

# Propuesta de Proyecto Final: Estimación de nota musical

Saul Ivan Rivas Vega

Aprendizaje Automatizado

3 de marzo de 2020

# 1. Descripción y delimitación del problema

Basado en atributos en el dominio de la frecuencia como son los coeficientes por ventana de muestreo de la transformada **constante-Q** estimar la nota musical de guitarra presente en un archivo de audio de una pieza musical monofónica con notas simples.

## 2. Objetivos

- Extraer los atributos frecuenciales por ventana del conjunto de datos.
- Entrenar un modelo predictivo para estimar la nota musical presente en una ventana dados sus atributos frecuenciales.
- Realizar un transcript de la melodía.

## 3. Justificación

Se han realizado múltiples estudios en la estimación de frecuencia fundamental o de tono [1–4]. En aplicaciones el ser en tiempo real como en [1] sería de gran utilidad, sin embargo los métodos con mayor precisión son los que analizan archivos pre-grabados como es el caso de [2,3], ambos superando al método estándar YIN [4] que se encuentra en múltiples bibliotecas de extracción de información musical como ESSENTIA [5].

El presente trabajo tiene como justificación el poder utilizar las propiedades fundamentales para la estimación del tono como en [4,6,7] y a su vez beneficiarse de los métodos de aprendizaje automatizado para ofrecer una opción para la estimación de notas musicales balanceando la eficiencia en la cantidad de atributos requeridos y la precisión de la estimación.

## 4. Base de datos a utilizar o estrategia para recopilarla

La base de datos es NSynth del grupo de investigación musical **Magenta** en Google [8], el dataset contiene 305,979 notas musicales, cada una con un distinto tono, timbre, y envoltura, obtenidos de 1,006 instrumentos grabando clips de monofónicos con una tasa de muestreo de 16kHz de 4 segundos con anotaciones de nota musical en el rango del formato MIDI (21-108) con 5 velocidades (25, 50, 75, 100, 127). La nota se mantuvo por 3 segundos dejando el último segundo para el decaimiento.

## 5. Análisis exploratorio de los datos

De los clips de audio del instrumento seleccionado se obtendrán los valores de la transformada **Constante-Q** utilizando la biblioteca Librosa para python [9] y exportando en formato csv los atributos frecuenciales y la clase para utilizar en el entrenamiento.

## Referencias

- [1] O. Das, J. O. S. Iii, and C. Chafe, “Real-time Pitch Tracking in Audio Signals with the Extended Complex Kalman Filter,” p. 7, 2017.
- [2] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, “CREPE: A Convolutional Representation for Pitch Estimation,” *arXiv:1802.06182 [cs, eess, stat]*, Feb. 2018, arXiv: 1802.06182. [Online]. Available: <http://arxiv.org/abs/1802.06182>
- [3] M. Mauch and S. Dixon, “PYIN: A fundamental frequency estimator using probabilistic threshold distributions,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 659–663, iSSN: 2379-190X.
- [4] A. de Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, Apr. 2002, publisher: Acoustical Society of America. [Online]. Available: <https://asa.scitation.org/doi/10.1121/1.1458024>
- [5] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, “ESSENTIA: an Audio Analysis Library for Music Information Retrieval,” Nov. 2013.
- [6] J. C. Brown, “Calculation of a constant Q spectral transform,” *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, Jan. 1991. [Online]. Available: <http://asa.scitation.org/doi/10.1121/1.400476>
- [7] J. Brown and M. Puckette, “An efficient algorithm for the calculation of a constant Q transform,” *Journal of the Acoustical Society of America*, vol. 92, p. 2698, Nov. 1992.
- [8] J. Engel, C. Resnick, A. Roberts, S. Dieleman, D. Eck, K. Simonyan, and M. Norouzi, “Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders,” *arXiv:1704.01279 [cs]*, Apr. 2017, arXiv: 1704.01279. [Online]. Available: <http://arxiv.org/abs/1704.01279>
- [9] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “librosa: Audio and Music Signal Analysis in Python,” Jan. 2015. [Online]. Available: <https://scinapse.io/papers/2191779130>