

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

8 de diciembre de 2016

Extracción de embocadura en Aliento/Arrugas: Introducción

En Aliento/Arrugas de Marcelo Toledo se utiliza como recurso expresivo tres tipos de embocadura para ejecución de la flauta. Se diferencian por el ángulo que forma el flujo de aire frente al filo de la embocadura. Se enlista a continuación los nombres de cada una, 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 "Toneless sound" with "Exhale" and "Inhale" markings. The Voice staff shows "sffz/p" and "f" dynamics. The system is labeled "bhc" (Blow Hole Covered) and "Blow Hole covert".
- (b) Breathy Embouchure:** The system is marked "C breathy". The Flute staff shows "mf" and "ff" dynamics. The Voice staff shows "Ksh" and "K" vocalizations. The system is labeled "bhc" (Blow Hole Covered) and "breathy".
- (c) Normal Embouchure:** The system is marked "LENTO, DELICATO E LONTANO" and "AGITATO CON FORZA". The Flute staff shows "ppp", "p", "mf/p", "f", and "ff" dynamics. The Voice staff shows "a", "mp", "mf/p", "f", and "ff" dynamics. The system is labeled "normal embouchure" and "bhc" (Blow Hole Covered).

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. Análisis de la naturaleza sonora del problema

Se cree que se puede separar con características que describan el nivel de periodicidad de la señal. Primer experimento: con Voicing y ZCR (? , Chapter 4) El compositor Color Material sonoro Embocadura

2. 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.

2.1. Experimentos

Se evalúa 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)*. En todos los casos se utilizan los parámetros por defecto ya que no es objetivo de este trabajo encontrar los valores óptimos de clasificación. La implementación se realiza mediante el módulo de *Python* llamado *Scikit Learn* (Pedregosa et al., 2011). En todos los casos los datos son preprocesados de manera de centrar en cero y escalar la varianza a uno, previo al clasificador.

2.2. Extracción de características

Se enlistan a continuación las características que se evalúan la extracción automática de embocadura. Además se describe brevemente sus principales características.

- Características Clásicas
- MFCC
- LPC
- Spectral Contrast (Jiang et al., 2002)

2.3. Resultados

Accuracy

Confusión de MFCC:

2.4. Experimentos 2

MFCC es un fierro pero apriori no tiene porque utilizar la información de periodicidad propiamente dicha. Parámetro relevante en la separación de estos tres tipos de embocadura. Se agrega voicing tipo yin De Cheveigné and Kawahara (2002)

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

2.4.1.

3. Trabajo a futuro

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

Referencias

- De Cheveigné, A. and Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930.
- Jiang, D.-N., Lu, L., Zhang, H.-J., Tao, J.-H., and Cai, L.-H. (2002). Music type classification by spectral contrast feature. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, volume 1, pages 113–116. IEEE.
- Klapuri, A. and Davy, M. (2007). *Signal processing methods for music transcription*. Springer Science & Business Media.
- 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.