

ESTUDIO DE LA POSICIÓN DE F1, F2 Y F3 EN SITUACIONES DE COARTICULACIÓN EMPLEANDO TÉCNICAS DE PREDICCIÓN LINEAL

[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: jrobi@eui.upm.es, 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

RESUMEN

En este artículo se muestran diversos mapas de formantes obtenidos a partir de grabaciones de sonidos pronunciados en situaciones de coarticulación. Las muestras de voz se han procesado empleando algoritmos de predicción lineal, extrayéndose los tres primeros formantes. Después se han creado mapas tridimensionales y bidimensionales que muestran la posición y evolución de estos formantes. En el estudio se utilizan cuatro hablantes y más de 3000 valores frecuenciales de posición de formantes, en un intento de validar la metodología y herramientas propias empleadas.

ABSTRACT

This article shows several formant maps obtained from speech signal recording produced using articulated VCV situations. The speech samples have been computed using Linear Prediction algorithms (LPC), obtaining the three first formants. Later, three-dimensional and two-dimensional maps have been created in order to show the formant positions and evolutions. In this study we have used four speakers and more than 3000 formant position frequencies, trying to validate the original developed methodology and tools.

1. INTRODUCCIÓN

Entre la información más representativa que permite la comprensión de un espectrograma de voz se encuentra la posición y evolución de los formantes existentes en los sonidos sonoros [PET52], [KAT95], [MAR95], 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].

El presente estudio pretende establecer las características típicas de algunos sonidos de la lengua castellana y presentar las variaciones que de forma natural se producen en el habla.

Los resultados conseguidos mediante técnicas LPC y transformaciones no lineales de señal, al presentarse de forma gráfica, facilitan la caracterización y clasificación de los sonidos en base a la disposición y evolución de sus formantes. Con los valores numéricos obtenidos se realiza un estudio detallado que establece mapas bidimensionales de los formantes hallados.

Resulta necesario establecer las porciones de sonidos que van a ser estudiadas. Un enfoque tradicional para su elección, se basa en la selección de los sonidos básicos y los alófonos más importantes que presenta el español, todo ello accesible mediante la bibliografía clásica de fonética española [QUI93], [MAR94]. En lugar de ello, se ha considerado más oportuno realizar grabaciones de sonidos rodeados de diferentes contextos, por ejemplo: 'epe', 'apa', 'ete', 'usu', etc. buscándose las variaciones de los sonidos desde un estudio basado totalmente en la fonética acústica.

Determinar el posible rango de frecuencias que puede adoptar cada formante en las vocales castellanas, ayudaría a la comprensión de los espectros de voz y al reconocimiento de sonidos aislados basado en la localización de sus formantes constitutivos. Este objetivo

requiere la utilización de una base de datos de muestras lo suficientemente grande y variada como para que garantice la generalidad de los resultados.

El estudio que aquí se presenta, pretende ser una referencia de cómo clasificar, obtener y procesar los datos de entrada (muestras de voz), con el fin de conseguir unos resultados claros y representativos. El número de hablantes empleados ha sido cuatro (dos hombres y dos mujeres), lo que hace imposible considerar los resultados como definitivos a pesar de las más de 3000 muestras de formantes utilizadas. Con las herramientas informáticas desarrolladas se podría abordar con facilidad un estudio más amplio que aunque resultaría tedioso, sería sencillo de llevar a cabo.

Aunque en la bibliografía tradicional habitualmente se compara únicamente la evolución de los dos primeros formantes [PET52], [HIL95], [MAR94], en este caso, confrontaremos los tres primeros, tomados de dos en dos en sus tres posibles combinaciones.

2. METODOLOGÍA EMPLEADA

Las muestras han sido tomadas para cada uno de los hablantes completando los siguientes grupos:

- 1.- Vocales adyacentes a sonidos bilabiales
- 2.- Vocales adyacentes a sonidos dentales e interdentales
- 3.- Vocales adyacentes a sonidos velares
- 4.- Vocales aisladas

Las consonantes seleccionadas han sido las nasales, oclusivas sordas y fricativas sonoras.

Las etapas fundamentales para la realización de este trabajo son:

1.- Grabación de las muestras, agrupadas atendiendo a:

- a) Hablante
- b) Grupo (bilabial, dental/interdental, palatal/velar, aisladas)
- c) vocal (a,e,i,o,u)

2.- Cálculo de los parámetros de predicción lineal de cada grabación.

3.- Obtención de los espectros de voz.

4.- Determinación de la posición de los formantes mediante los algoritmos de tratamiento de señal desarrollados para realizar este estudio.

5.- Selección de los instantes de tiempo más significativos en cada grabación.

6.- Inserción ordenada en una base de datos de los valores frecuenciales de los tres primeros formantes para todos los instantes de tiempo seleccionados en la etapa anterior.

7.- Realización de selecciones sobre la base de datos, atendiendo a diferentes combinaciones de los parámetros establecidos en la primera etapa.

8.- Creación de mapas tridimensionales que sitúen a los sonidos en el espacio F1, F2, F3.

9.- Proyección de cada mapa tridimensional sobre los tres planos posibles (F1-F2, F1-F3, F2-F3).

10.- Agrupación y visualización de resultados.

La figura 1 presenta tres ejemplos de espectros obtenidos tras la consecución del paso 4. En este caso, se muestra en la gráfica superior izquierda los formantes realzados correspondientes a los sonidos 'apa', 'aβa', 'ama'. En la gráfica inferior se representa la secuencia 'aka', 'aya', 'aηa', y en la derecha: 'ata', 'aða', 'ana'.

De cada uno de los espectros con formantes realzados conseguidos en el paso 4, se seleccionan en la etapa 5 unos 15 instantes de tiempo significativos, lo que nos lleva a la obtención de unos 3600 valores espectrales de formantes (15 x 5 vocales x 4 grupos x 4 hablantes x 3 formantes).

El resultado final del trabajo se representa mediante visualizaciones de los sonidos sobre planos obtenidos tras realizar proyecciones que parten del espacio de los formantes F1, F2, F3.

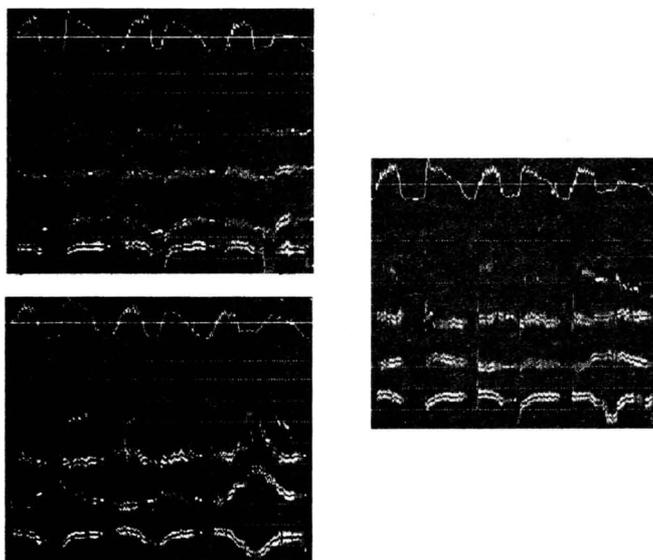


Figura 1

Espectro superior: 'apa aβa ama'. Espectro inferior: 'aka aya aηa'. Espectro de la derecha: 'ata aða ana'.

3. PRIMEROS RESULTADOS

En primer lugar explicaremos el significado de las figuras que vamos a presentar en este artículo. Como referencia tomamos la figura 2. El gráfico tridimensional situado en la esquina superior izquierda muestra con diferentes tonos de grises las posiciones de los formantes. Cada dimensión representa un formante (F1, F2, F3). El resto de gráficas de la figura se corresponden con las posibles proyecciones bidimensionales del gráfico tridimensional:

- Gráfico superior derecho: plano F2-F1 (F2 horizontal, F1 vertical)
- Gráfico inferior derecho: plano F2-F3 (F2 horizontal, F3 vertical)
- Gráfico inferior izquierdo: plano F1-F3 (F1 horizontal, F3 vertical)

Los formantes de esta figura de ejemplo pertenecen a diversos sonidos vocálicos situados en segmentos VCV pronunciados por un hablante masculino.

A la vista del plano F1-F2 (superior derecha) de la figura, se puede distinguir con facilidad las posiciones de los formantes de las cinco vocales castellanas en las realizaciones del habla que han sido tomadas para crear este ejemplo. De derecha a izquierda (de mayor a menor frecuencia en F2) observamos los sonidos 'i', 'e', 'a', 'o', 'u'.

Para comprender algo mejor como los gráficos representados expresan la posición (y en parte la evolución) de los formantes, tomemos como ejemplo y referencia la vocal 'a' reflejada en el plano F1-F2 (superior derecha) de la figura 3. En el caso de que la vocal sea anterior o posterior a una consonante bilabial (color gris oscuro), tanto el primero como el segundo formante bajan en frecuencias. Cuando la vocal es adyacente a una consonante velar (color negro), F1 baja en frecuencias, mientras que F2 sube. En el caso dental/interdental (el segundo tono de gris) los formantes están bastante estables, con tendencia a bajadas en F1 y subidas en F2. Este comportamiento

refleja los resultados esperados, que a modo de ejemplo podemos observar en la evolución de los formantes que muestra la figura 1.

En las figuras 4, 5 y 6 se agrupan resultados pertenecientes a dos hablantes masculinos. La figura 4, al igual que en cada caso individual, diferencia las vocales según el grupo de sus consonantes adyacentes.

En la figura 5 los resultados se agrupan por vocales. Se puede apreciar como, aunque la separabilidad no se pierde en el plano F1-F2, las posiciones de cada vocal pueden variar bastante al introducirse diferentes hablantes. De nuevo, los planos F1-F3 y F2-F3 pueden ayudar a realizar identificaciones en caso de dudas en el plano F1-F2.

La figura 6 representa cada hablante con un color; en este caso, el hablante identificado con el color gris claro presenta menores frecuencias en los tres formantes que el hablante identificado por el gris oscuro. Otra característica importante es la gran diferencia que existe en las posiciones de las vocales 'i', 'e', 'a', que denotan una articulación muy diferente entre estos dos hablantes en las posiciones no abocinadas de la boca.

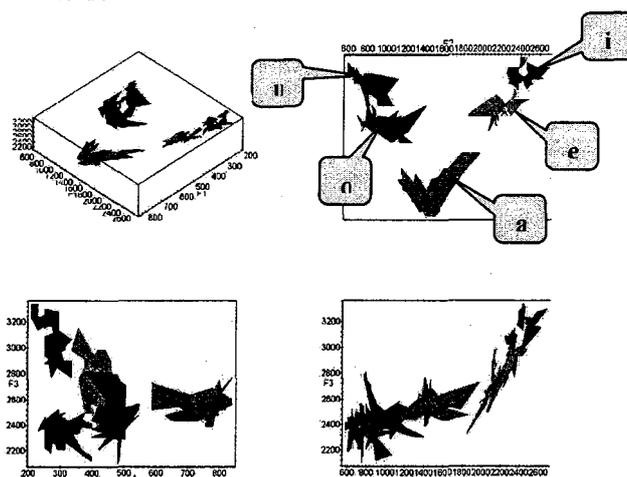


Figura 2

Ejemplo de posiciones frecuenciales de los tres primeros formantes en las vocales del castellano empleando un hablante masculino .

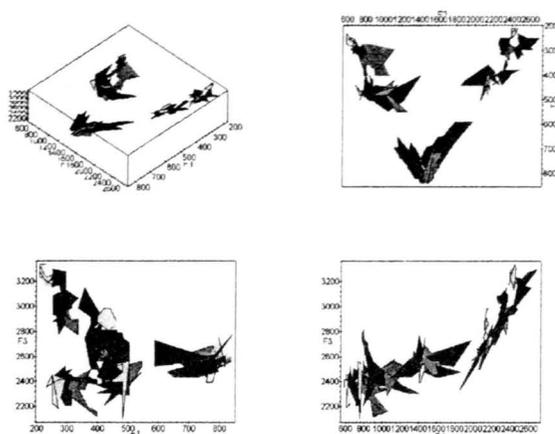


Figura 3

Ejemplo de posiciones frecuenciales de los tres primeros formantes obtenidos para un hablante masculino en vocales aisladas (el gris más claro), vocales adyacentes a consonantes bilabiales (gris oscuro), a consonantes dentales/interdentales (el segundo tono de gris), a consonantes velares (negro).

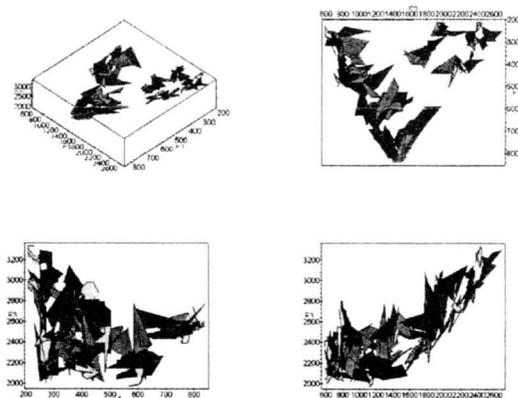


Figura 4

Ejemplo de posiciones frecuenciales de los tres primeros formantes obtenidos en dos hablantes masculinos sobre vocales aisladas (el gris más claro), vocales adyacentes a consonantes bilabiales (gris oscuro), a consonantes dentales/interdentales (el segundo tono de gris), a consonantes velares (negro).

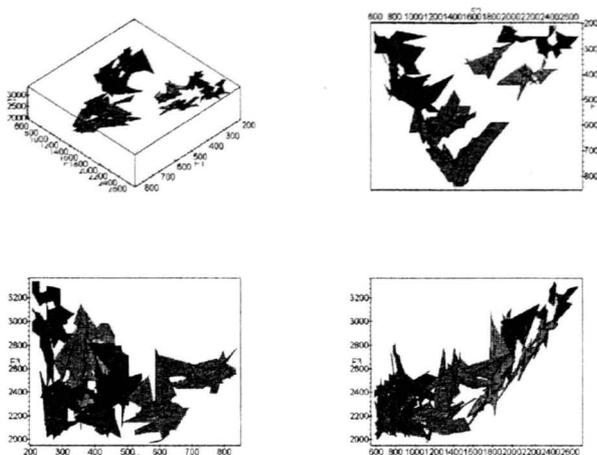


Figura 5
Ejemplo de posiciones frecuenciales de los tres primeros formantes en las vocales del castellano empleando dos hablantes masculinos .

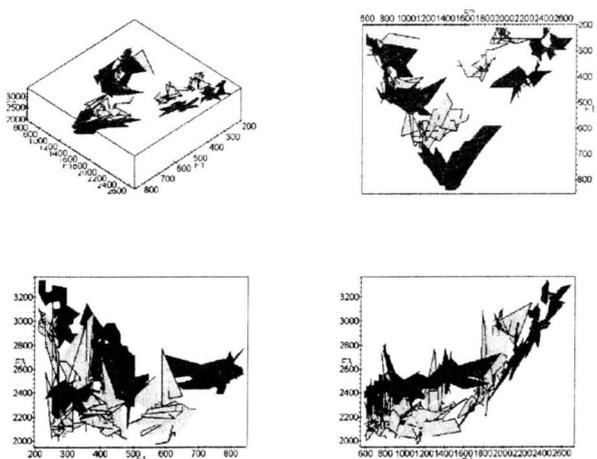


Figura 6
Posiciones de los formantes obtenidos en los dos hablantes masculinos.
Tono oscuro: hablante masculino 1. Tono claro: hablante masculino 2.

4. RESULTADOS CON CUATRO HABLANTES

La figura 7 presenta la posición de las vocales de los cuatro hablantes (2 masculinos y 2 femeninos) empleados en este estudio. La separabilidad ha disminuido debido a la variedad de las muestras obtenidas y sobre todo a la utilización de varios hablantes; sin embargo, todavía es posible distinguir con facilidad la mayor parte de los casos. Los planos en los que interviene el tercer formante ayudan a la identificación vocálica.

En la figura 8 se presentan las vocales diferenciadas por hablantes. Al haberse unificado los diferentes rangos de las escalas que abarca cada hablante por separado, el resultado ha sido que en los casos masculinos (con frecuencias más bajas) las vocales se han trasladado (comprimido) hacia los orígenes de frecuencias de cada plano representado.

El fenómeno expresado en el párrafo anterior, explica como es posible que, por ejemplo, la 'a' del hablante 'gris claro' se confunda con la 'o' de los demás. Sencillamente estamos comparando valores absolutos de frecuencias, cuando en realidad lo que nos interesa son sus posiciones relativas.

La figura 8 ilustra muy claramente la incidencia del tono de la voz del hablante en las posiciones absolutas de sus formantes. La relación es directa e incide en F1, F2 y F3. la solución a este problema consiste en aplicar un sencillo factor de corrección a los valores frecuenciales obtenidos si conocemos a priori las características de cada hablante, en caso contrario, este factor lo obtendríamos de una correcta estimación de la frecuencia fundamental del hablante.

La figura 9 presenta los planos F1-F2 correspondientes a las figuras 7 y 8, pero con características de visualización diferentes: puntos y contornos, esto permite hacerse una idea aproximada de cual es la posición media de los formantes, que por ejemplo, en el caso de la 'o', normalmente no tiene frecuencias tan altas en F2 como parecía en la figura 7.

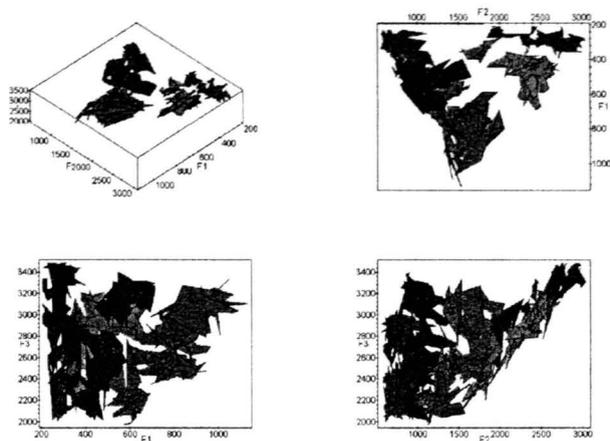


Figura 7

Ejemplo de posiciones frecuenciales de los tres primeros formantes en las vocales del castellano empleando a los 4 hablantes (2 masculinos y 2 femeninos)

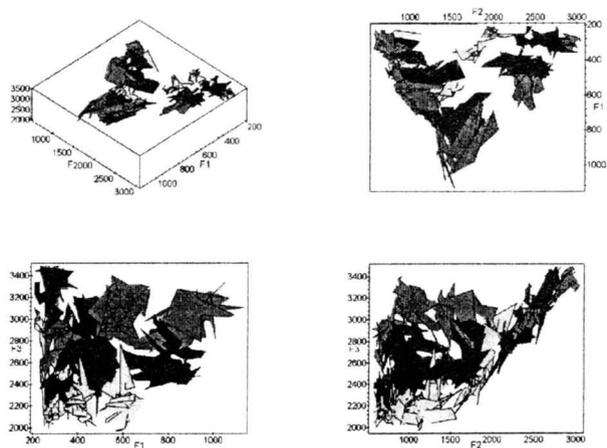


Figura 8

Posiciones de los formantes obtenidos empleando los 4 hablantes (2 masculinos y 2 femeninos). Segundo tono de gris: hablante femenino 1. Negro : hablante femenino 2. Gris oscuro : hablante masculino 1. Gris más claro : hablante masculino 2.

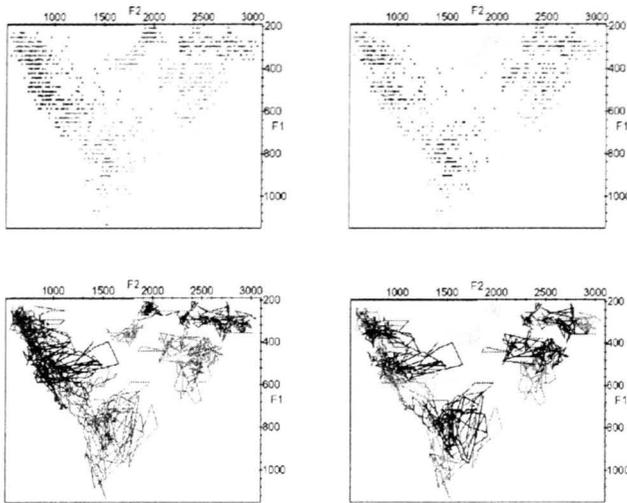


Figura 9
 Planos F1-F2 correspondientes a las figuras 7 y 8 presentados
 como nube de puntos y por superficies transparentes.

5. CONCLUSIONES

Es posible realizar una adecuada distinción vocálica acudiendo a comparaciones en el plano F1-F2 y apoyándose en caso de necesidad en los planos F1-F3 y F2-F3.

La representación visual de los formantes en planos de proyección (F1-F2, F1-F3, F2-F3), nos ofrece una visión complementaria a los espectros, que ayuda a clasificar e identificar los sonidos vocálicos.

La metodología propuesta se basa en clasificaciones de datos estáticos de formantes de voz, pero ayuda a la comprensión de la evolución de estos formantes.

Para poder caracterizar los espectros de voz, es necesario comprender la evolución de los formantes en las vocales adyacentes a los sonidos consonánticos, sin embargo, éstas evoluciones, aunque siguen unas reglas y pautas generales, presentan mucha variación entre distintas realizaciones del habla.

El tono de voz de cada persona influye directamente en la altura de los formantes que presentan los espectros, esto hace necesario una fase de normalización para conseguir una correcta caracterización de los parámetros espectrales de mayor interés.

Para poder realizar análisis representativos del habla, es necesario obtener, confrontar y comparar un gran número de grabaciones de voz de diferentes hablantes pronunciando sonidos básicos rodeados de diferentes contextos

6. BIBLIOGRAFÍA

- [HIL95] J. Hillenbrand, L.A. Getty, M.J. Clark, K. Wheeler, "Acoustic characteristics of American English vowels", *Journal of the Acoustic Society of America*, Vol. 97 (5), Mayo 1995, pp. 3099-3111
- [KAT95] W.F. Katz, P.F. Assmann, K.M. Jenouri, "Identification of natural and synthesized vowels produced by children and adults: Effects of fundamental frequency variation", *130th Meeting: Acoustic Society of America, Speech Communication: Studies of Voice*, 4pSC2, 1995

- [MAR94] E. Martínez Celdrán, *Fonética*, Martínez Celdrán E., Teide, 1994
- [MAR95] E. Martínez, R. Estapa, L. Romera, "La invariación acústica: Introducción", *Estudios de Fonética Experimental*, Vol 7, 1995, pp. 13-24
- [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
- [TOK93] S. Tokuma, "Some arguments on vowel formant shift", *Speech, Hearing and Language: Work in Progress*, UCL, Vol. 7, 1993, pp. 233-254