# **Revisión de literatura**

Como punto de partida a la teoría que soporta este trabajo de investigación, se hace una definición de lo que es una señal de audio en su forma más general y como la percibe un ser humano, esta señal se puede definir como una función del nivel de presión del sonido dependiente del tiempo. Para capturar una señal de audio es posible utilizar un dispositivo micrófono, dispositivo que convierte los niveles de presión del sonido en voltajes eléctricos, esta señal de voltajes eléctricos está definida para todos los instantes de tiempo, y por lo tanto es considerada una señal de tiempo continuo y se representa como Y(t).

Para gestionar una señal de voltaje Y(t) en una computadora, es necesario capturar una versión de la misma pero representable en una dimensión digital, lo que quiere decir Y(t) se debe convertir en y[n], donde y[n] es una versión de tiempo discreto y de valores cuantizados de Y(t), la siguiente grafica representa una señal de audio analógica y su representación en digital:

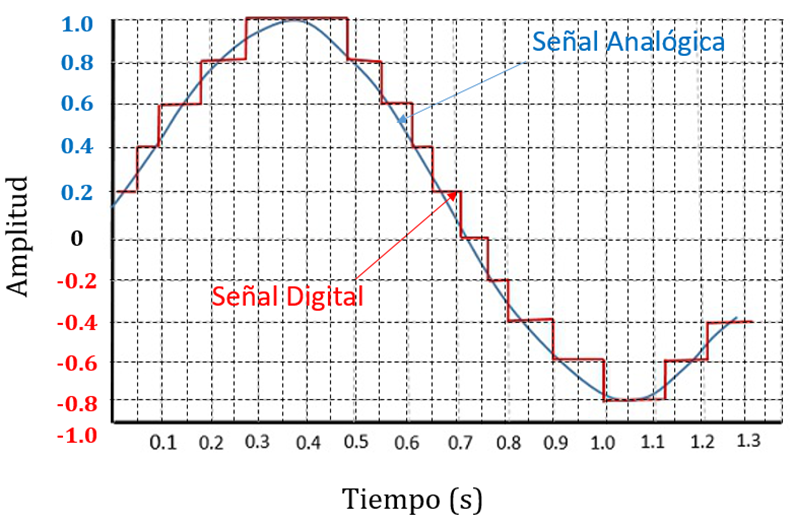


Figura xx. Señal Analógica, Y(t)*,* vs Señal Digital, y[n]. Fuente de elaboración propia.

Una señal discreta en el tiempo se puede entender como una señal que solo está definida para ciertos instantes de tiempo. Esto es, si la señal fue muestreada (Frecuencia de Muestreo, Fs) a 20 Hz, es decir veinte muestras en un segundo, cada una de las muestras está separada de la otra en 0.05 s. En la Figura 2 se puede observar el ejemplo de una señal analógica muestreada a Fs = 20 Hz.

Este tiempo se conoce como periodo de muestreo, Ts. La relación entre la frecuencia de muestreo y el periodo de muestreo viene dada por la siguiente ecuación ():

Fs=1/Ts (1)El ser humano solo puede percibir componentes frecuenciales en la señal desde los 20 a los 20.000 Hz. Luego entonces es necesario muestrear este tipo de señales por lo menos al doble de la máxima frecuencia (Teorema de Nyquist), esto es, a 40.000 Hz. En la práctica, es muy común encontrar que las señales de audio musical se muestrean a 44.100 Hz. Lo anterior significa que la separación temporal entre muestras es de 2.2676x10-5 s (0.02267 ms ≈ 22.7 µs).

Por otro lado, la cuantización hace referencia a que los valores de amplitud de la señal de audio también se han discretizados al igual que el tiempo. Usualmente, la amplitud se fija dentro de un rango, entre un valor máximo, Amax, y uno mínimo, Amin, simétrico respecto al valor cero. Este rango se divide en M partes. Por ejemplo, en la Figura XX se ha dividido el rango de -1.0 a 1.0 en 11 partes simétricas respecto al 0. Si M es una potencia de 2, la amplitud de cada muestra cuantizada puede ser fácilmente representada mediante un código binario de longitud, NB, igual a log2(M) bits. Dado que en este escenario M sería siempre un número impar, es imposible conservar al cero como un valor de cuantización y a la vez tener un número simétrico de pasos de cuantización. En la práctica se fija un paso menos en los valores positivos. Longitudes típicas de códigos binarios en señales de audio son 16, 24 y 32 bits. Estas longitudes permiten tener 216 = 65536, 224 = 16.777.216 y 232 = 4.294.967.297 divisiones del rango, respectivamente. A su vez se tienen, por ejemplo, separaciones entre valores de amplitud cuantizadas, ΔQ, para un rango entre -1.0 y 1.0, y NB = 32 de:

ΔQ=(A\_max-A\_min)/2^NB (2)

Ejemplo:

ΔQ=(A\_max-A\_min)/2^NB =(1.0-(-1.0))/2^32 ≅4.66x〖10〗^(-10)

Generalmente, las señales de audio pueden ser mono (un canal) o estéreo (dos canales). Para propósitos de desarrollo de algoritmos MIR se utilizarán, en este documento, señales de audio de un canal. En caso de disponer de señales audio estéreo se seleccionará uno de los canales de esta señal.

# **8. ANTECEDENTES Y ESTADO ACTUAL DEL TEMA**

Para la primera etapa de este trabajo se hace uso de tecnologías de procesamiento de la señal, experimentando concretamente con señales no estacionarias finitas. Estas señales son procesadas utilizando un conjunto de herramientas y algoritmos que están enmarcados dentro de un área dedicada a la recuperación de información musical, conocida como Music Information Retrieval (MIR), un campo de investigación multi-disciplinar donde se combinan diferentes saberes como la ingeniería y la música.

Existe un organismo dedicado a la evaluación formal de los algoritmos diseñados por las diferentes comunidades, como lo es el Music Information Retrieval Evaluation eXchange (MIREX), administrado por el laboratorio The International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) y la Universidad de Illinois Urbana de Chanpaign (UIUC) (Downie, 2008, p. 247)[[1]](#footnote-1).

En la siguiente tabla, recuperada del artículo de J. Stephen Downie, se describen las tareas evaluadas por el MIREX en los años 2005, 2006 y 2007.

1. Downie, JS (2008). The music information retrieval evaluation exchange (2005-2007): Una ventana a la investigación de información musical recuperación. Acústica Ciencia y Tecnología, 29 (4), 247-255 [↑](#footnote-ref-1)