

# Tercera Entrega - Proyecto 2016

PROCESAMIENTO DIGITAL DE SEÑALES DE AUDIO

MAESTRÍA EN INGENIERÍA ELÉCTRICA del *Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República.*

*Juan Braga*

24 de octubre de 2016

## Extracción de embocadura en Aliento/Arrugas

En Aliento/Arrugas de Marcelo Toledo se utiliza como recurso compositivo tres tipos de embocadura para ejecución de la flauta. Se diferencian por el ángulo que forma la embocadura frente al flujo de aire en la ejecución del instrumento. Desde la audición se puede percibir las diferencias en el material sonoro generado. Se enlista a continuación los nombres para cada caso, manteniendo su denominación en Inglés (idioma utilizado en la partitura de la obra).

- *Blow Hole Covered*: El flujo de aire ingresa directo al tubo de la flauta, sin generar turbulencia contra el filo de la embocadura.
- *Breathy Embrochure*: Caso intermedio entre las otras dos embocaduras.
- *Normal Embrochure*: Embocadura clásica de la flauta, donde el el flujo de aire frente al filo de la embocadura genera la excitación tonal.

## Definición del problema y estrategia de resolución

Se propone la extracción automática de la embocadura a través del análisis computacional de grabaciones de la obra. En este caso se utilizará una estrategia de resolución del tipo de reconocimiento de patrones, en particular de clasificación supervisada.

## Datos

Se cuenta con 3 grabaciones de diferentes intérpretes de la obra Aliento/Arrugas. Los intérpretes son: Pablo Somma, Emma Resmini y Ulla Suokko. Los archivos de audio se etiquetaron utilizando el software *Sonic Visualizer* dividiendo los fragmentos de audio en 5 clases:

- Silencio.
- Silencio con respiración del intérprete.
- Sonido generado con *Blow Hole Covered*.
- Sonido generado con *Breathy Embrochure*.
- Sonido generado con *Normal Embrochure*.

## Extracción de características y clasificación

Se propone entonces las utilización de dos características de naturaleza distinta para la extracción automática de embocadura:

- *Mel-frequency Cepstral Coefficients (MFCC's)*.
- *Linear Predictive Coding (LPC's)*.

En cuanto a los algoritmos de clasificación se utiliza:

- *Linear SVM*.
- *Random Forest*.

A continuación se presentan los resultados para la estrategia planteada en la extracción automática de embocadura a partir del audio.

## Resultados

Para el cálculo de características se utilizan ventanas de análisis de  $23ms$  y saltos del 50 % del largo de la ventana. Se dejaron afuera del análisis las clases: *Silencio* y *Silencio con respiración del intérprete*.

Para el cómputo de *LPC's* se utiliza un orden de  $p = 26$  generando vectores de características de dimensión 26. En este caso se utiliza una implementación vista en el curso realizada en *Matlab*.

Por otro lado para los *MFCC's* se utilizan 40 bandas-mel, determinando un vector de dimensión 40 para estas segundas características. En este caso se utiliza la implementación en *Python* de *Librosa* (McFee et al., 2015).

En cuanto a los algoritmos de clasificación se utiliza el software *Weka* (Hall et al., 2009), en particular la implementación *SMO* con *PolyKernel* para *Linear SVM* y *Random Forest* ambos con valores por defecto.

A continuación en la Tabla 1 se observan los resultados. Además se agregó para comparación cualitativa la clasificación para la extracción de características vista en la entrega pasada. Los valores fueron determinados con *10-fold cross validation*.

	LinearSVM	RandomForest
MFCC's	83.6 %	91.9 %
LPC's	69.8 %	76.1 %
Voicing	76.7 %	74.6 %

Cuadro 1: Instancias correctamente clasificadas según características y algoritmo de clasificación.

## Conclusiones

Las características que mejor se comportan en la resolución del problema son las *MFCC's*. Por otro lado *LPC's* de orden  $p = 26$  y las características de *Voicing* planteadas en la entrega anterior tienen un desempeño similar.

## Referencias

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th Python in Science Conference*.