



FIA 2018

**XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18-
24 al 26 de octubre**

EXTRACCIÓN DE LA CURVA DE TONO DE UN CANTANTE MEDIANTE COMBINACIÓN DE ALGORITMOS CLÁSICOS

PACS: 43.75.+a.

García Barrios, Guillermo¹; Sáenz Lechón, Nicolás¹; Gutiérrez Arriola, Juana M^a¹; Pedrero González, Antonio²; Fraile Muñoz, Rubén¹; Gómez Alfageme, Juan José¹; Osma Ruiz, Víctor¹
¹CITSEM de la Universidad Politécnica de Madrid (UPM). ²Grupo de Acústica Arquitectónica Edificio La Arboleda. Campus Sur. Calle Alan Turing 3. 28031 Madrid.

Madrid

España

+34914524900 ext 20787

garciabarrios.g12@gmail.com

Palabras Clave: canto, frecuencia fundamental, tono, análisis de la señal de voz

ABSTRACT

The algorithms of automatic extraction of the fundamental frequency contour often present errors that must be corrected with a subsequent manual or automatic processing. In this communication we present a system for extracting the fundamental frequency contour using the results of six classical algorithms that are combined to avoid errors. The algorithm has been proven to extract the fundamental frequency contour of a singer performing a song and results are satisfactory.

RESUMEN

Los algoritmos de extracción automática de la curva de tono presentan errores que, a menudo, deben corregirse con un procesamiento posterior manual o automático. En esta comunicación se presenta un sistema de extracción de la curva de tono utilizando el resultado de seis algoritmos clásicos que se combinan para evitar errores y proporcionar los valores más fiables de los seis que se han calculado. El algoritmo se ha probado para extraer la curva de tono de un cantante interpretando una canción y los resultados son satisfactorios.

1. INTRODUCCIÓN

El análisis de la frecuencia fundamental de la señal cantada es de interés en varias aplicaciones como: separación de canto, clasificación del género musical, síntesis de canto, transcripción automática y estudio de las diferencias de interpretación entre cantantes [1], [2]. La extracción

FIA 2018

XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18-
24 al 26 de octubre

de la frecuencia fundamental ha sido ampliamente estudiada para la señal de voz hablada y la traslación de estos algoritmos a la voz cantada presenta sus particularidades. En [1] se comparan seis algoritmos para la extracción de la frecuencia fundamental y se concluye que no hay algoritmo que de un resultado óptimo en todos los aspectos. El algoritmo RAPT es el que tiene menor número de errores graves, es decir, cuando la frecuencia detectada varía en más de un semitono respecto a la real. El algoritmo de Praat es el que mejor determina los límites de las zonas sonoras y el algoritmo YIN es el que tiene menor error en la detección de la frecuencia fundamental cuando no hay errores graves.

Teniendo en cuenta estos resultados y los algoritmos de extracción de frecuencia fundamental implementados por el grupo de investigación se propone realizar el análisis de la frecuencia fundamental de un cantante interpretando una canción sin hacer una adaptación específica a la señal cantada. Posteriormente se va a realizar un postproceso de los resultados obtenidos de seis algoritmos clásicos para obtener el contorno final.

2. MATERIALES

La base de datos utilizada corresponde a la descrita en [3], donde seis cantantes de coro fueron grabados individualmente cantando ocho canciones pertenecientes al repertorio de canto mozárabe. Las grabaciones se realizaron en una cámara anecoica con un par de micrófonos de alta calidad y bajo ruido. La señal de audio se digitalizó con una frecuencia de muestreo de 44.1kHz.

3. MÉTODOS

El algoritmo que se utiliza para extraer la curva de tono o frecuencia fundamental tiene los siguientes pasos:

- a) Detección de zonas sonoras
- b) Extracción de la frecuencia fundamental (F0) con seis algoritmos.
- c) Postprocesado de los resultados para obtener el contorno de F0

3.1 Detección de zonas sonoras

Para detectar las zonas sonoras se realiza primero una detección de voz siguiendo el algoritmo descrito en [4] que realiza una normalización, una compresión para mejorar la detección cuando hay ruidos cortos de gran amplitud, un filtrado para quitar la componente de continua y el ruido de altas frecuencias y un enventanado. Para cada ventana de voz se calcula el parámetro [4]:

$$COPER = \sum_{n=1}^N |x[n] \cdot |x[n]| - x[n-1] \cdot |x[n-1]|| \quad (1)$$

Se define un umbral que se puede ajustar manualmente de forma que si el parámetro supera el umbral hay voz.

FIA 2018

XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18-
24 al 26 de octubre

Posteriormente, en las zonas donde se ha detectado voz se realiza una división en zonas sordas y sonoras basada en el coeficiente de autocorrelación y en la relación de energía en baja y alta frecuencia.

En la figura (1) se observa el resultado de la aplicación del algoritmo de detección de zonas sonoras a un segmento de una canción.

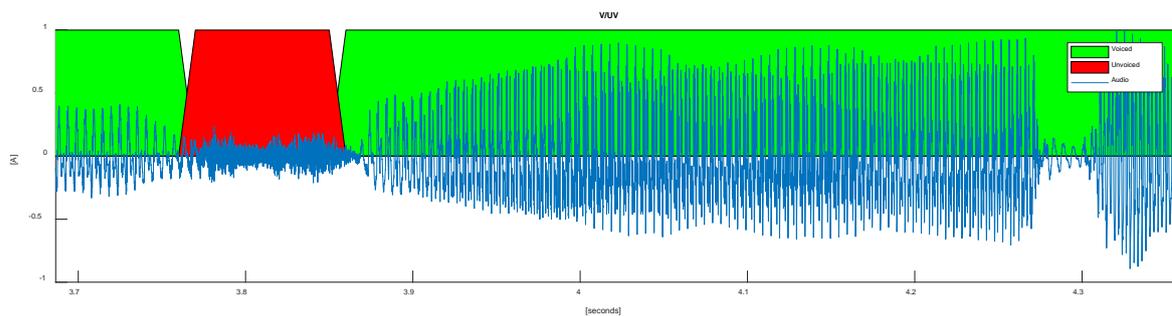


Figura 1. Resultado de aplicar el algoritmo de detección sordo/sonoro a un segmento de una canción. En verde las zonas sonoras y en rojo las zonas sordas.

3.2 Extracción de la frecuencia fundamental

Para extraer la frecuencia fundamental se utilizan seis algoritmos:

- Algoritmo descrito por Rabiner y Schafer en [5]. A partir de la autocorrelación del residuo LPC (Linear Prediction Coefficients) de la señal se calcula el periodo fundamental. Se divide la señal en ventanas de 60ms con un desplazamiento de 30ms.
- Algoritmo descrito por Rabiner y Schafer en [6]. Calcula la autocorrelación de la señal recortada mediante un recorte estático. La primera ventana tiene una duración de 34ms y el resto es cinco veces T_0 , siendo T_0 el periodo fundamental detectado. No hay solapamiento entre ventanas.

FIA 2018

XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18-
24 al 26 de octubre

- Algoritmo descrito por Rabiner y Schafer en [7]. Utiliza el cepstrum para obtener el periodo fundamental. Se usan ventanas de 40ms con un desplazamiento de 20ms.
- Algoritmo utilizado por Kasuya et al. en [8]. Utiliza la autocorrelación de la señal recortada para obtener tres posibles candidatos a periodo fundamental del segmento. Se realiza un postprocesado para el elgir el mejor candidato. Se utilizan ventanas de 40ms con un desplazamiento de 20ms.
- Algoritmo propuesto por Boyanov et al. en [9]. Se calcula el periodo fundamental en tres dominios: temporal, frecuencial y cepstral y se combinan los tres valores para obtener el resultado final. La duración de las ventanas es de tres veces el periodo detectado en el segmento anterior y no hay solapamiento entre ventanas. La primera ventana tiene una duración fija de 34ms.
- Algoritmo descrito por Childers en [10]. Se trabaja sobre el cepstrum del residuo de la predicción LPC de la señal. La duración de las ventas es de 34ms con un desplazamiento de 17ms.

3.3 Postprocesado

Una vez obtenido el tono fundamental con los seis algoritmos se modifican los resultados para tener un valor del tono cada 10ms, se repiten o eliminan valores según sea necesario. A continuación se obtiene la inversa del tono fundamental para obtener la frecuencia fundamental. Los resultados se almacenan en una matriz de seis filas como se muestra en la figura 2.

1	190.0862	190.0862	188.4615	188.4615	188.4615	188.4615	188.4615	188.4615	178.5425	178.5425	178.5425	178.5425	178.5425	178.5425
2	188.4615	188.4615	188.4615	188.4615	188.4615	188.4615	180	180	180	180	180	180	202.2936	202.2936
3	191.7391	189.2704	189.2704	189.2704	189.2704	189.2704	185.2941	185.2941	185.2941	166.4151	166.4151	166.4151	0	0
4	189.2704	189.2704	188.4615	188.4615	188.4615	188.4615	188.4615	188.4615	0	0	0	0	0	0
5	191.7391	190.0862	189.2704	189.2704	188.4615	189.2704	189.2704	177.8226	0	0	0	0	0	0
6	190.9091	190.0862	190.0862	190.0862	190.0862	190.0862	0	0	0	0	0	0	0	0

Figura 2. Resultado de los algoritmos de extracción de frecuencia fundamental. Cada fila corresponde al valor obtenido por un algoritmo y cada columna a un instante de tiempo, separados 10ms.

FIA 2018

XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18-
24 al 26 de octubre

El valor 0 que se observa en algunas celdas de la matriz significa que el algoritmo no ha detectado frecuencia fundamental en ese segmento de la señal.

Al observar los resultados se puede aseverar que todos los algoritmos presentan errores debido a distintas circunstancias. En la figura 3 se muestra un tramo de señal en la que se pueden observar errores en todos los algoritmos aplicados.

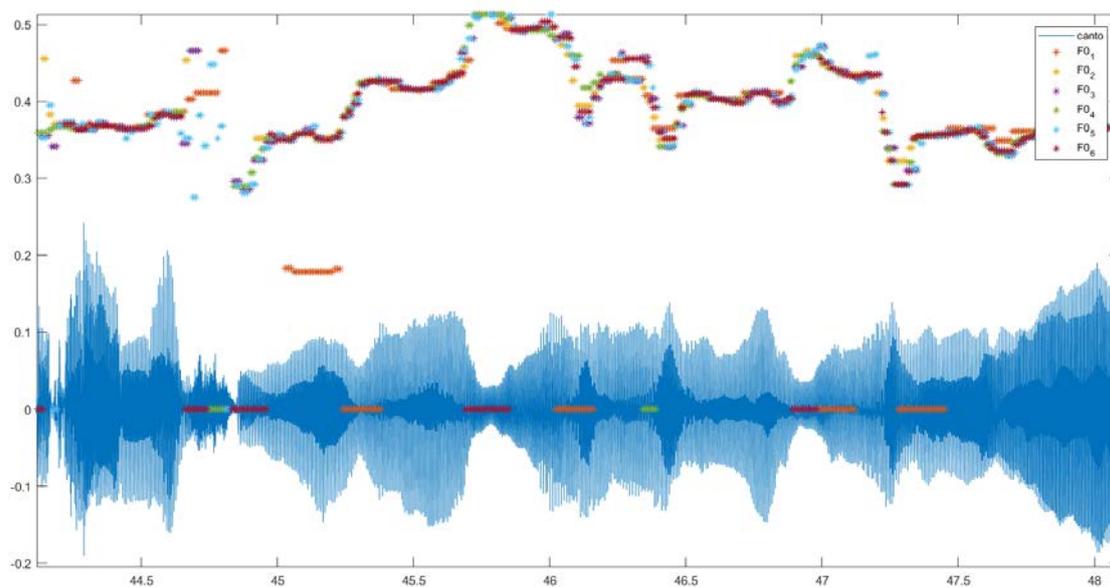


Figura 3. Resultado de los algoritmos de extracción de frecuencia fundamental. Los valores de F0 están normalizados a 400Hz para poder representarlos junto con la señal de audio.

Para corregir los errores se obtiene la mediana en cada instante de tiempo y el resultado se muestra en la figura 4.

FIA 2018

XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18-
24 al 26 de octubre

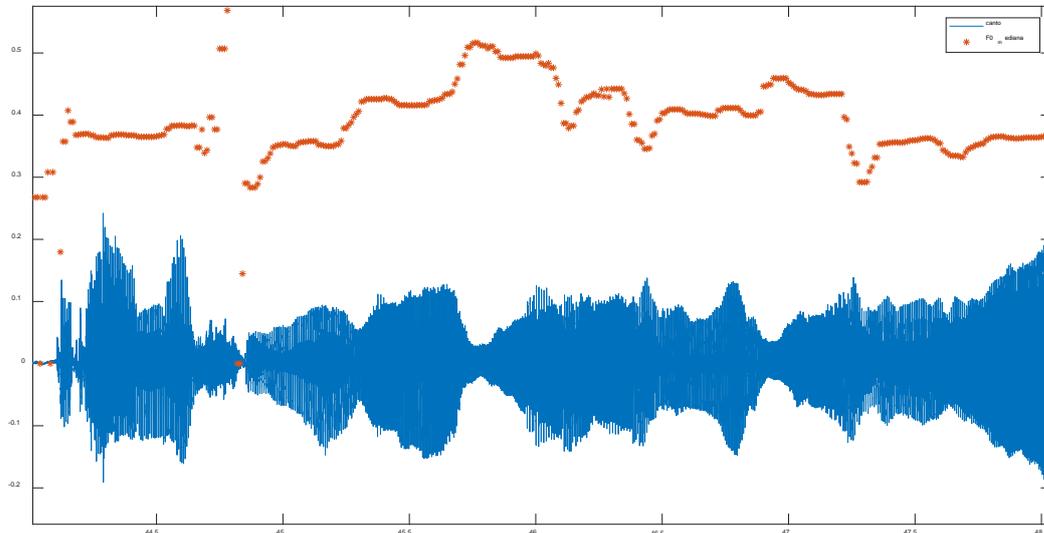


Figura 4. Contorno de F0 después de haber calculado la mediana de los seis valores obtenidos por los algoritmos. El valor de F0 está normalizado a 400Hz para poder representarlo junto con la forma de onda.

Finalmente se pone a 0 el tono de los segmentos sordos de acuerdo a la detección sordo/sonoro detectada en el primer paso.

4. CONCLUSIONES

En esta comunicación se ha presentado un método de extracción de tono para audios de canto mozárabe basado en algoritmos clásicos de extracción de frecuencia fundamental en habla. No se han adaptado los parámetros de los algoritmos sino que se han combinado los resultados de seis algoritmos para obtener una curva de tono libre de fallos.

Los resultados se han utilizado para modificar las curvas de tono, intercambiarlas entre cantantes y sintetizar un nuevo audio y no se han detectado errores graves.

5. REFERENCIAS BIBLIOGRÁFICAS

[1] Babacan, O.; Drugman, T.; d'Alessandro, N.; Henrich, N.; Dutoit, T. "A comparative study of pitch extraction algorithms on a large variety of singing sounds" Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, pp7815-7819, Vancouver, Canada, May, 2013



FIA 2018

XI Congreso Iberoamericano de Acústica; X Congreso Ibérico de Acústica; 49º Congreso Español de Acústica -TECNIACUSTICA'18- 24 al 26 de octubre

[2] Hsu, C.L.; Chen, L.Y.; Jang, J.S.R.; Li, H.J. "Singing pitch extraction from monaural polyphonic songs by contextual audio modeling and singing harmonic enhancement" Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), pp201-206. Kobe, Japan. October 2009.

[3] Pedrero, A.; Pollow, M.; Dietrich, P.; Behler, G.; Vorländer, M.; Díaz, C.; Díaz, A. Mozarabic Chant anechoic recordings for auralization purposes. Proceedings of the VIII Congreso Iberoamericano de Acústica, Évora, Portugal, October 1-3, 2012. En CD-ROM.

[4] PERALTA REYES, F.; COTRINA ATENCIO, A. *Algoritmo Coper para la detección de actividad de voz*. Facultad de Ingeniería Electrónica (UNMSM), 2002.

[5] Rabiner, L.R.; Schafer, R.W. Digital processing of speech signals, Prentice Hall, NJ, 1978, pp. 372-378, pp. 150-158, pp. 447-450

[6] Rabiner, L.R.; Schafer, R.W. Digital processing of speech signals, Prentice Hall, NJ, 1978, pp. 372-378, pp. 150-158, pp. 150-158

[7] Rabiner, L.R.; Schafer, R.W. Digital processing of speech signals, Prentice Hall, NJ, 1978, pp. 372-378, pp. 150-158, pp. 372-378

[8] Kasuya, H.; Ogawa, S.; Kikuchi, Y.; Ebihara, S. An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology", Speech Communication, vol. 5, 1986, pp. 171-81

[9] Boyanov, B.; Hadjitodorov, S.; Doskov, D.; Teston, B. Robust hybrid pitch detector for pathological voice analysis. Proceedings of Larynx'97, Marseille, France, 1997. Pp. 55-58

[10] Childers, D.G. Speech processing and synthesis toolboxes, John Wiley & Sons, Inc., NY, 2000

[11] MATLAB Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States