

CARTA DE FORMANTES Y TRIÁNGULO VOCÁLICO DEL CASTELLANO
EN EL PUERTO DE SAGUNTO.

José Romero, Juan Sánchez.

Laboratorio de Acústica
Departamento de Física Aplicada
Universidad de Valencia.

1. INTRODUCCIÓN

En este comunicado presentamos la carta de formantes y el triángulo vocálico de fonemas vocálicos tónicos en sílaba libre (A. Quilis 1), tanto de informantes masculinos y femeninos, residentes del Puerto de Sagunto, (Valencia), utilizando una tarjeta convertidor analógico digital y un programa de análisis de la voz elaborado en el Laboratorio de Acústica de la Universidad de Valencia.(20) y comparamos los resultados obtenidos con otros autores de la Lengua Castellana.

El avance de la electrónica digital en estos últimos años ha permitido avances espectaculares en los procesos de síntesis y reconocimiento de la voz. Sin embargo el reconocimiento de la voz y la síntesis son imposibles sin el estudio acústico previo y detallado de la lengua. Esta necesidad ha motivado estudios fonéticos tanto acústicos como fisiológicos en todos los idiomas. (3) (4). En particular, por lo que se refiere a la acústica del castellano, cabe destacar los estudios realizados por A. Quilis (1) A.M. Borzone (2), etc.

En el marco de realizar estudios fonéticos similares en las ciudades de la Comunidad Valenciana, presentamos los resultados obtenidos en esta investigación.

2. PROCESO DE LA FONACIÓN

La acústica describe el proceso del habla como la respuesta de un sistema de filtros sobre una fuente de sonido. El sonido se inicia por el paso estrecho de un flujo de corriente de aire procedente de los pulmones. Las cuerdas vocales regulan el paso de la citada corriente de aire mediante la abertura y cierre de la glotis. Los pulmones, la tráquea, la laringe, la faringe y las cavidades oral y nasal, constituyen el sistema de filtros de nuestro aparato fonatorio.

La fonación, comporta en dos etapas independientes. La primera tiene lugar en los pliegues vocálicos: el aire que sale de los pulmones excita los pliegues vocálicos y produce una sucesión muy rápida de aperturas y cierres de los mismos. El valor de la frecuencia F_0 de esta vibración de los pliegues vocálicos presenta pocas variaciones para las diferentes vocales, siendo más propio de la persona que de la vocal (éste es precisamente uno de los parámetros que nos permiten diferenciar entre vocales masculinas y femeninas). El dominio espectral glotal periódica: viene definida completamente por una serie de máximos que aparecen en la región de bajas frecuencias. Este sonido glotal sufre un proceso de modulación en su paso desde los pliegues vocálicos hasta los labios. Este proceso de modulación son diferentes para cada vocal, dado que la configuración del tracto vocálico articulatorio es diferente en cada caso.

Como consecuencia, el sonido producido por los pliegues vocálicos, prácticamente idéntico para todas las vocales, resulta ampliado o atenuado de forma diferente para unas frecuencias u otras en función de la vocal que se está articulando y, por lo tanto, el sonido que sale de los labios resulta diferente para cada vocal.

Desde el punto de vista acústico, las diferentes vocales se pueden caracterizar por los valores de las frecuencias, los anchos de banda y las intensidades de sus formantes (resonancias propias del tracto vocálico).

3. TRATAMIENTO DIGITAL DE LA SEÑAL

Al llegar la voz al micrófono provoca una señal analógica continua $f(t)$ cuyo valor en cada instante corresponde a la tensión producida en los bornes del micrófono. Dado que con este tipo de señales no se pueden utilizar los algoritmos implementables en un ordenador, para llevar a cabo nuestro trabajo es necesario convertir dicha señal analógica en una señal digital $s(n)$ constituida por una secuencia discreta de valores de la variable independiente n .

Este proceso anterior se lleva a cabo fácilmente mediante un conversor analógico-digital (A/D). Estos conversores poseen dos características fundamentales, que determinan la calidad del proceso de conversión: la frecuencia de muestreo de la señal analógica y los bits de resolución de la conversión.

La frecuencia de muestreo de dicho conversor debe ser ajustada de forma tal que en el proceso de conversión no se pierda información relevante. En nuestra investigación dicha frecuencia de muestreo ha sido fijada en 12.8 kHz lo que nos ha permitido analizar espectralmente las voces humanas hasta un fondo de escala de 6 kHz (el rango de frecuencias cubierto por este trabajo está comprendido entre 12.5 Hz y 5.0 kHz, un intervalo más que suficiente para registrar completamente los cuatro primeros formantes de todas las vocales).

Por otra parte, hemos optado por utilizar un conversor de 12 bits. Este conversor transforma la señal analógica en una señal digital cuyas muestras pueden tomar un total de 40960 valores diferentes, lo cual representa un nivel de resolución completamente satisfactorio para nuestros objetivos.

4. ANÁLISIS ESPECTRAL

La deducción de los valores de los parámetros vocálicos mencionados anteriormente se ha llevado a cabo utilizando dos de los algoritmos más corrientes en el tratamiento digital de señales, la Transformada Rápida de Fourier y la Predicción Lineal para la envolvente del espectro SPL(8) (9) (10). En particular, la Transformada Rápida de Fourier (FFT) se ha utilizado en el algoritmo que se conoce con el nombre de "Cepstrum"(12)-(18)

5. INFORMANTES Y PALABRAS UTILIZADAS

Se han analizado un total de 41 informantes, y se tomó nota sobre el nombre del Informante; Lugar de Grabación; Fecha; Edad; Sexo; Profesión; Nivel de estudios; Procedencia; Tiempo que vive en el Puerto de Sagunto. En la encuesta figuraban; 20 hombres y 21 mujeres. Sus edades estaban comprendidas entre 15 y 56 años. Las grabaciones se realizaron en el Puerto de Sagunto (zona de alta inmigración (21), (22). El número total de informantes considerado en esta investigación es del mismo orden que el utilizado en otros trabajos similares a estos (19) (23) y en cualquier caso resulta adecuado para un estudio general de la caracterización vocálica.

Previamente, para estudiar la variabilidad de un informante, se analizaron cinco fonemas vocálicos en el mismo contexto de palabras, entre dos consonantes oclusivas: bilabiales de forma átona la "p". Las palabras eran (*): papá, pepe, pipa, popa, pupa.

Cada informante leyó una lista de 31 palabras diferentes (1), y son las siguientes:

(i) vivo, Aquí, Amarillos, día, cerillas.

(e) bebe, cerveza, perro, negras, trece.

(a) baba, aparato, cava, cada, caza, gafas, pasas, rama, parra, tacha, vaya.

(o) bobo, yo, todo, ocho, doce.

(u) veintiuno, virtud, sepultura, puas, ruso.

La lectura se hizo de una forma secuencia, pausada y con la pronunciación lo más natural posible. Las vocales se presentan en diferentes contextos tónico y átono.

6. REALIZACIÓN Y PROCESADO DE LAS GRABACIONES

Las grabaciones se realizaron con una casete de mano con una cinta de audio de tipo II, se realizaron en condiciones usuales de habla: unas en una sala de grabación, otras en el interior de un coche, otras en una habitación de una casa, y otras en la calle. Teniendo en cuenta que el ruido era despreciable esto se comprobó al analizar el espectro de la señal, la relación (S/N).

7. RESULTADOS Y DISCUSIÓN

7.1 Variabilidad de la voz humana.

Con ello nos referimos al hecho de que una misma persona, cuando articula varias veces un mismo fonema, no obtiene resultados absolutamente idénticos, sino en cada caso el sonido producido es ligeramente diferente. Estas variaciones dependen de muchos factores: dificultad para pronunciar el sonido, condiciones ambientales, estado emocional y físico del informante, etc.

En la figura siguiente se puede ver los valores medios de las formantes de cada vocal

Se puede observar a partir de la tabla que la variabilidad es muy importante siendo la desviación típica relativa del 20% para la primera formante, del 5% para la segunda formante, del 3% para el tercero y el cuarto.

7.2. Formantes.

En la tabla siguiente se exponen los valores medios de las frecuencias centrales de las regiones de las formantes de todas las vocales de los informantes por sexos, así como su desviación típica relativa es del orden del 20% para todas las formantes. lo que hace pensar que a partir de la segunda formante sus valores son característicos de cada informante.

	F ₁	desv.tip.	F ₂	desv. tip.	F ₃	desv.tip.	F ₄	desv.tip.
a	708	106	1317	91	2669	40	3292	108
e	587	63	1715	71	2478	93	3461	87
i	467	60	1980	41	2831	94	3554	215
o	563	113	1198	48	2603	75	3253	111
u	360	50	985	46	2398	71	3344	92

También se han calculado los anchos de banda de las formantes, es decir, el ancho de banda para el que se presentan una caída de 3 dB, y los niveles relativos de los picos de las formantes.

Tablas de valores medios de las formantes y sus desviaciones típicas para cada vocal tomada en 10 palabras sueltas para la "a" y cinco para el resto de vocales siendo el número de informantes de 20, para cada sexo

HOMBRES								MUJERES									
	F ₁	d.t.	F ₂	d.t.	F ₃	d.t.	F ₄	d.t.		F ₁	d.t.	F ₂	d.t.	F ₃	d.t.	F ₄	d.t.
a	656	82	1350	140	224	329	3230	512	a	691	219	1402	359	2337	3183	3183	600
e	491	96	1550	307	2346	398	3416	529	e	597	103	1680	465	2517	3385	3385	563
i	413	79	1900	317	2743	397	3560	479	i	470	103	2099	642	3020	3844	3844	516
o	470	110	1125	287	2073	453	2999	554	o	567	125	1251	347	2221	3131	3131	685
u	439	120	1139	293	2112	528	3085	585	u	518	130	1204	310	2289	3341	3341	646

7.3. Triángulo vocálico por sexos.

Estos corresponden a la representación cartesiana de la segunda formante, en función del primero, y es la forma más habitual de caracterizar las vocales de un idioma. A partir de ellas se puede reconocer la gran variabilidad en la nube de puntos para cada vocal, e incluso en el solapamiento de ellas. Las circunferencias corresponden cada una a una vocal según se indica, la nube de puntos cuyo radio es aproximadamente la desviación típica y cuyo centro es el valor medio.

8. COMPARACIÓN CON RESULTADOS DE OTROS AUTORES

Con ánimo de comparar con los trabajos de otros autores (23) (24) se representan en la tabla siguiente las formantes primera y segunda obtenidas por: (3) (4) (5) (6) (7) para cada sexo.

Los valores obtenidos por otros autores se encuentran dentro de la dispersión obtenida.(con la cadena de grabación que se ha dispuesto es suficiente para este trabajo)

vocal	Quilis	Andreu	Celdrán	J.L. Ramon	Nuestras	Sánchez
a	688	754	650	733(114)	656(82)	708(106)
e	316	497	451	465(88)	491(96)	587(63)
i	210	346	347	278(78)	413(79)	467(60)
o	332	571	496	510(116)	470(110)	563(113)
u	218	418	485	306(122)	439(120)	360(50)

vocal	Quilis	Andreu	Celdrán	J.L. Ramon	Nuestras	Sánchez
a	1218	1328	1300	1203(178)	1350(140)	1317(91)
e	2090	1930	1847	1972(246)	1550(307)	1715(71)
i	2519	2146	2193	2254(318)	1900(317)	1980(41)
o	931	1176	1007	909(116)	1125(287)	1198(48)
u	721	9254	1008	774(124)	1139(293)	1198(48)

Los números entre paréntesis son desviaciones típicas, todos los informantes eran hombres excepto para las de Quilis que son promediados hombres y mujeres; todas corresponden a vocales tónicas en sílaba libre excepto las de J.L. Ramón que son vocales sueltas, y las de J. Sánchez que son átonas.

Podemos concluir diciendo: que la variabilidad en un informante es grande, siendo la desviación típica de un determinado informante del 20% para la primera formante y menor del 5% para las tres siguientes formantes. La desviación típica relativa para un grupo de informantes del Puerto de Sagunto es del orden del 20%. Esto hace pensar que a partir de la tercera formante son valores característicos de la persona. En cuanto a la diferencia entre sexos, por causa de la gran dispersión, se pueden considerar iguales, lo que nos llevaría a confeccionar un sólo triángulo vocálico para ambos. Lo mismo ocurre con la diferencia tónica átona, y como cabría esperar no hay una variación en las formantes que las distinga. Los valores de la primera formante de la "i" y de la "u" obtenidos en el Puerto de Sagunto son algo altos comparados con los otros autores, aunque estos no citan la dispersión. Tampoco recogen resultados acerca de las tercera y cuarta formante. Otro resultado obtenido al comparar las formantes de las vocales en nuestro trabajo es que la "o" y la "u" se solapan en todos las formantes

9. BIBLIOGRAFÍA

- (1) Quilis. Fonética Acústica de la Lengua Española
- (2) M. A. Borzone. Manual de Fonética acústica. Hachette (1980)
- (3) G. Fant. "Acoustic theory of speech production". Mouton. The Hague (1960)
- (4) J.L. Flanagan. "Speech analysis, synthesis and perception". Springer-Verlag. Berlin (1972).
- (5) R. Cerdá. "El timbre vocálico en catalán". C.S.I.C. Madrid (1972)
- (6) J. Martí. "Parámetros vocálicos del catalán". Folia Phonetica 1, 23 (1984).
- (7) J. Martí. "Estudi acústic del català i síntesi automàtica per ordinador". Tesis Doctoral. Universidad de València (1986).
- (8) L. R. Rabiner y R. W. Schafer. "Digital processing of speech signals". Prentice Hall. New York (1978).
- (9) J. Markhoul. "Spectral linear prediction: properties and applications". IEEE Transactions A.S.S.P.23.283(1975).
- (10) A. J. Markel y A. H. Gray. "Lineal prediction of speech". Springer-Verlag. Berlin (1980)
- (11) A. M. Noll. "Cepstrum pitch determination". Journal Acoust. Soc. Amer. 41. 293 (1967).
- (12) M. M. Sondhi. "Estimation of vocal tract areas: the need for acoustical measurement". IEE Transactions ASSP.27. 268(1979).
- (13) H. Wakita. "Estimation of vocal tract shapes from acoustical analysis of the speech wave: the state of art". IEE Transactions A.SSP.27.281(1979).
- (14) F. Lonchamp. J.p. Lefebvre y J. P. Zerling. "Comparaison de deux méthodes d'estimation des fonctions d'aires vocaliques". XI Congrès International d'Acoustique. 4. 23(1983)
- (15) D.E. Newland. Vibraciones aleatorias y análisis espectral. A. C.(1983)
- (16) A. Michael Noll. Cepstrum pitch determinación. JASA. Vol. 41 núm.2 (1976) pág 283-296.
- (17) Programs for D.S.P. Digital Signal Processing Committee. IEEE. A.S.S.P. Society. IEEE Press.
- (18) J.S Lienard. "Les processus de la communication parlée". Ed Masson. (1977).
- (19) M. Morse. "Vibration and sound". 2nd. Ed. Mc. Graw Hill. (1948)
- (20) Luís Faus, Amado García. Análisis Acústico y Articulatorio del Vocalismo Valenciano. Rev de Acústica 1992
- (21) José Dionisio Martínez "Algunos aragonismos léxicos en el habla castellana".
- (22) Josep-Dionis Martínez. "Valencianismes lèxics en la Parla Valençiana del Port de Sagunt. -Breu estudi poblacional y lingüística". Revista Braçal. 1990.
- (23) J.L. Ramón Et Al . Método de estudio de las frecuencias fundamentales y análisis espectrográfico de los dos primeros formantes de las vocales castellanas. Acta Otorrinolaring Esp. 30. 399-414-1979.
- (24) M. Celdran E 1981. Fonética Ed. Teide.