

Informe Final - 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

6 de diciembre de 2016

Extracción de embocadura en Aliento/Arrugas: Introducción

En Aliento/Arrugas de Marcelo Toledo se utiliza como recurso compositivo tres tipos de embocadura para ejecución de la flauta. En la ejecución del instrumento se diferencian por el ángulo que forma la embocadura frente al flujo de aire. 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). Además en la Figura 1 se observa su notación en la partitura de Aliento/Arrugas.

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

Figure 1 displays three musical systems (a, b, c) illustrating different flute embouchure techniques and their corresponding vocalizations. Each system includes a Flute staff and a Voice staff.

- (a) Blow Hole Covered:** The system is marked "INTENSO E CON FORZA!" and "Lento". The Flute staff shows a "bhc" (Blow Hole Covered) embouchure. The Voice staff shows "Exhale" and "Inhale" markings, with dynamic markings like *sfz/p*, *f*, *p*, *ff*, and *fz*. The Voice part includes the syllables "Jh hak - sh".
- (b) Breathy Embouchure:** The system is marked "C breathy". The Flute staff shows a "bhc" embouchure. The Voice staff shows "Exhale" and "Inhale" markings, with dynamic markings like *mf*, *ff*, *sfz/p*, and *f*. The Voice part includes the syllables "Ksh - K", "Ksh - K", and "Ksh - Ks ts Ks ts Ks".
- (c) Normal Embouchure:** The system is marked "LENTO, DELICATO E LONTANO" and "AGITATO CON FORZA". The Flute staff shows a "normal embouchure". The Voice staff shows "Exhale" and "Inhale" markings, with dynamic markings like *ppp*, *p*, *mf/p*, *f*, and *ff*. The Voice part includes the syllables "a - a - a - a" and "mf/p - f/p - f - f - ff".

Figura 1: Notación de las embocaduras se observa en la parte superior de los sistemas. (a) *Blow Hole Covered*. (b) *Breathy Embouchure*. (c) *Normal Embouchure*. Fragmentos extraídos de la partitura de Aliento/Arrugas.

Definición del Problema

Se propone la extracción automática del tipo de embocadura a través del análisis computacional de grabaciones de la obra.

Datos

Se cuenta con 5 grabaciones de diferentes intérpretes de la obra *Aliento/Arrugas*. Los intérpretes son: Pablo Somma, Emma Resmini, Claire Chase, Juan Pablo Quinteros 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 Embouchure*.
- Sonido generado con *Normal Embouchure*.

Para el presente informe se trabaja con las tres últimas clases de la lista anterior.

1. Estrategia de resolución

Se trata de resolver el problema con un enfoque de reconocimiento de patrones. Se procesa el audio como un *Bag of Frames* a partir del computo de características. Se utilizan tres algoritmos clásicos de clasificación para evaluación del desempeño de las características en la extracción automática de embocadura a partir de las muestras de audio.

1.1. Análisis de la naturaleza sonora del problema

Color Material sonoro Embocadura

1.2. Extracción de características

MFCC LPC SPECTRAL CONTRAST CLASSICS SPECTRALS

Se evaluó el desempeño de características de diversa naturaleza en la extracción del tipo de embocadura.

Todos los experimentos se realizan con *5-fold cross validation* donde los folds son las diferentes interpretaciones de la pieza musical. De esta forma se asegura que frames provenientes de la misma grabación no sean usados para train y test en un mismo experimento.

Se utilizan tres clasificadores distintos para minimizar el bias que pueda existir entre los datos y un algoritmo en particular. Se trabaja con los algoritmos: *Random Forest (trees=10)*, *Support Vector Machine (kernel lineal)* y *K-Nearest Neighbors (k=10)*. La implementación utiliza el módulo de *Python* llamado *Scikit Learn* (Pedregosa et al., 2011)

Se identifica nuevamente el problema en la separación

Comparar confusión entre 2x2 y 3x3 en la separación de BHC y BREATHY

Discusión de agregar voicing, ver como mejora la matriz de confusión

1.2.1.

2. Trabajo a futuro

- Segmentación de audio, umbralizando los silencios y las respiraciones
- Salir del bag of frames, para utilizar la redundancia temporal

Referencias

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.