

Informe de resultados: señales reales

Leandro Perren - Matias Meurzet

5 de noviembre de 2015

1. Introducción

En este documento se presentan los resultados de obtener señales grabadas en ambientes reales cotidianos, tanto de ronquidos como de otros sonidos (aquí los definiremos como ruidos) que no son ronquidos y que pueden estar presentes en las señales obtenidas en entornos reales cotidianos, como por ejemplo:

1. Ambientales
2. Tos, estornudos
3. Vehículos
4. Voz
5. Música

El objetivo es poder evaluar características temporales (energía, zcr, rms, factor cresta) en las señales grabadas en entornos reales cotidianos. Una vez extraídas las características necesarias se podrá, entrenar y evaluar distintos modelos/clasificadores con la intención de construir una aplicación para dispositivos móviles capaz de detectar episodios de ronquidos y diferenciarlos de otras interferencias (tos, vehículos, voces, música, etc.) que pudieran acoplarse en la grabación del sueño de la persona.

2. Procedimiento

Para realizar las tareas antes mencionadas se procedió primeramente a grabar tanto ronquidos como ruidos por medio de los micrófonos de dispositivos móviles con sistema operativo Android (Tablet y celular) y utilizando la aplicación llamada «Easy Voice Recorder». Estas señales se grabaron con una frecuencia de 16000Hz, luego se utilizó el software Audacity para editar dichas señales extrayendo manualmente episodios de ronquidos como así también ruidos. En el caso de estos últimos se grabaron y editaron manualmente distintos posibles ruidos que podrían presentarse durante una grabación en ambientes normales. Una vez obtenidas estas señales se extrajeron las características temporales (energía,

zcr, rms, factor cresta, energía/zcr), tanto de los ronquidos como de los ruidos para poder obtener resultados y estadísticas que nos permitan evaluar las mismas como características representativas. Luego se generó una base de datos de características con las que se entrenarán los clasificadores.

2.1. Análisis de ronquidos

Durante el proceso de obtención de las señales (ronquidos) se llevaron a cabo distintas extracciones, las cuales fueron enumeradas como extracción 1 y 2, cada una de estas extracciones consiste de una grabación de una persona mientras duerme con el fin de captar en ella los ronquidos que pudieran producirse. Posteriormente cada una de estas extracciones fue analizada manualmente mediante el software Audacity y se extrajeron algunos segmentos de ronquidos presentes en la grabación creando distintos archivos llamados «ronquido_16000Hz_nro» todos de formato .wav los cuales contienen un episodio de ronquido cada uno.

1. Datos de la extracción 1 Esta primera extracción consiste de una grabación con una frecuencia de muestreo de 16000Hz de una persona de sexo masculino. De la misma se extrajeron 5 episodios los cuales fueron guardados en formato .wav y fueron nombrados como «ronquido_16000Hz_1», «ronquido_16000Hz_2», «ronquido_16000Hz_3», «ronquido_16000Hz_4» y «ronquido_16000Hz_5».
2. Datos de la extracción 2 Esta segunda extracción consiste de una grabación con una frecuencia de muestreo de 16000Hz de una persona de sexo masculino. De la misma se extrajeron 3 episodios los cuales fueron guardados en formato .wav y fueron nombrados como «ronquido_16000Hz_6», «ronquido_16000Hz_7» y «ronquido_16000Hz_8».

2.2. Análisis de ruidos

De igual manera a lo realizado en el caso de los ronquidos para la obtención de ruidos se llevaron a cabo extracciones que consistieron en grabar una serie de sonidos ambientales que no son ronquidos y que comunmente se pueden encontrar en señales grabadas en entornos cotidianos. Se realizaron grabaciones a una persona hablando en una radio, una persona tosiendo y el sonido de un telefono. De estas grabaciones se extrajeron fragmentos de dichos sonidos y se guardaron los mismos con los siguientes nombres, «ruido_16000Hz_habla», «ruido_16000Hz_tos», «ruido_16000Hz_telefono».

3. Extracción de características

Estos archivos creados cuyo contenido son ronquidos y otros sonidos fueron utilizados para extraerles características temporales con el fin determinar si las mismas son representativas de los ronquidos y que nos permitan diferenciarlos de otros sonidos (voz, musica, tos, vehiculos, etc). Las características temporales

extraídas en este trabajo son energía, zcr, rms, factor cresta, energía/zcr. Estas características fueron calculadas tomando para cada archivo ventanas deslizantes de 100ms cada una con un solapamiento del 50 %. A continuación se detalla la manera en que se calcula cada una de estas características y los resultados obtenidos para cada una de ellas.

3.1. Energía

El cálculo de la energía se realizó de acuerdo a la siguiente fórmula:

$$e_n = \sum_{i=1}^N x_i^2 / N \quad (1)$$

Donde e_n representa la energía del frame n correspondiente a una ventana de 100ms, x_i representa la amplitud de la muestra i-esima del frame n y N es la cantidad de muestras que tiene el frame en cuestión.

Como se puede observar en las figuras 1 y 2 se calculó la energía para los cinco ronquidos de la extracción 1 (ver Figura 1.) y para los 3 de la extracción 2 (ver Figura 2.), correspondientes ambos a una persona de sexo masculino. En la figura 3 se puede observar los gráficos de la energía para las señales de sonidos no deseados, las cuales corresponden a una persona tosiendo, un sonido de teléfono y por último el sonido de personas hablando. En todos estos gráficos se representó la energía para cada frame usando ventanas deslizante de 100ms con solapamiento de 50 %.

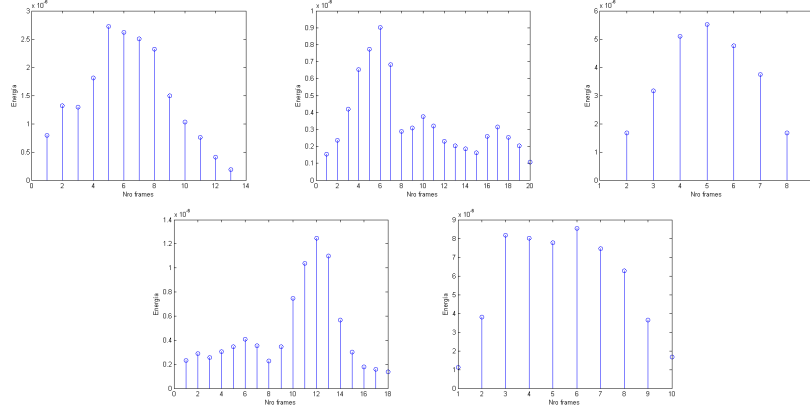


Figura 1: Energía de los ronquidos de la extracción 1

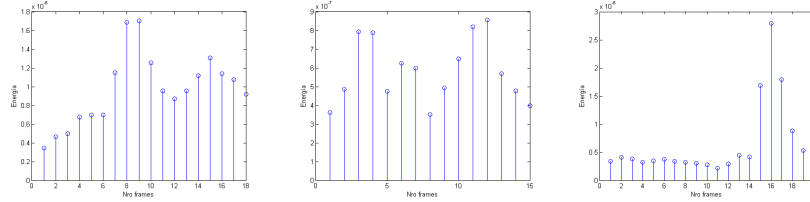


Figura 2: Energía de los ronquidos de la extracción 2

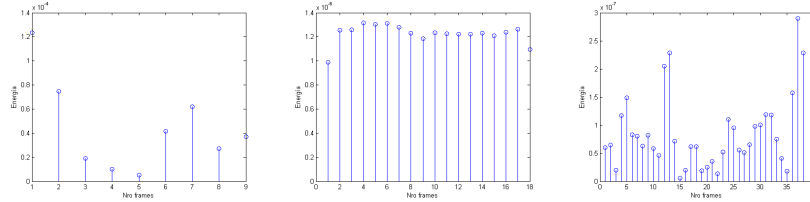


Figura 3: Energía de los ruidos: tos, telefono, habla

3.2. ZCR

La tasa de cruces por ceros (zero-crossing rate) es una medida de la cantidad de cambios de signos que sufre una señal, o lo que es igual el numero de veces que la señal toma el valor cero, y por lo tanto nos sirve de medida de la frecuencia que tiene la misma. El calculo del ZCR se realizó usando la siguiente formula:

$$zcr = \sum_{i=2}^N x_i * x_{i-1} \quad (2)$$

En las figuras 4 y 5 se encuentran graficados los resultados del calculo del ZCR para los cinco ronquidos de la extracción 1 (ver Figura 4.) y para los 3 de la extracción 2 (ver Figura 5.). En la figura 6 se pueden observar los gráficos

del ZCR para las señales de sonidos no deseados, las cuales corresponden a una persona tosiendo, un sonido de teléfono y por último el sonido de personas hablando. En todos estos gráficos se representó el ZCR para cada frame usando ventanas deslizante de 100ms con solapamiento de 50 %.

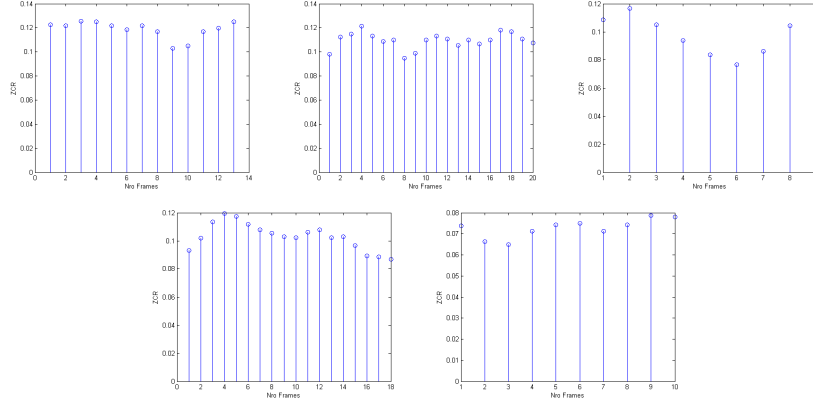


Figura 4: ZCR de los ronquidos de la extracción 1

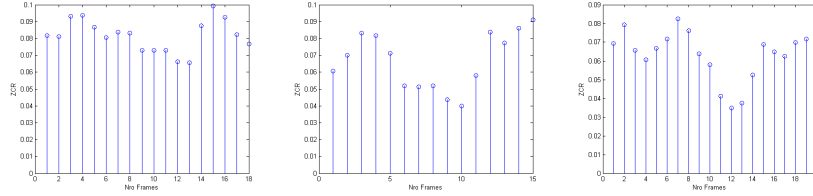


Figura 5: ZCR de los ronquidos de la extracción 2

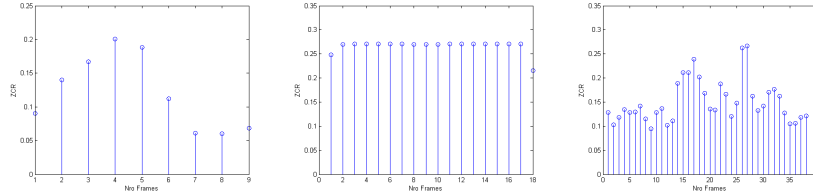


Figura 6: ZCR de los ruidos: tos, telefono, habla

3.3. Root Mean Square

El valor Root Mean Square se calcula tomando la raíz cuadrada de la media aritmética de los valores al cuadrado de cada muestra. Nos permite obtener un valor promedio de la señal sin tener en cuenta los efectos del signo de las muestras.

$$rms_n = \sqrt{\sum_{i=1}^N x_i^2 / N} \quad (3)$$

Donde rms_n representa el RMS del frame n correspondiente a una ventana de 100ms, x_i representa la amplitud de la muestra i-esima del frame n y N es la cantidad de muestras que tiene el frame en cuestión.

En la figura 7 y 8 se pueden observar los resultados del calculo del RMS para los ronquidos de las extracciones 1 y 2, y en la figura 9 podemos observar lo mismo para los tres ruidos.

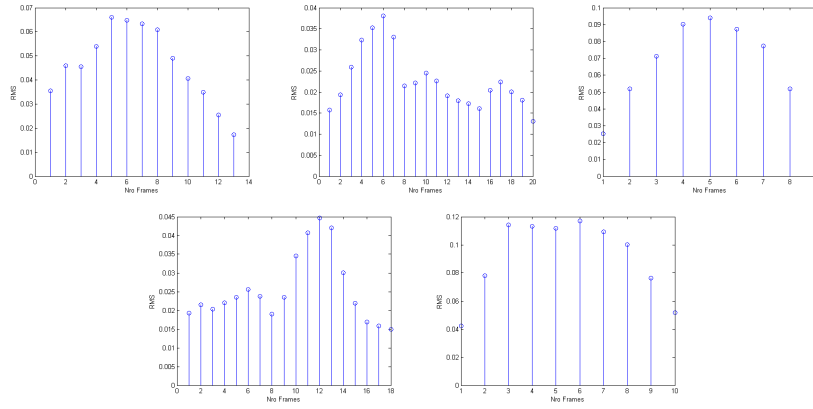


Figura 7: RMS de los ronquidos de la extracción 1

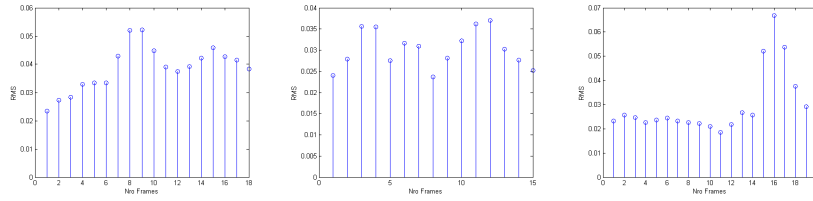


Figura 8: RMS de los ronquidos de la extracción 2

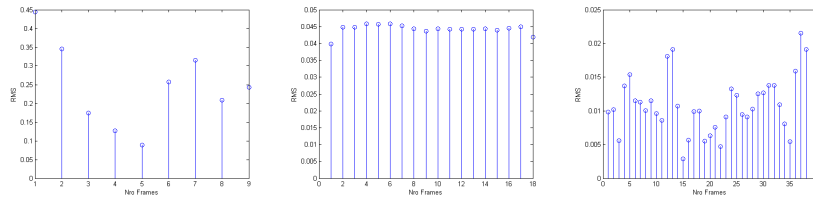


Figura 9: RMS de los ruidos: tos, telefono, habla

3.4. Factor cresta

El factor cresta es una medida obtenida haciendo el cociente entre el máximo valor absoluto de una señal dividido por su root mean square, dando un valor mayor en aquellas señales que tienen una serie de valores picos donde las magnitudes en esos valores tienden a ser mucho mayores que en el resto.

$$cresta_n = \max |x_n|/rms_n \quad (4)$$

Donde x_n representa cada muestra del frame n y rms_n es el root mean square del frame n.

En las figuras 10 y 11 se pueden observar los resultados del calculo del Factor Cresta para los ronquidos de las extracciones 1 y 2, y en la figura 12 podemos observar lo mismo para los tres ruidos.

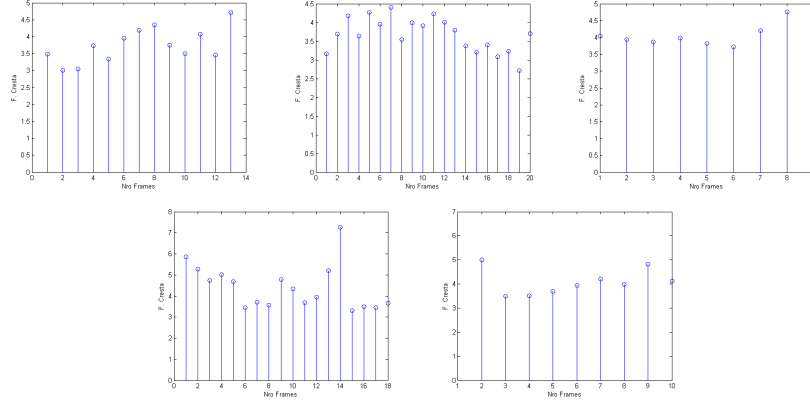


Figura 10: Factor cresta de los ronquidos de la extracción 1

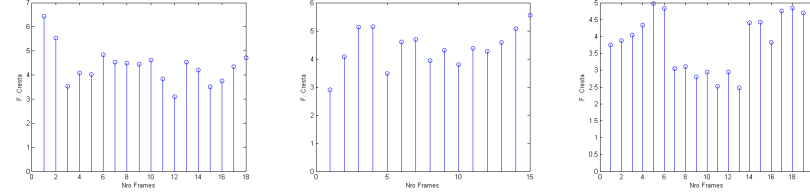


Figura 11: Factor cresta de los ronquidos de la extracción 2

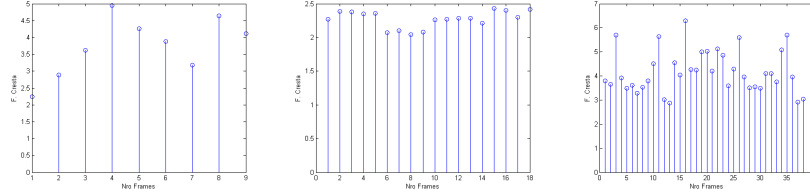


Figura 12: Factor cresta de los ruidos: tos, telefono, habla

3.5. Energía/ZCR

Esta medida es el resultado de calcular el cociente entre la energía de un frame y su ZCR. El uso de esta medida es muy comun en el análisis de voz para diferenciar sonidos sordos de sonoros, ya que los sonidos sordos se caracterizan por tener baja energía y un alto ZCR y lo inverso se da en los sonidos sonoros, por lo que este cociente refuerza este efecto.

$$EnergiaZCR_n = e_n / ZCR_n \quad (5)$$

Donde e_n y ZCR_n es la energía y el ZCR del frame n. En las figuras 13 y 14 se pueden observar los resultados del calculo de Energia/ZCR para los ronquidos de las extracciones 1 y 2, y en la figura 15 podemos observar lo mismo para los tres ruidos.

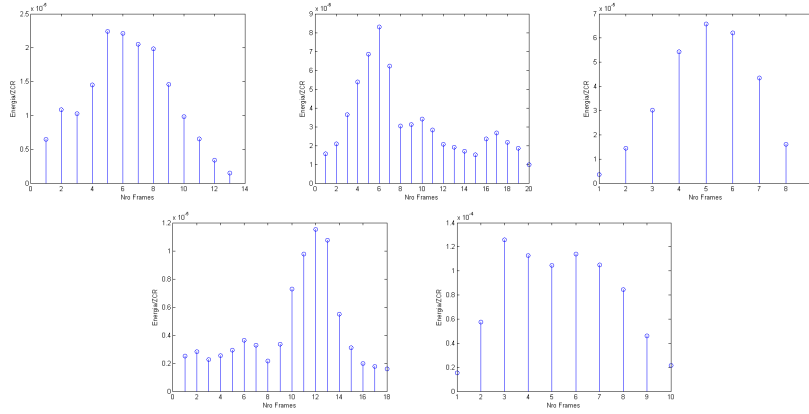


Figura 13: Energía/ZCR de los ronquidos de la extracción 1

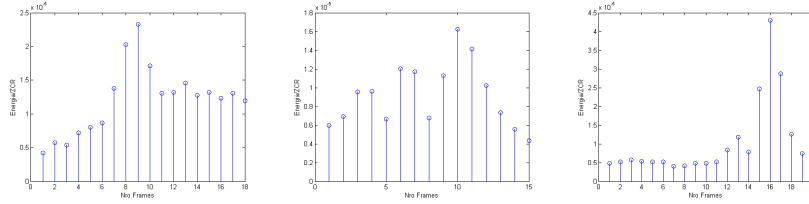


Figura 14: Energía/ZCR de los ronquidos de la extracción 2

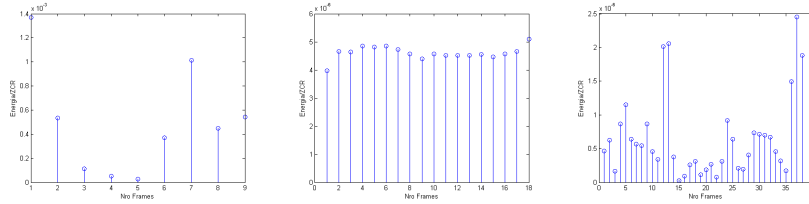


Figura 15: Energía/ZCR de los ruidos: tos, telefono, habla

3.6. Tabla de comparación de características entre ronquidos y diferentes ruidos

Habiendo enumerado y detallado el calculo de cada una de las características temporales usadas en este trabajo, se realizó el calculo de la media y la varianza para cada característica. El resultado de estos calculos podemos verlo en la siguiente tabla.

Tipo señal	Energia		ZCR		RMS	
	Media	Varianza	Media	Varianza	Media	Varianza
Ronquido 1	1.4824e-006	7.3062e-013	0.1187	5.0813e-005	0.0464	2.3648e-004
Ronquido 2	3.5131e-007	5.0034e-014	0.1095	4.2967e-005	0.0227	4.7432e-005
Ronquido 3	2.9647e-006	3.7739e-012	0.0983	1.8845e-004	0.0646	6.4752e-004
Ronquido 4	4.5723e-007	1.1735e-013	0.1032	8.9711e-005	0.0256	8.3111e-005
Ronquido 5	5.6499e-006	8.0302e-012	0.0728	2.0309e-005	0.0914	7.6039e-004
Ronquido 6	9.7302e-007	1.4449e-013	0.0819	8.8927e-005	0.0387	6.2551e-005
Ronquido 7	5.8264e-007	2.8707e-014	0.0668	2.8193e-004	0.0302	1.9703e-005
Ronquido 8	6.5582e-007	4.6976e-013	0.0631	1.7632e-004	0.0297	1.7543e-004
Ruido telefono	1.2261e-006	6.0263e-015	0.2663	1.8814e-004	0.0443	2.0889e-006
Ruido habla	8.5400e-008	4.2314e-015	0.1510	0.0019	0.0109	1.7872e-005
Ruido tos	4.4394e-005	1.3986e-009	0.1209	0.0030	0.2450	0.0124

Cuadro 1: Comparación entre características temporales.

Tipo señal	F. Cresta		Energía/ZCR	
	Media	Varianza	Media	Varianza
Ronquido 1	3.7394	0.2565	1.2521e-005	5.0502e-011
Ronquido 2	3.6741	0.2079	3.1830e-006	3.9167e-012
Ronquido 3	4.1288	0.1606	3.2874e-005	5.9082e-010
Ronquido 4	4.4144	1.1021	4.3764e-006	1.0342e-011
Ronquido 5	4.2944	0.6827	7.8678e-005	1.6449e-009
Ronquido 6	4.3660	0.5877	1.2099e-005	2.5444e-011
Ronquido 7	4.3985	0.4867	9.2240e-006	1.1552e-011
Ronquido 8	3.8218	0.7375	1.0471e-005	1.0884e-010
Ruido telefono	2.2721	0.0155	4.6062e-006	5.3290e-014
Ruido habla	4.1869	0.7475	6.5134e-007	3.5590e-013
Ruido tos	3.7534	0.7477	4.9615e-004	2.0231e-007

Cuadro 2: Comparación entre características temporales.

En la figura 16 se encuentran representados distintos gráficos de barras donde cada gráfico representa una característica y las barras del 1 al 8 corresponden a los ocho ronquidos utilizados en el trabajo y las barras 9,10 y 11 corresponden a los ruidos telefono, habla y tos.

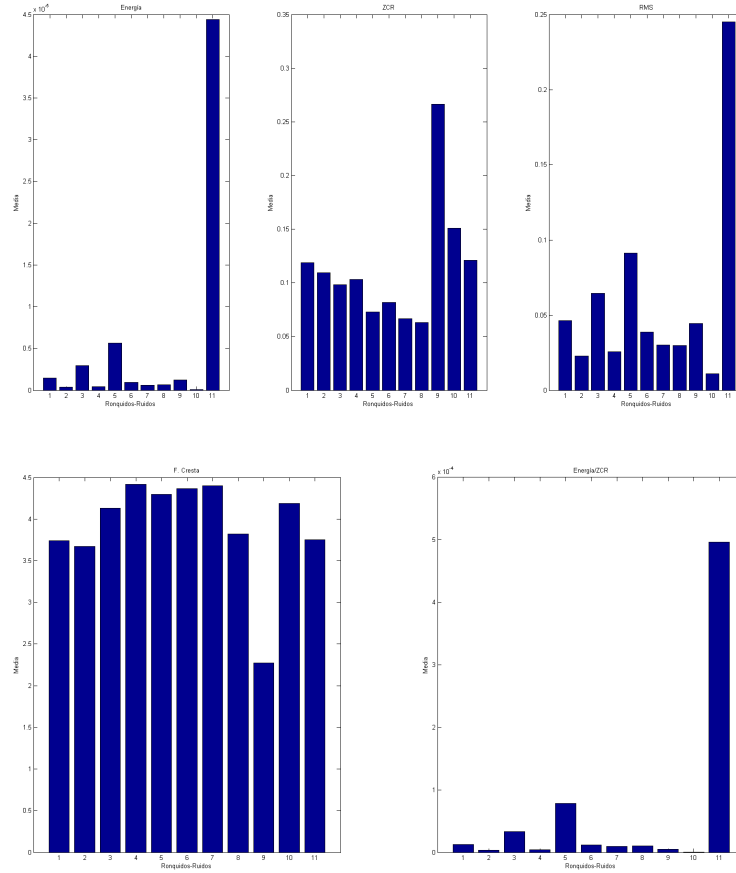


Figura 16: Comparación características temporales

Analizando estos resultados por característica podemos ver que en el caso de la energía y el RMS arrojan resultados similares entre sí, ya que la diferencia entre estas características está en que el RMS ignora los efectos de los valores negativos de la señal. En ambos casos podemos observar una menor variación de la media entre los ruidos, en cambio si observamos las barras 10 y 11 (habla, tos) vemos una diferencia abrupta entre ellas y con los ruidos. Esta diferencia visible entre las medias correspondientes a los ruidos y a las de los ruidos permiten concluir que la energía es una característica representativa y podremos tenerla en cuenta a la hora de entrenar un clasificador. En el caso del ZCR los resultados son mas visibles pudiéndose ver como las barras 9,10,11 correspondientes a los tres ruidos presentan valores mas altos que en las de los ruidos, con lo cual esta característica también es representativa y podrá ser utilizada en la diferenciación de ruido con otras señales no deseadas. Respecto del factor cresta tal como podemos apreciar en el gráfico de barras esta

característica presenta valores casi similares tanto entre los ronquidos como en los ruidos, el único valor visiblemente diferente se da con la barra 9 correspondiente al sonido de un telefono, dados esta poca variación esta característica no será tomada en cuenta ya que no aporta diferencias entre los ronquidos de los ruidos. Con el cociente energía/ZCR podemos observar como para el caso de los ruidos (barras 9,10 y 11) tenemos valores diferentes entre si y a su vez distintos respecto de los ronquidos, por ejemplo se puede ver que tanto el 9 y el 10 son mas pequeños que el resto y que en el caso del 11 es mucho mayor. Esto se debe a que el cociente permite potenciar haciendo que valores como el 9 cuya energía es similar al ronquido 1 pero que sin embargo el tener un zcr mayor hace que el cociente para el 9 sea menor al cociente del ronquido 1. Por lo tanto como a nivel de la energía no podíamos distinguir un ronquido como el 1 de un ruido como el 9 este cociente si permite distinguir entre ellos, es por esto que esta característica sera utilizada para este trabajo.