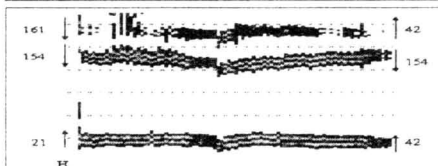


[r]

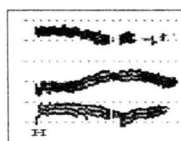
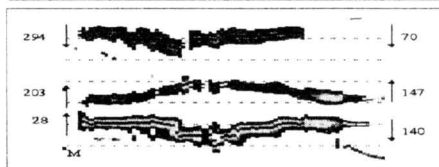
iri



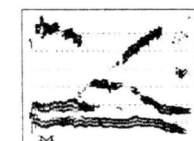
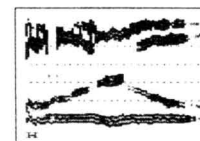
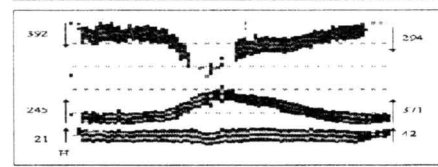
ere

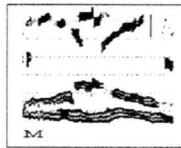
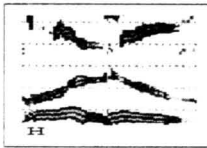
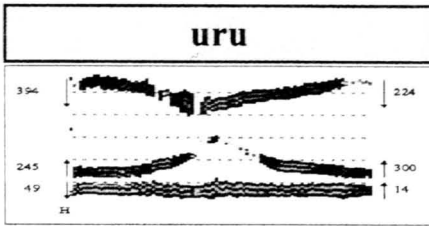


ara

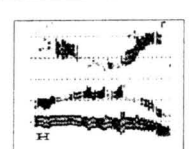
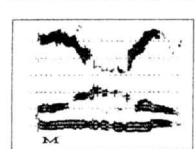
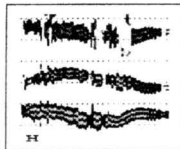
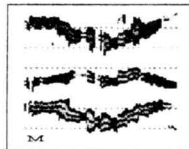
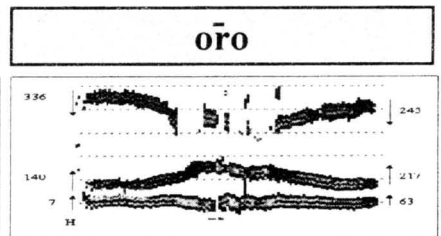
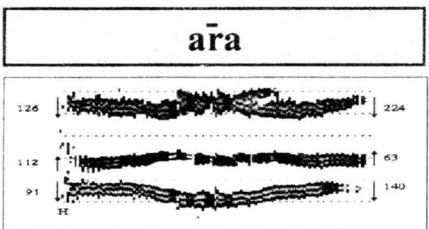
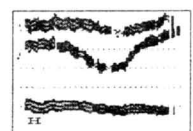
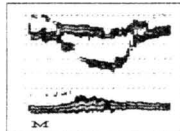
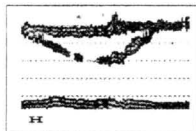
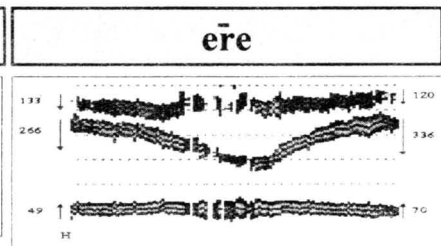
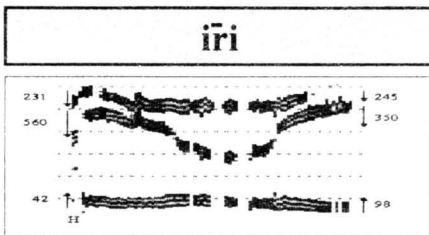


oro

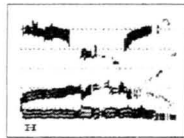
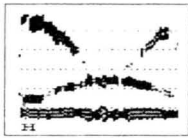




**[r]**

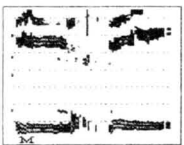
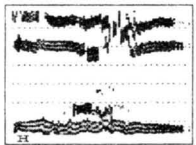
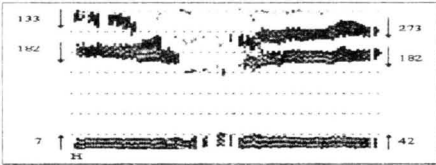


**ūru**

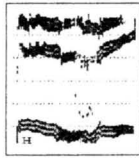
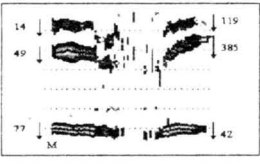
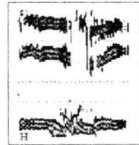
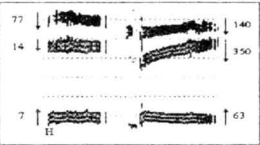


**[θ]**

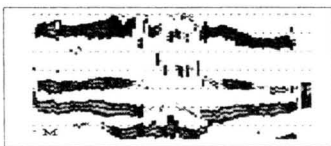
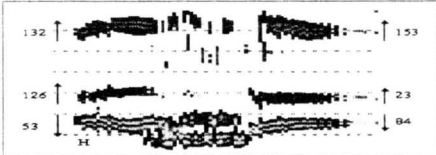
**iθi**



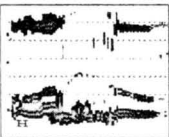
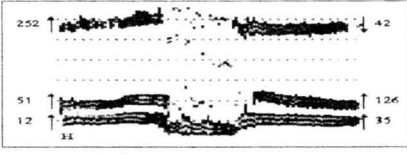
**eθe**



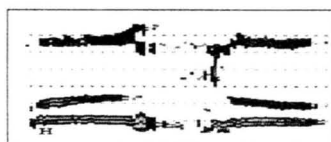
**aθa**



**oθo**

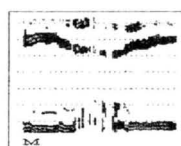
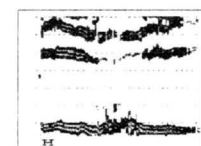
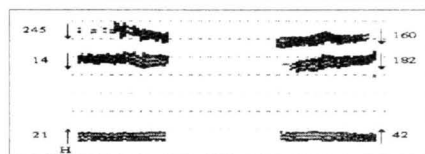


**uθu**

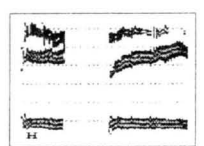
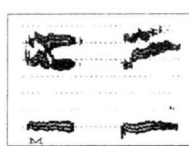
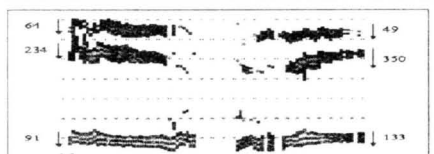


**[s]**

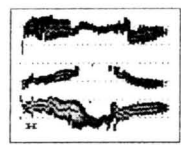
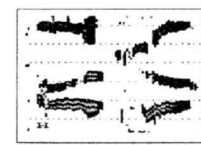
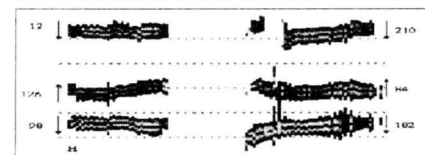
**isi**



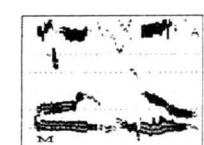
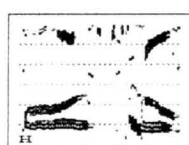
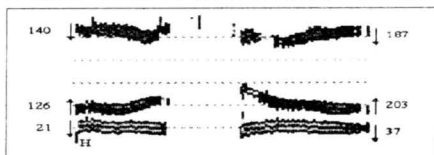
**ese**



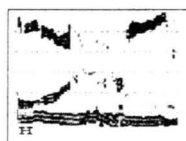
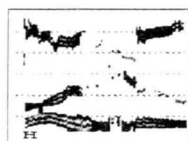
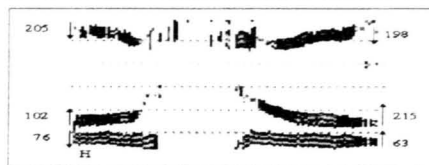
**asa**



**oso**

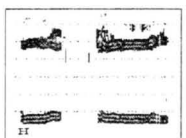
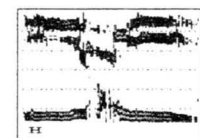
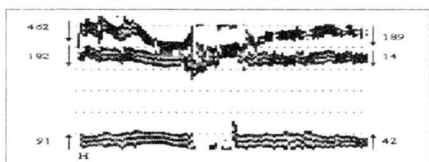


**usu**

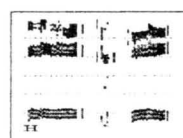
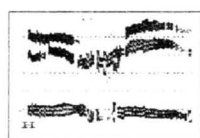
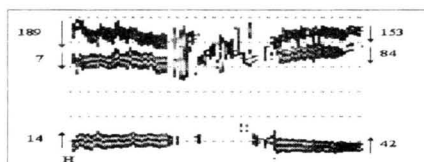


**[X]**

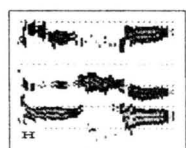
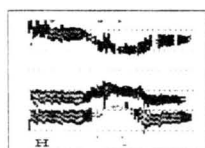
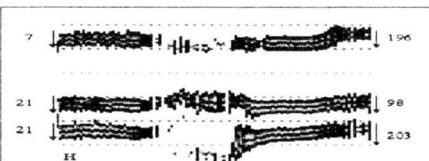
**ixi**



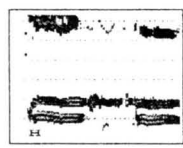
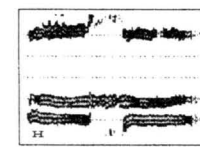
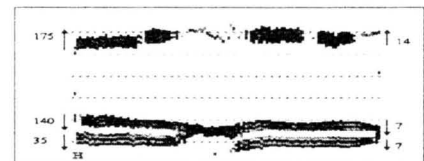
**exe**



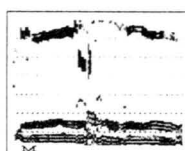
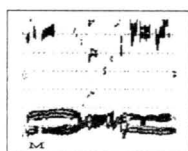
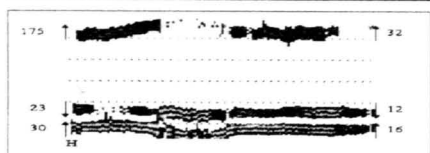
**axa**



**oxo**

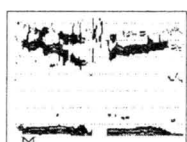


**uxu**

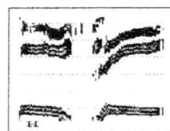
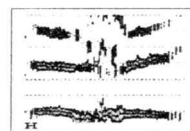
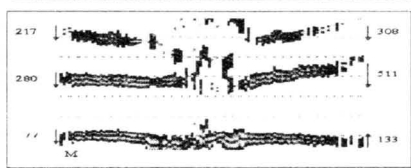


**[f]**

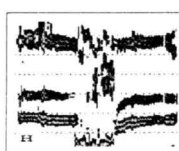
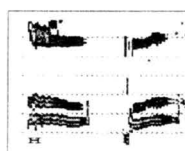
**ifi**



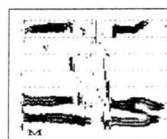
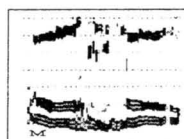
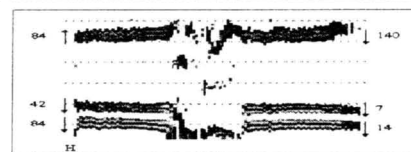
**efe**



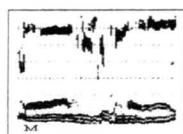
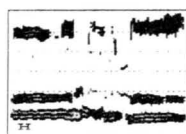
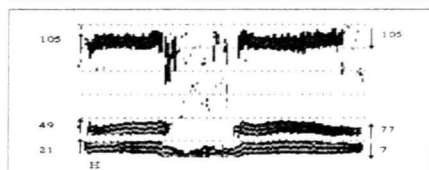
**afa**



**ofo**

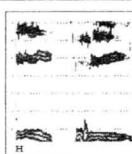
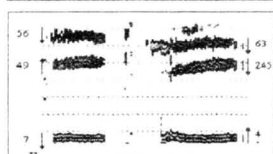


**ufu**

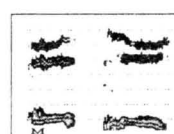
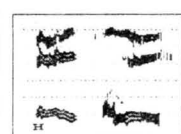
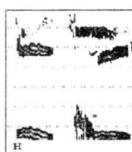
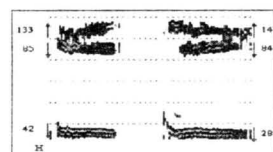
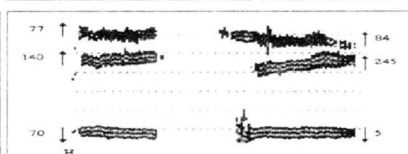


**[tʃ]**

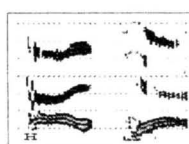
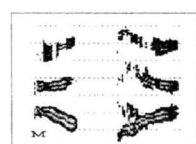
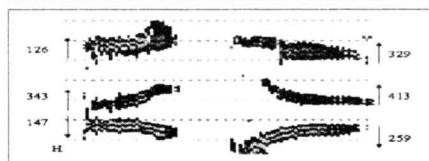
**itʃi**



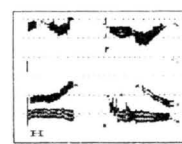
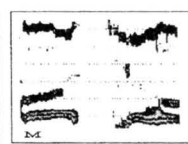
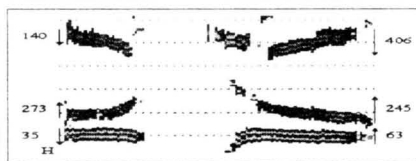
**etʃe**

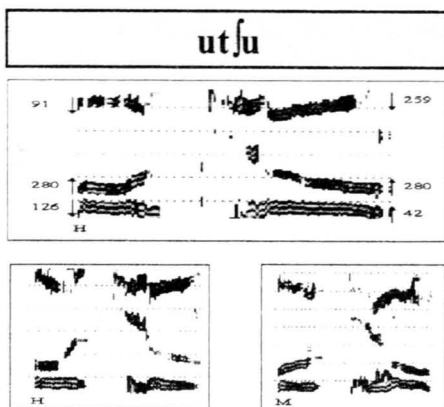


**atʃa**



**otʃo**





#### 4. CONCLUSIONES

En el trabajo realizado sobre la evolución de los formantes, se obtiene como primera y más importante conclusión la constatación del hecho de que existen grandes diferencias entre el estudio de las posiciones estáticas de los formantes y el de sus evoluciones. Mientras que la posición de los formantes se puede determinar con relativa facilidad, la obtención de sus evoluciones a lo largo del tiempo resulta más costosa y menos fiable de lo que cabría esperar.

Existe una gran variabilidad en la evolución de los formantes de las vocales coarticuladas con sonidos consonánticos, esta variabilidad nos impide ofrecer unos resultados generales y fiables; de hecho, los resultados de este trabajo se pueden tomar como una pauta, pero no como una regla de comportamiento de los formantes de las vocales junto a sonidos consonánticos. Un estudio profundo en esta dirección daría cabida sin lugar a dudas a un importante avance en el campo de la fonética acústica del idioma analizado.



La utilización de varios hablantes en las grabaciones genera resultados dispares, produciéndose ciertas dispersiones en las evoluciones de formantes obtenidas. Las evoluciones de los formantes, aunque siguen unas reglas y pautas generales, presentan mucha variación entre distintas realizaciones del habla. Para poder realizar análisis representativos es necesario obtener, confrontar y comparar un gran número de grabaciones de voz de diferentes hablantes pronunciando sonidos básicos rodeados de diferentes contextos.

En este artículo se ofrece una gran cantidad de casos de estudio en forma de espectros de voz. Estos datos, junto a la metodología y herramientas empleados, pueden servir de base para enfocar un trabajo de objetivos más ambiciosos en el campo de la fonética.

Las cualidades de los métodos y algoritmos desarrollados han sido validadas en los estudios aquí realizados, pudiéndose afirmar que se proporciona un buen soporte para el análisis de diversas características espectrales del habla, especialmente aquellas relacionadas con la posición y la evolución de los formantes del habla.

## 5. BIBLIOGRAFÍA

- [MAS75] D.W. Massaro, "Preperceptual images, processing time, and perceptual units in speech perception", *Understanding language (Academic Press New York)*, 1975, pp. 125-150
- [PAR86] T. Parsons, *Voice and speech processing*, Mc Graw Hill, 1986
- [PET52] G.E. Peterson, H.L. Barney, "Control methods used in a study of the vowels", *Journal of the Acoustic Society of America*, Vol. 24, 1952, pp. 175-184

- [QUI93] A. Quilis, *Tratado de fonología y fonética españolas*, Gredos, 1993
- [RAB93] L.R. Rabiner, *Fundamentals of speech recognition*, Biing-Hwang Juang, Prentice Hall, 1993
- [RAN95] M. Rangoussi, A. Delopoulos, "Recognition of unvoiced stops from their time-frequency representation", *International Conference on Acoustics, Speech and Signal Processing*, Vol. 1, 1995, pp. 792-795
- [REP78] B.H. Repp, "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants", *Perception Psychophysics*, Vol. 24, 1978, pp. 471-485
- [ROW92] C. Rowden, *Speech processing*, Mc Graw Hill, 1992
- [TOK93] S. Tokuma, "Some arguments on vowel formant shift", *Speech, Hearing and Language: Work in Progress*, UCL, Vol. 7, 1993, pp. 233-254

