

Procesamiento digital de señales de audio

Introducción al procesamiento de audio

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

① Introducción

Modalidad y contenido del curso
Temática y aplicaciones

② Muestreo y cuantización

Muestreo
Cuantización

Modalidad y contenido del curso

Modalidad

- Tipo: grado y posgrado, Créditos: 8
- Horario:
 - martes y jueves de 10:00 a 12:00
 - Laboratorio de Software del IIE
- Aprobación:
 - entrega de ejercicios obligatorios (individual)
 - proyecto final sobre un problema establecido (en parejas)
- Práctico:
 - 4 repartidos, una semana luego de cada bloque temático

Modalidad y contenido del curso

Contenido

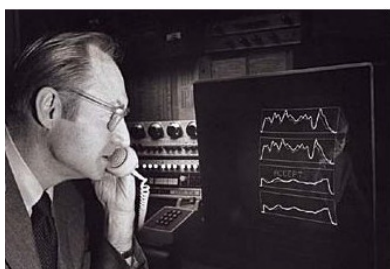
- *Introducción al procesamiento de audio*
- *Modelos para señales de voz y audio*
- *Percepción auditiva*
- *Filtros digitales en audio, síntesis de sonido, efectos*
- *Análisis de tiempo corto de señales de audio*
- *Procesamiento tiempo-frecuencia*
- *Análisis por modelado espectral*
- *Análisis Homomórfico y por Predicción Lineal*
- *Codificación de voz y audio*
- *Recuperación de información musical (MIR)*

Temática y aplicaciones

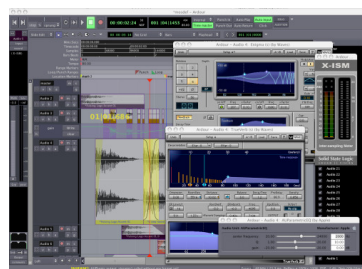
Temática

- voz hablada: gran desarrollo en telecomunicaciones (desde 1960s)
- audio: conjunto más amplio de tipos de señal de audio
 - música, sonidos ambientales, de origen animal, maquinaria, etc
- algoritmos específicos permiten mejores resultados (e.g. codificación)

objetivo: comprender los fundamentos de la producción y percepción de sonido y cómo las técnicas de procesamiento digital de señales pueden contribuir a resolver los diversos escenarios de aplicación

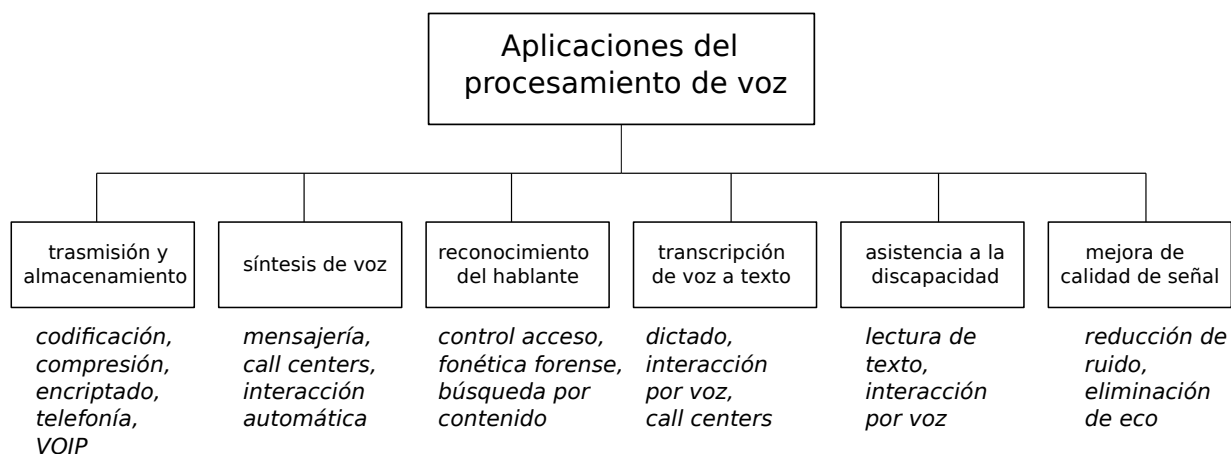


J. Flanagan, Bell Labs, 1972



Digital Audio Workstation, Ardour

Procesamiento de voz



[Rabiner and Schafer, 2011]*

*Rabiner, L. R. and Schafer, R. W. (2011). *Theory and Applications of Digital Speech Processing*

Procesamiento de voz



[Rabiner and Schafer, 2011]*

*Rabiner, L. R. and Schafer, R. W. (2011). *Theory and Applications of Digital Speech Processing*

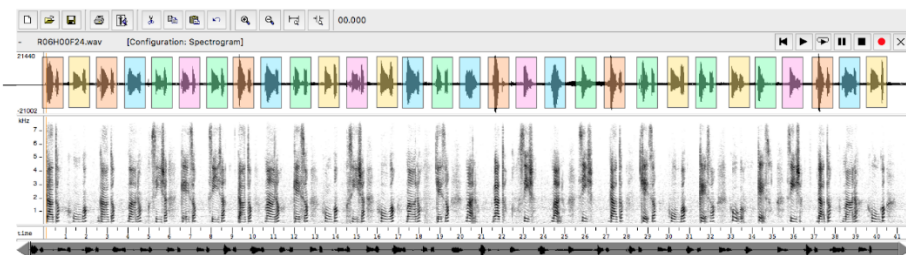
Procesamiento de voz

- detección temprana de dificultades en el aprendizaje de la lectura



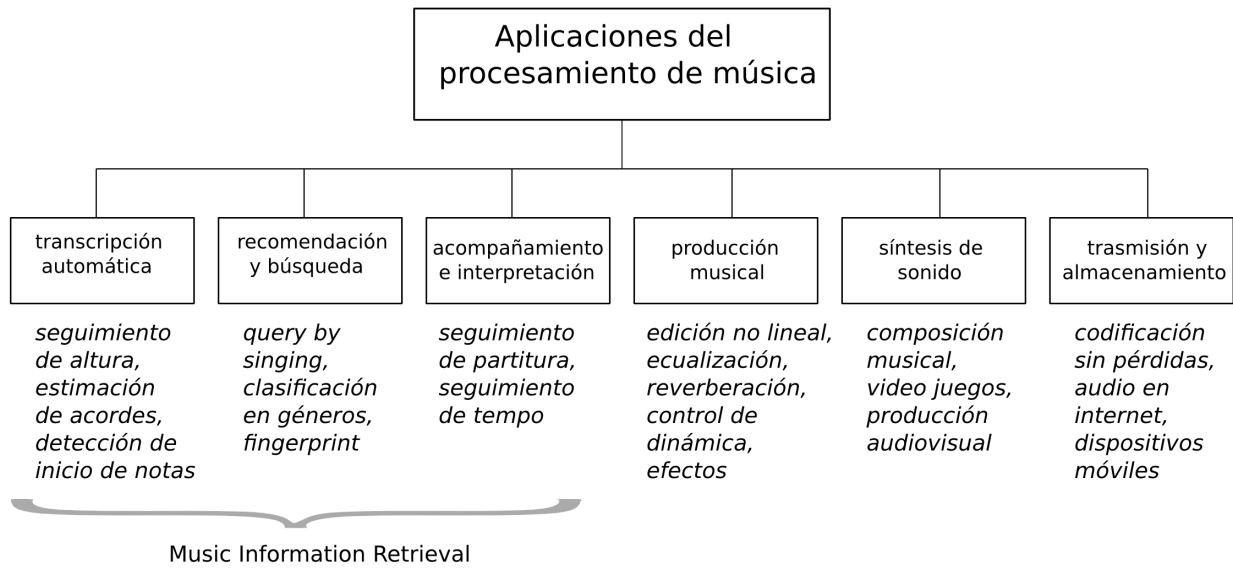
audio original

- reconocimiento automático de palabras, identificación de errores



grupos: 0 1 2 3 4

Procesamiento de música



Procesamiento de música



+



VOCALOID[™]
singing synthesizer - Yamaha

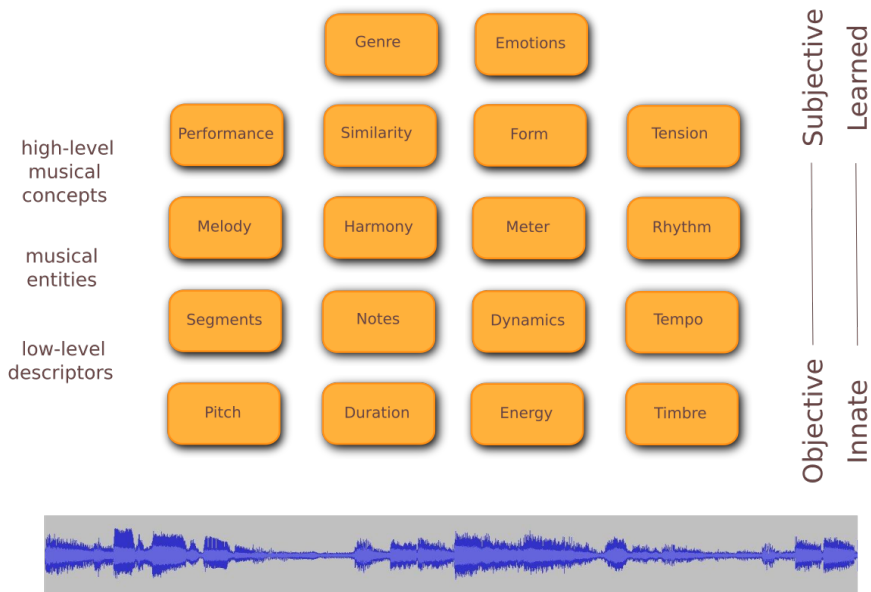
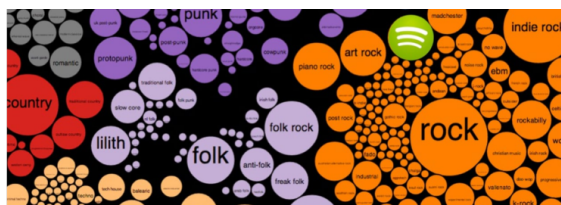




Paul Lamere, Director of Developer Platform at The Echo Nest.

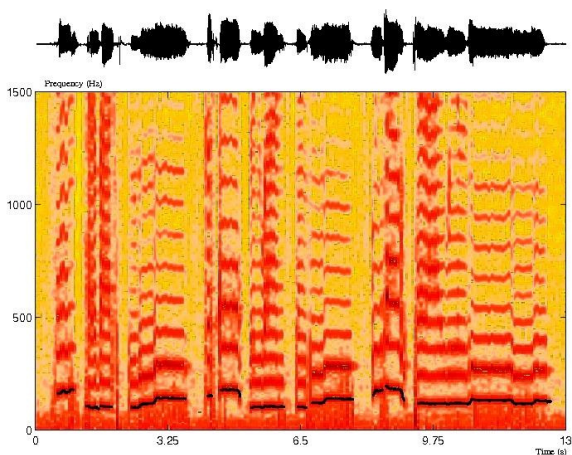



Xavier Serra, Director of Music Technology Group, UPF.

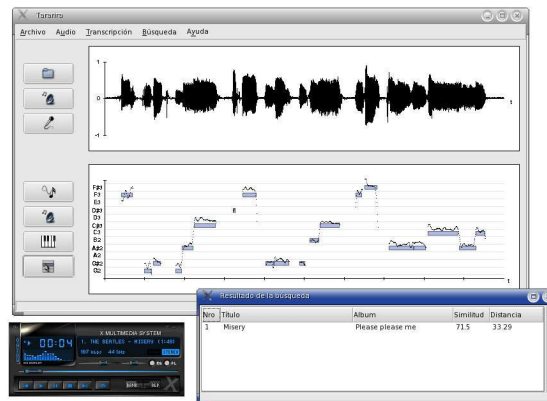
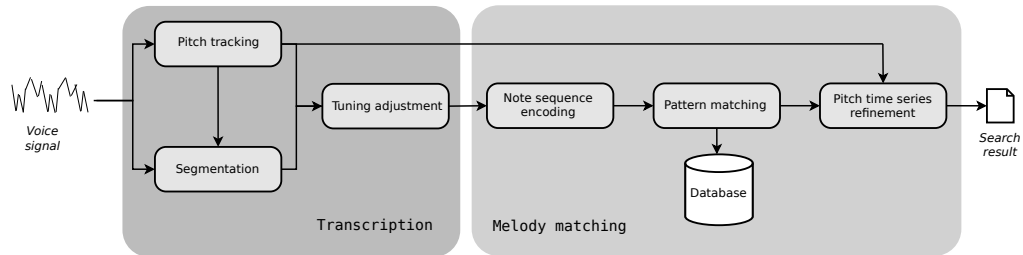


Procesamiento de música

- representación de la señal: temporal, espectral, tiempo-frecuencia
- descriptores de bajo nivel: transformación que captura aspecto particular
- entidades musicales: combinando y agregando descriptores de bajo nivel
- modelos semánticos: capturan similitudes y diferencias entre conceptos

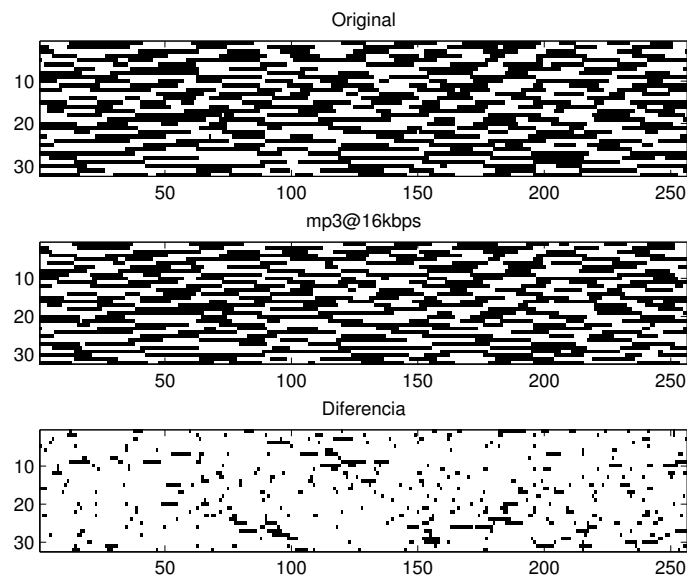


Aplicaciones: Query by Humming



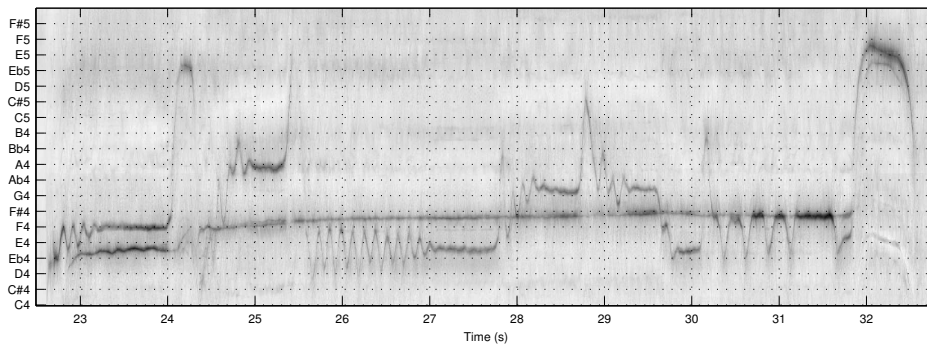
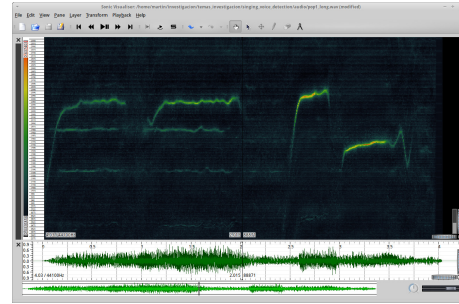
Aplicaciones: Audio Fingerprint

- identificación de audio (voz, música) a partir de un fragmento breve
- **huella digital** robusta a transformaciones como compresión y ruido
- funcionando en aplicaciones comerciales y dispositivos móviles



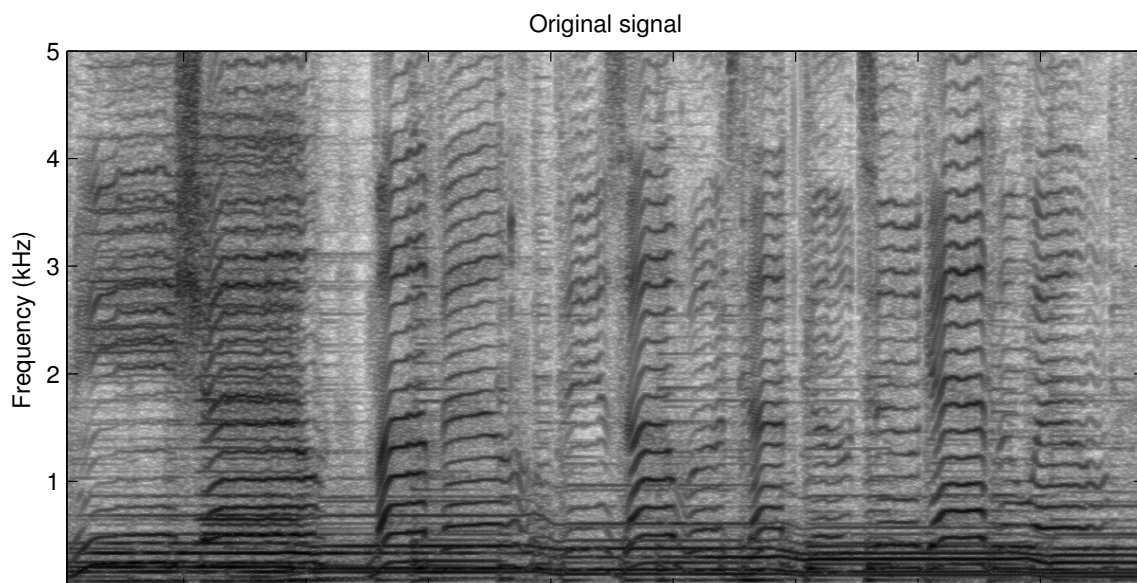
Aplicaciones: Análisis de interpretación

- software de representación de contenido melódico
- análisis musicológico de interpretación (afinación, rasgos expresivos, etc.)



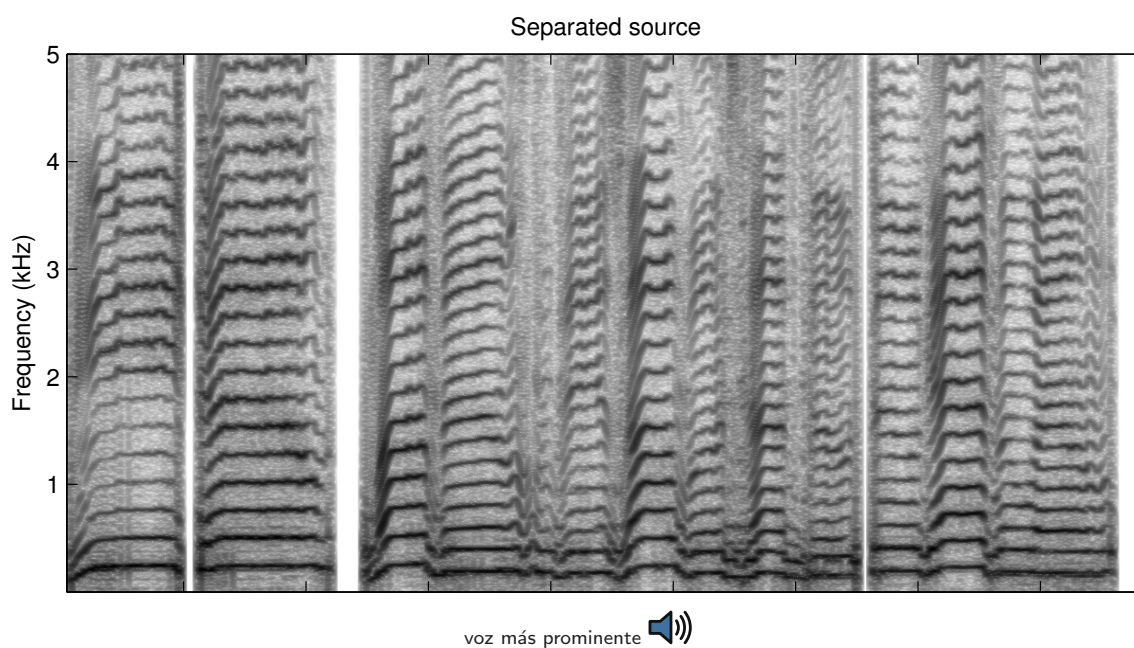
🔊 Bulgarian harvest song - Diaphonic chant of the Shope country

Aplicaciones: Separación de fuentes

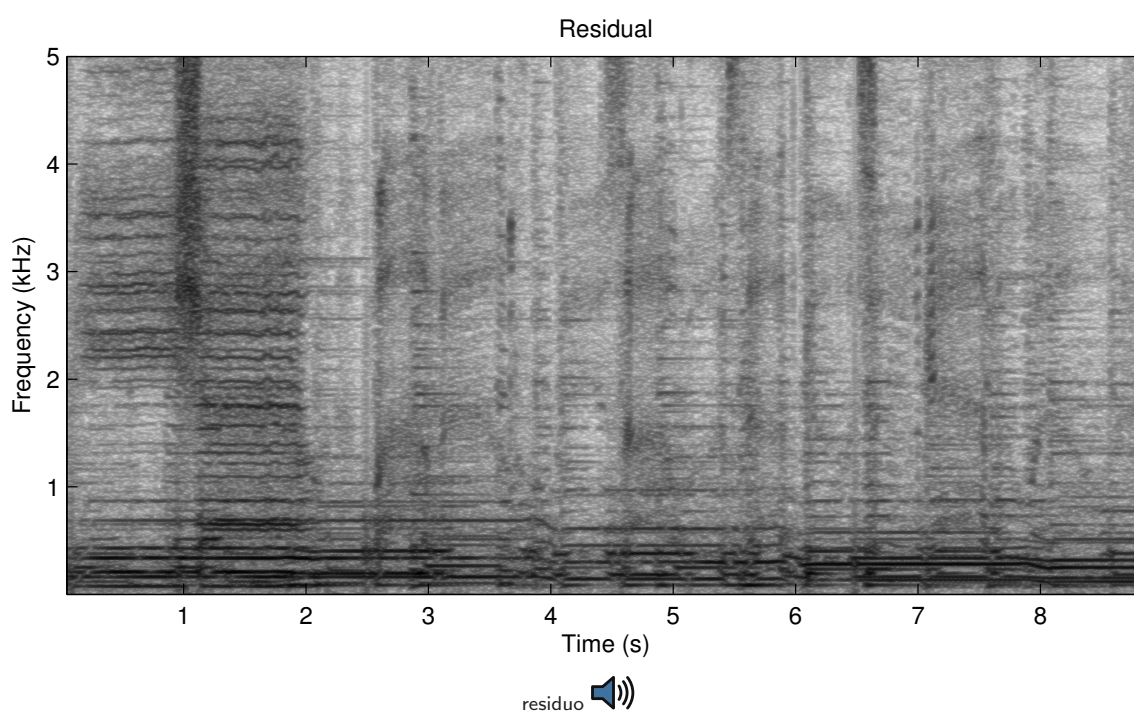


señal original 🔊

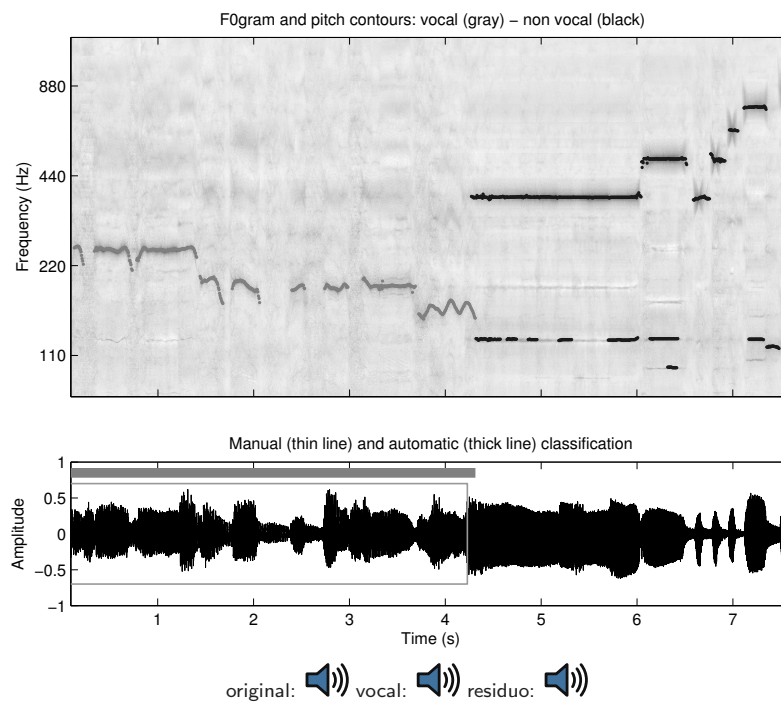
Aplicaciones: Separación de fuentes



Aplicaciones: Separación de fuentes



Aplicaciones: Separación de fuentes y detección de voz



Aplicaciones: Reconocimiento de cantantes

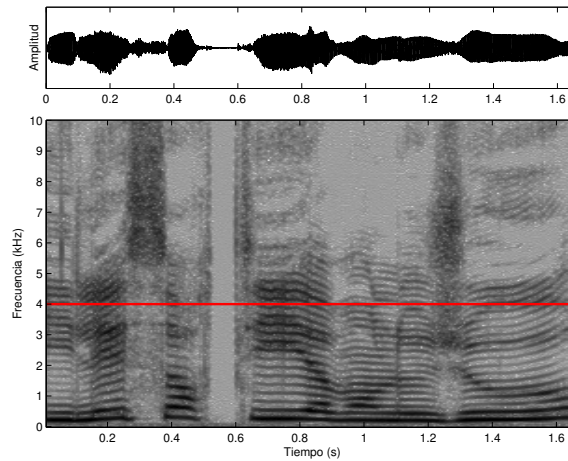


ejemplo

Muestreo

Frecuencia de muestreo

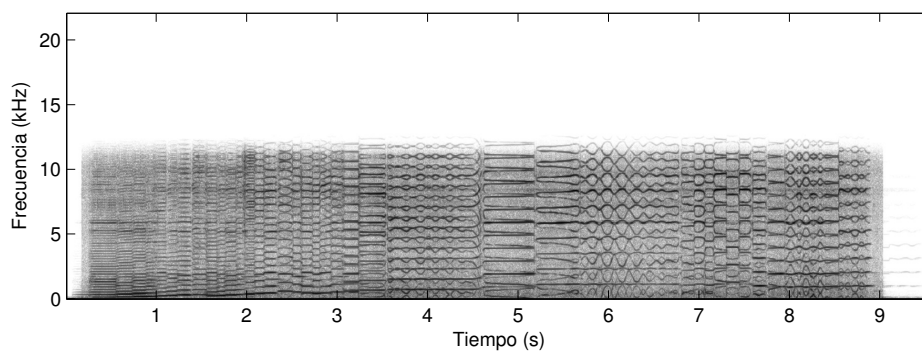
- teorema de muestreo: $f_s > 2f_{max}$
- en audio f_s de 8 a 192 kHz, dependiendo de la aplicación
- standards: CD 44.1 kHz, DVD-Audio y Blu-ray hasta 192 kHz



Muestreo

Aliasing

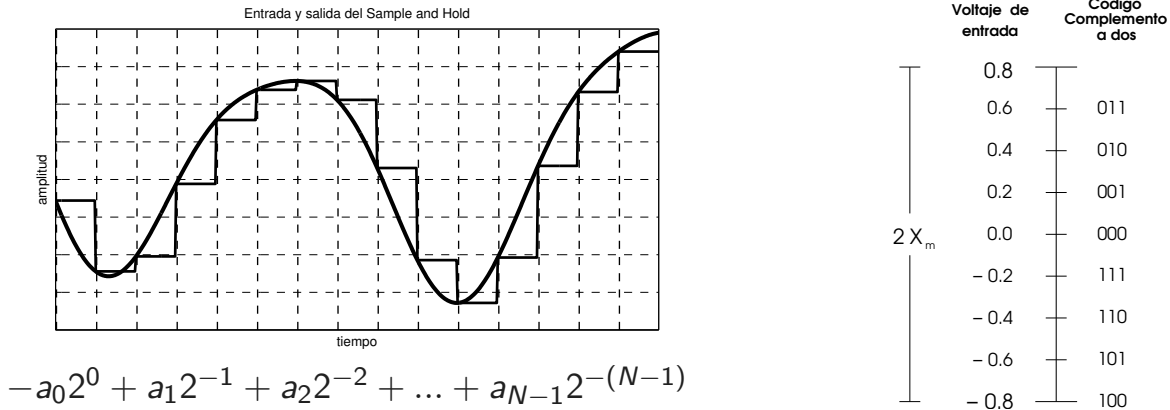
- señal a muestrear de banda limitada (filtro antialias)
- si esto no se cumple sobreviene *aliasing*
- no es un problema práctico, salvo en ocasiones (e.g. síntesis)



Cuantización

Cuantización uniforme

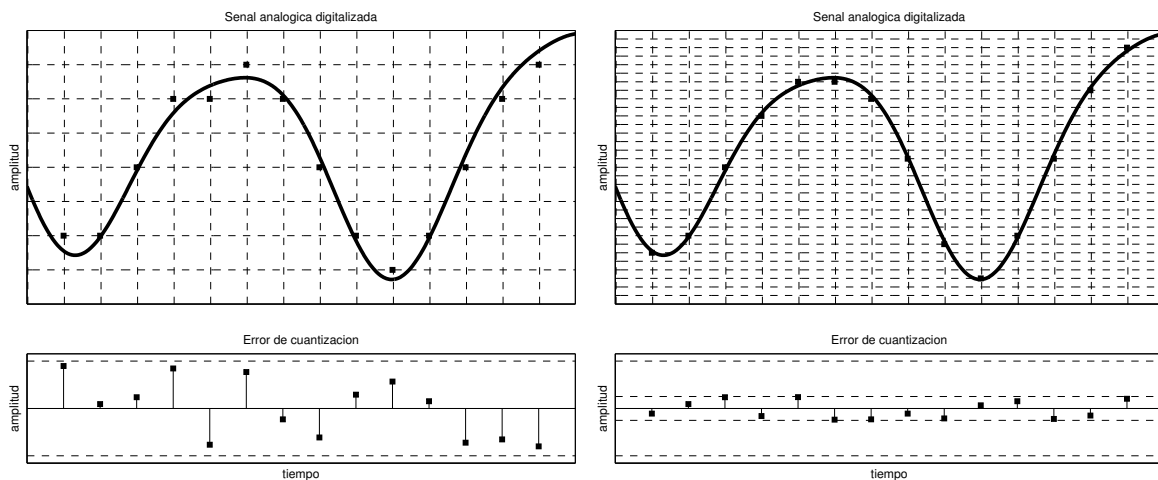
- cuantización: aproximar amplitud a precisión finita (redondeo)
- cantidad de niveles: 2^N , con N largo de palabra
- paso de cuantización: $Q = \frac{2X_m}{2^N}$, con X_m amplitud de pico máxima



Cuantización

Error de cuantización

- error: aproximación introduce pérdida de información
- amplitud del error: acotado a $Q/2$, varía entre $-Q/2 : Q/2$
- decrece al aumentar N largo de palabra (e.g. 8, 16, 24 bits)



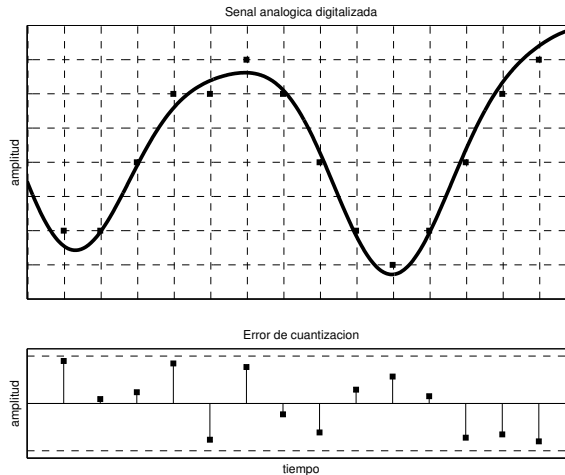
Cuantización

Modelo de error de cuantización

hipótesis: error no correlacionado con la señal ni consigo mismo

modelo: ruido de distribución uniforme, $U(-\frac{Q}{2}, \frac{Q}{2})$

señal compleja de gran amplitud \rightarrow poco audible (enmascarado)



Cuantización

Relación señal a ruido

hipótesis: variaciones de gran amplitud, espectro de banda ancha
sinusoide de amplitud máxima, ruido uniforme

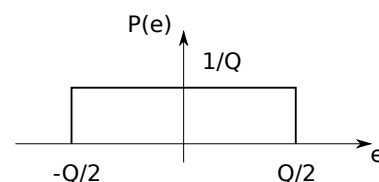
SNR crece 6 dB por bit, determina rango dinámico manejable

$$S_{rms} = \frac{S_{peak}}{\sqrt{2}} = \frac{Q2^{n-1}}{\sqrt{2}}$$

$$E_{rms} = \left[\frac{1}{Q} \int_{-\frac{Q}{2}}^{\frac{Q}{2}} e^2 de \right]^{\frac{1}{2}} = \left[\frac{Q^2}{12} \right]^{\frac{1}{2}}$$

$$\frac{S}{E} = \left[\frac{S_{rms}}{E_{rms}} \right]^2 = \frac{3}{2} 2^{2n}$$

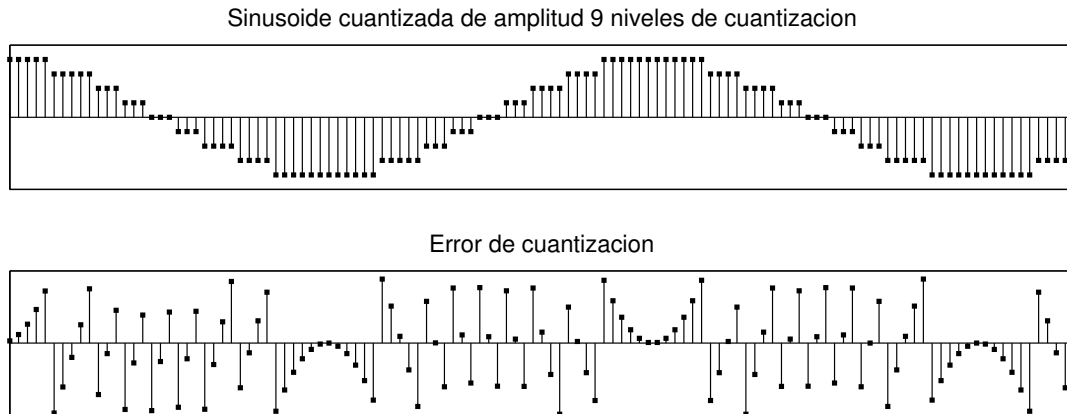
$$\frac{S}{E}(\text{dB}) = 20 \log \left[\sqrt{\frac{3}{2}} 2^n \right] = 6.0206n + 1.7609$$



Cuantización

Distorsión

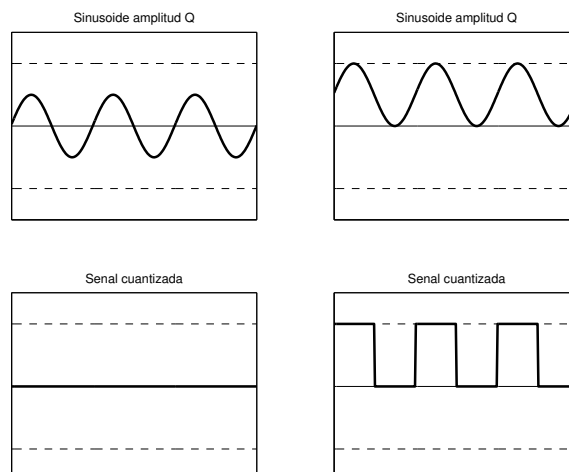
- señal compleja de gran amplitud: error asimilable a ruido blanco
- pocos pasos de cuantización: correlación entre señal y error
- deja de percibirse como ruido, efecto de distorsión más nocivo



Cuantización

Distorsión y dither

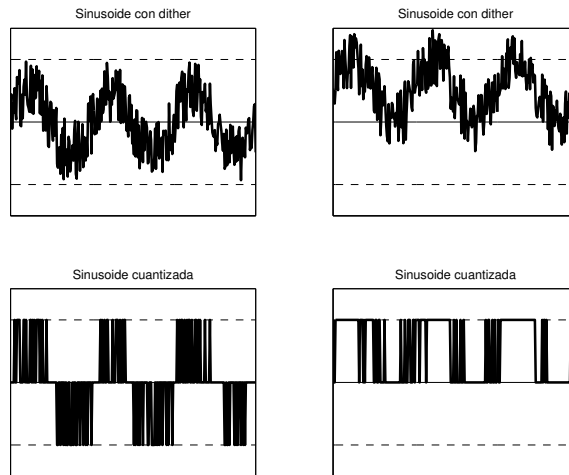
- sinusoides de amplitud Q centradas en nivel y paso de cuantización
- cuantización introduce severa distorsión (armónicos y aliasing)



Cuantización

Distorsión y dither

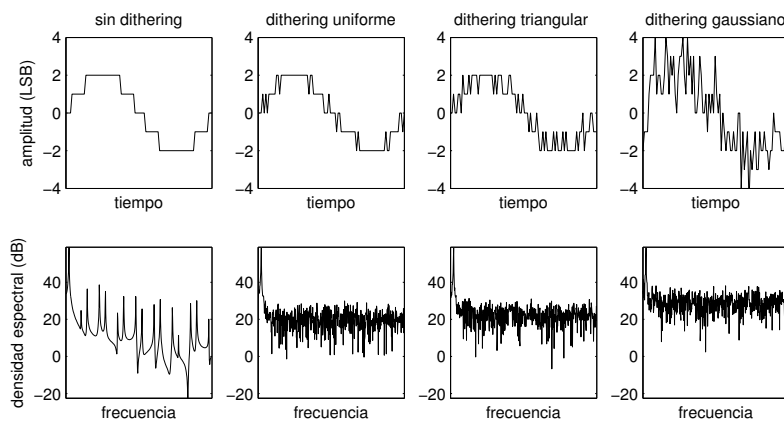
- agregando ruido de bajo nivel se elimina la correlación
- baja la SNR, pero disminuye distorsión (y aumenta resolución)



Cuantización

Dither

- tipos de dither: pdf rectangular, triangular, gaussiana*
- compromiso entre piso de ruido y eliminación de distorsión
- aplicación: cuantización, re-cuantización (e.g. 24 a 16 bits)



*Pohlmann, K. C. (2000). *Principles of Digital Audio*

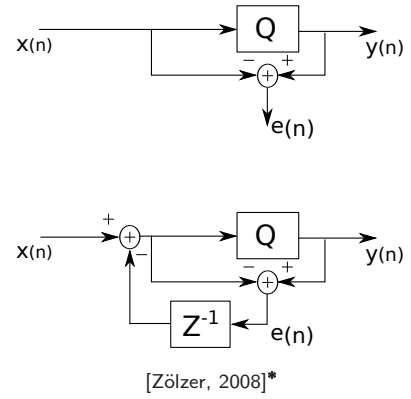
Cuantización

Noise shaping

- se puede procesar el dither para hacerlo menos audible (e.g. pasaltos)
- manipular ruido de cuantización y dither con criterios psicoacústicos

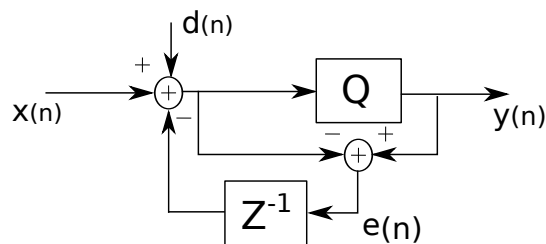
$$\begin{aligned}
 y(n) &= [x(n)]_Q = x(n) + e(n) \\
 y(n) &= [x(n) - e(n-1)]_Q \\
 &= x(n) - e(n-1) + e(n) \\
 &= x(n) + \hat{e} \quad \hat{e} = e(n) - e(n-1)
 \end{aligned}$$

filtro pasa altos: $H(z) = \frac{z-1}{z}$



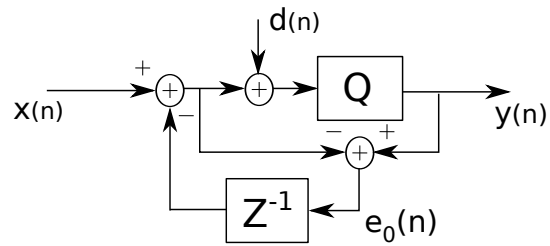
*Zölzer, U. (2008). *Digital Audio Signal Processing*

Noise shaping



$$\begin{aligned}
 y(n) &= [x(n) + d(n) - e(n-1)]_Q \\
 &= x(n) + d(n) - e(n-1) + e(n) \\
 \hat{e} &= d(n) + e(n) - e(n-1)
 \end{aligned}$$

Noise shaping






$$\begin{aligned}y(n) &= [x(n) + d(n) - e_0(n-1)]_Q \\ &= x(n) + d(n) - e_0(n-1) + e(n) \\ e_0(n) &= y(n) - (x(n) - e_0(n-1)) \\ &= d(n) + e(n) \\ y(n) &= x(n) + d(n) - d(n-1) + e(n) - e(n-1)\end{aligned}$$

Software

- Sonic Visualiser <http://www.sonicvisualiser.org/>
- Audacity <http://audacity.sourceforge.net/>

Referencias

-  Pohlmann, K. C. (2000).
Principles of Digital Audio.
-  Rabiner, L. R. and Schafer, R. W. (2011).
Theory and Applications of Digital Speech Processing.
-  Zölzer, U. (2008).
Digital Audio Signal Processing.