

Departamento de Computación, FCEyN, UBA

Procesamiento del Habla

Agustín Gravano

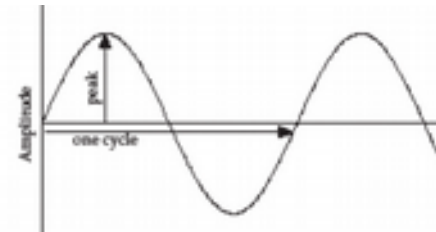
2do Cuatrimestre 2019

Acústica. Repaso de la clase anterior

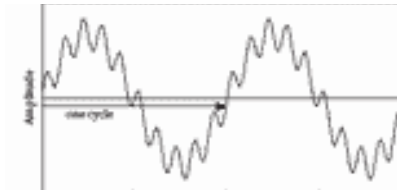
- **Sonidos periódicos y aperiódicos**

- Ondas periódicas simples

- Ciclo, período (T), frecuencia (f)

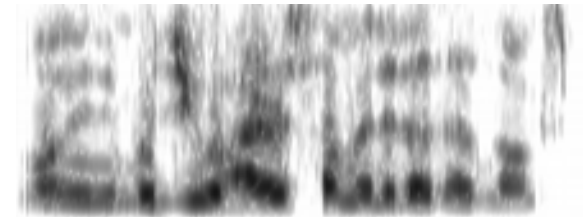


- Ondas periódicas complejas



- Ruido blanco, ondas transitorias

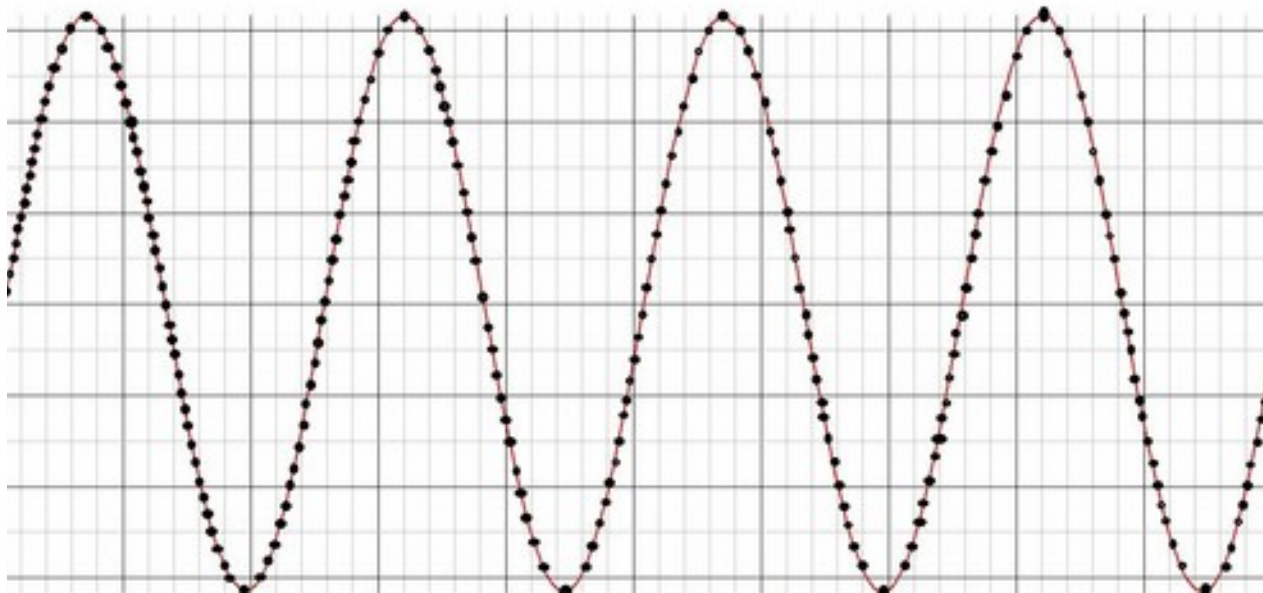
- Análisis de Fourier, FFT, Espectrograma



- **Ejercicios**

Procesamiento Digital de Señales

- **Señal analógica (continua)**: La línea de tiempo tiene valores de amplitud con precisión **infinita** en **todos** los puntos.
- **Señal digital (discreta)**: La línea de tiempo tiene sólo una **secuencia** de valores de amplitud con precisión **finita**.



Procesamiento Digital de Señales

- Un **micrófono** convierte oscilaciones de presión en el aire (sonido) en oscilaciones de voltaje.
 - Los dispositivos analógicos (discos de vinilo, cassettes) las guardan como señales continuas.
 - Los dispositivos digitales (computadoras, CDs) las convierten y guardan como señales discretas.
- Conversión Analógica-Digital (Digitalización)
 - 1) **Muestreo**: Discretización del tiempo.
 - 2) **Cuantización**: Discretización de la amplitud.

Conversión Analógica-Digital

- **Tasa de muestreo** (*sampling rate*)
 - ¿Cada cuánto hay que tomar muestras de la señal?
 - **Teorema de Nyquist-Shannon:** Para capturar la periodicidad de una onda con frecuencia f , es necesaria una tasa de muestreo **mayor que $2f$** .

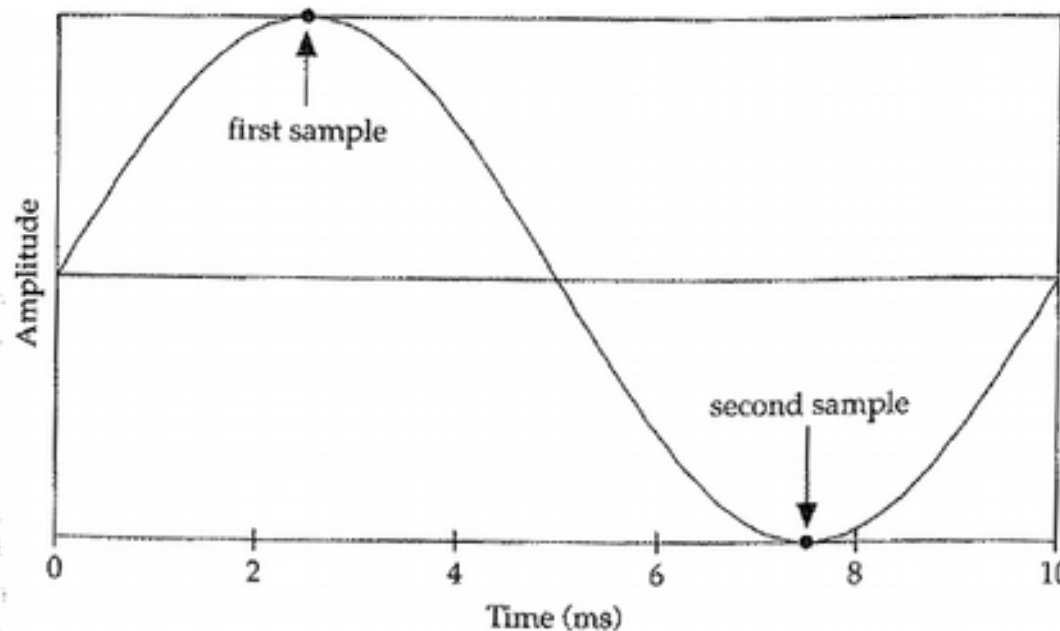


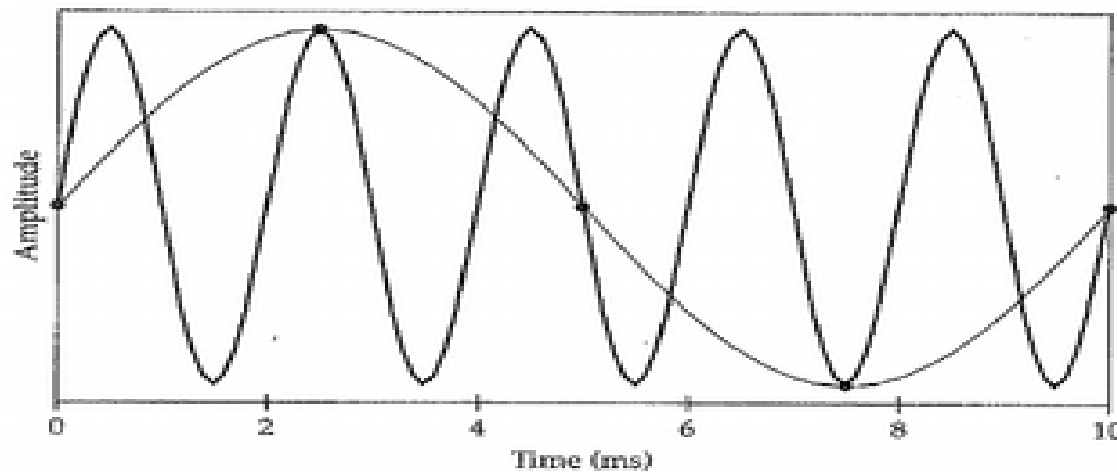
Figure 2.2 This figure illustrates why it takes two samples to capture the periodicity of a sine wave.

Conversión Analógica-Digital

- Balance entre muestreo y almacenamiento
 - Oído humano: máxima frecuencia ~20kHz
 - 44.1kHz: calidad de CD de audio
 - ¿Pero realmente necesitamos guardar 44k muestras por segundo si queremos almacenar habla?
 - Teléfono: [300 Hz, 4 kHz] (muestreo = 8 kHz).
 - Algunos sonidos del habla (fonos) tienen frecuencias mayores a 4 kHz: [s], [f].
 - Un tasa de muestreo de 16 kHz suele alcanzar para el procesamiento del habla.
- `sox --info IN.WAV`
 - Sample Rate : 16000

Conversión Analógica-Digital

- Error de muestreo: **aliasing**.
 - Ejemplos ópticos: rueda, turbina, agua, helicóptero.
 - Ocurre cuando la señal contiene frecuencias mayores a la **frecuencia de Nyquist** (mitad de la tasa de muestreo).



- Solución: Filtro anti-aliasing (ej.: *oversampling*).

Filtros Acústicos

- Bloquean sonidos de ciertas frecuencias.
 - **Filtro pasa-bajos (*low-pass*)**: Bloquea las componentes con frecuencia mayor a un umbral.
 - **Filtro pasa-altos (*high-pass*)**: Bloquea las componentes con frecuencia menor a un umbral.
 - **Filtro pasa-banda (*band-pass*)**: Bloquea las componentes con frecuencia por fuera de una banda.
- `sox IN.WAV OUT.WAV sinc FREQ`
 - Ejemplos de valores para FREQ:

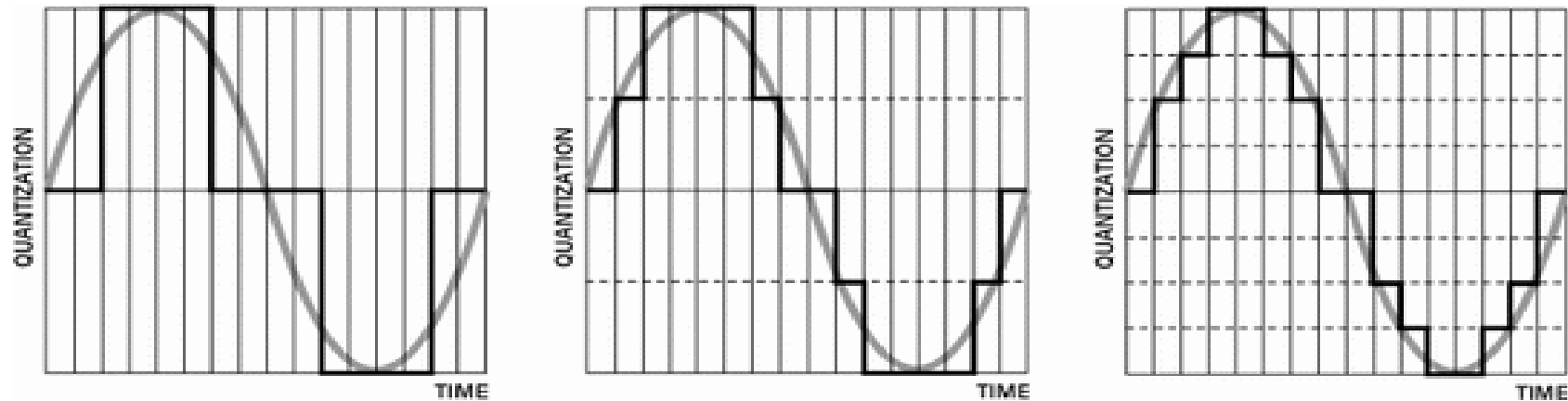
-4000	<i>low-pass</i>
4000	<i>high-pass</i>
3000-4000	<i>band-pass</i>

Conversión Analógica-Digital

- **Cuantización**
 - Las computadoras no tienen precisión infinita.
 - ¿Cuán precisas deben ser las muestras de amplitud que tomamos de la señal?
 - Ej.: 8, 12, 16, 32 bits por muestra
 - 256, 4096, 65536, 4294967296 niveles de amplitud.
- `sox --info IN.WAV`
 - Precision : 16-bit
- ¿Cuántos niveles es necesario distinguir?

Conversión Analógica-Digital

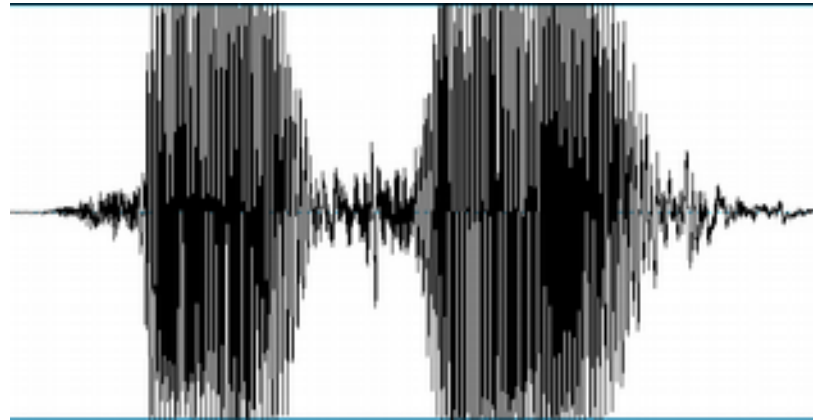
- Balance entre precisión de cuantización y almacenamiento.
 - Los **errores de cuantización** se reducen aumentando la precisión, pero a costa de más espacio.



- La elección depende de los datos y de la aplicación.
- Habla: 16kHz, 16bits suele ser razonable.

Conversión Analógica-Digital

- Problema derivado de la cuantización:
 - **Saturación digital** (*clipping*): La amplitud de la señal es mayor al rango representable.



- Solución #1: Redefinir los niveles de amplitud.
- Solución #2: Disminuir la amplitud de la fuente.

Ejercicios

- Escuchar bach.wav (44.1kHz, 16bits)
 - Fragmento de *Partita en Sol Mayor* de J. S. Bach.

```
play FILENAME
```

- Bajar sampling rate a 16, 8, 4 kHz y comparar.

```
sox IN.WAV -r FREQ OUT.WAV
```

- Subir sampling rate de 4 kHz a 44.1kHz.
 - ¿Por qué no vuelve a estar en buena calidad?
- Aplicar filtro high-pass de 8kHz a:
 - Audio original a 44.1 kHz.
 - Audio resampleado a 16 kHz.

- Crear espectrogramas de los audios y comparar.

```
sox IN.WAV -n spectrogram -o OUT.PNG
```

Variables Acústicas: Intensidad



Intensidad

- Ejemplo: `hola.wav`
- Nivel de presión del sonido.
- Puede medirse en:
 - Unidades de presión (Pa).
 - Unidades de voltaje (V).
- Es más frecuente usar **decibeles** (dB).
 - Escala logarítmica relativa a un nivel de referencia.
 - $20 \log_{10} (P / P_0)$ dB
 - Nivel de referencia $P_0 = 20$ micropascales = 2×10^{-5} Pa
 - Umbral de audición humana: “silencio”.

Percepción de la Intensidad

Evento	Presión (Pa)	Intensidad (dB)
Silencio	2×10^{-5}	0
Susurro	200	20
Oficina silenciosa	2K	40
Conversación	20K	60
Colectivo	200K	80
Subte	2M	100
Trueno	20M	120
DAÑO	200M	140

Cálculo de la Intensidad

- Sean x_i ($i = 1 \dots N$) muestras de la amplitud de (parte de) una señal.

- Amplitud RMS (*root mean square*) = $\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$

- **Intensidad** = $20 \log_{10} \frac{\text{RMS}}{P_0}$

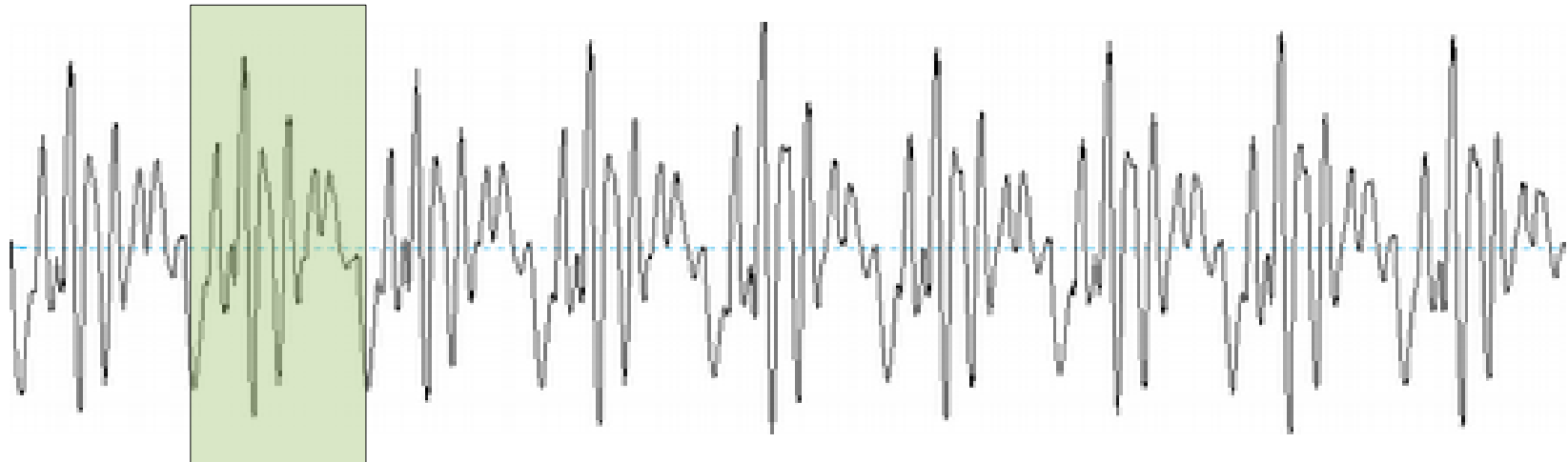
donde P_0 es el nivel de referencia para el silencio.

Variables Acústicas: Nivel tonal (*Pitch*)



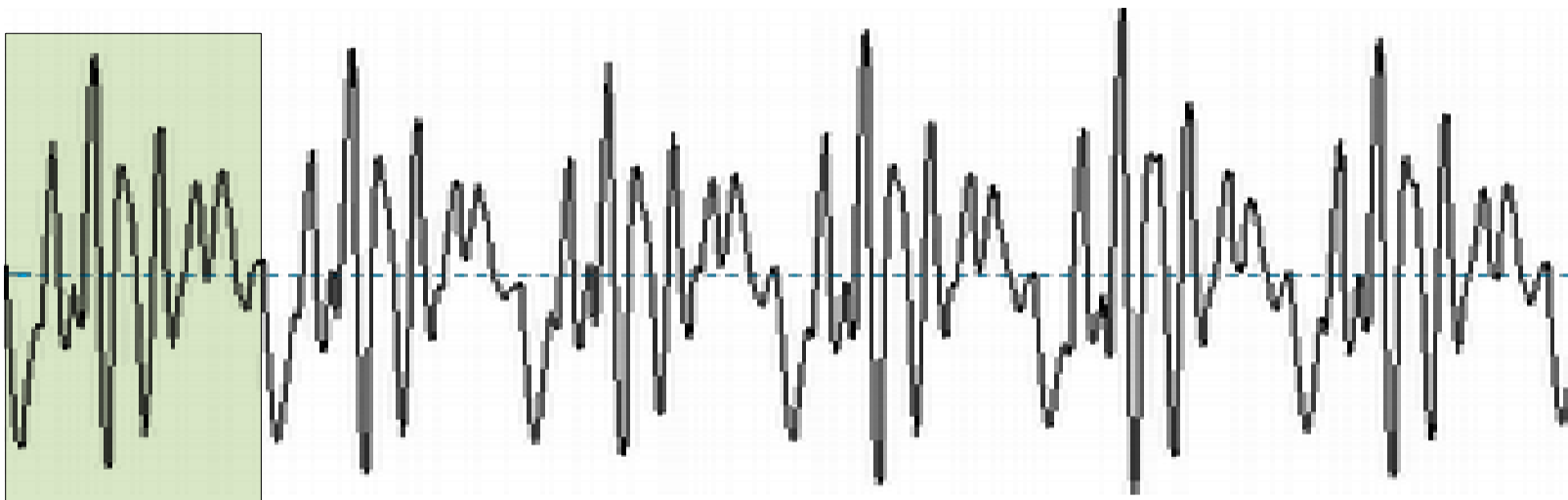
Percepción del Nivel Tonal (*Pitch*)

- Ejemplo: a.wav
- Frecuencia fundamental (F0): Frecuencia más baja de una onda periódica.
 - Tasa a la cual se repite el patrón complejo más chico.



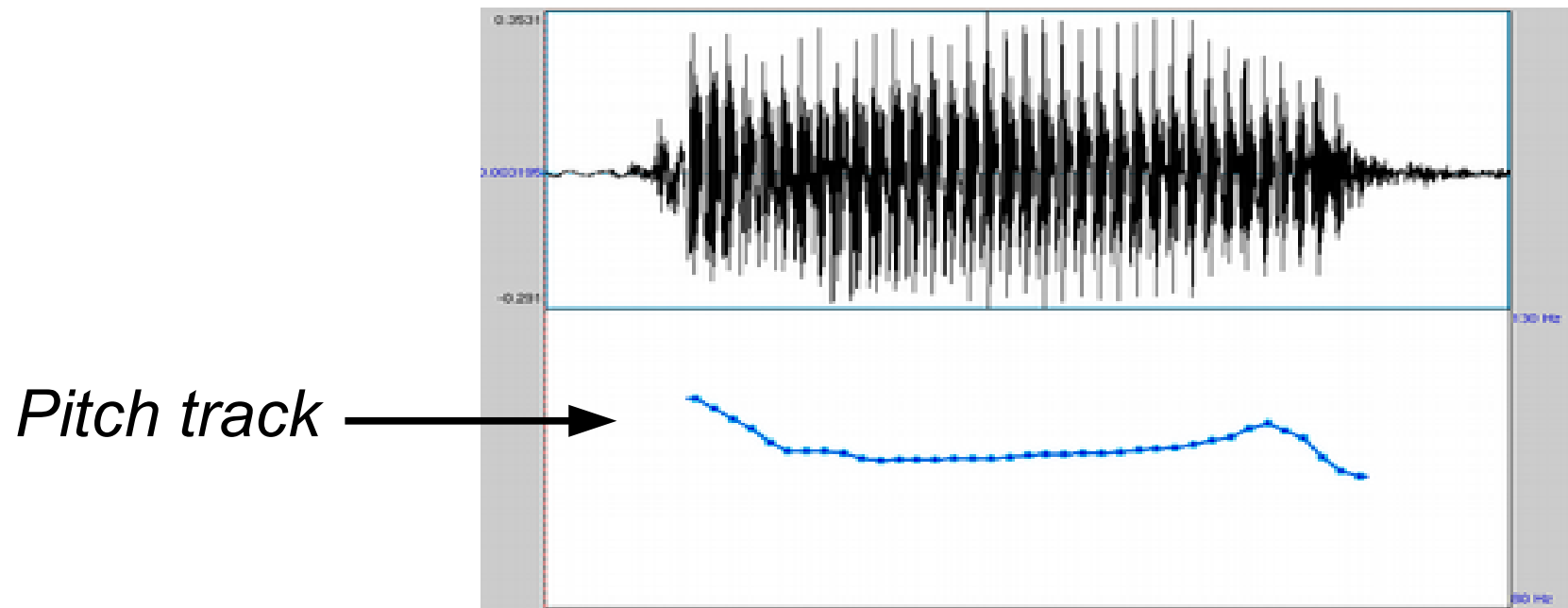
Estimación del Tono (*Pitch Tracking*)

- Método de **auto-correlación**
 - Una onda periódica se correlaciona consigo misma, dado que cada ciclo se parece mucho al siguiente.
 - Deslizar una copia de la onda hacia la derecha, hasta encontrar un punto de máxima correlación. El offset encontrado corresponde a la duración del período (T). La inversa ($1/T$) es la F_0 .



Estimación del Tono (*Pitch Tracking*)

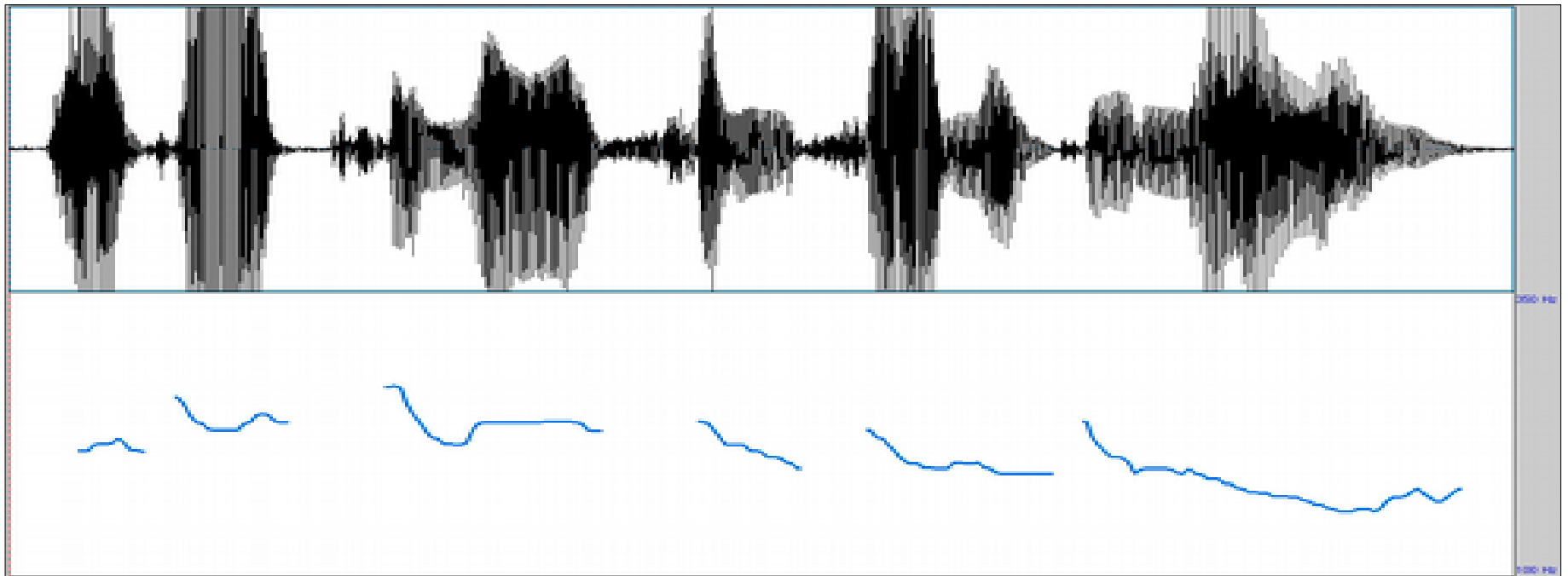
- Gráfico $F_0 \times \text{tiempo}$ = *Pitch track*
- Relacionado a la percepción del **nivel tonal**.



- Rango: mujeres 100-500Hz, hombres 75-300Hz.

Estimación del Tono (*Pitch Tracking*)

- Funciona bien para fonos sonoros: vocales, [m], [b], [l], etc. (ondas periódicas compuestas).
- Funciona mal para fricativas, oclusivas sordas, etc.: [s], [f], [t], [k], [tʃ] (sonidos aperiódicos).



Praat

En una terminal, ejecutar:

`praat`

Ignorar warnings en la terminal.

Cerrar la ventana “Praat Picture”. No la vamos a usar.

Intensidad y nivel tonal en Praat

Ejercicio 1:

- Abrir `/home/gravano/ph/archivos-senales/lamparita.wav`
- Hacer click en *View & Edit*.
- Menú *Intensity*
 - 1) Activar *Show intensity*. En *Intensity Settings* poner 50-100dB
 - 2) Seleccionar un segmento de habla.
 - 3) Click en *Intensity Listing* y en *Get intensity*.
- Menú *Pitch*
 - 1) Activar *Show pitch*. En *Pitch Settings* poner 75-500 Hz.
 - 2) Seleccionar un segmento de habla.
 - 3) Click en *Pitch listing* y en *Get pitch*.
- Para la primera y segunda /u/ (en “subí” y “un”), estimar a mano su F0 usando solamente la forma de onda. Comparar con los cálculos de Praat.

Pitch track vs. Espectrograma

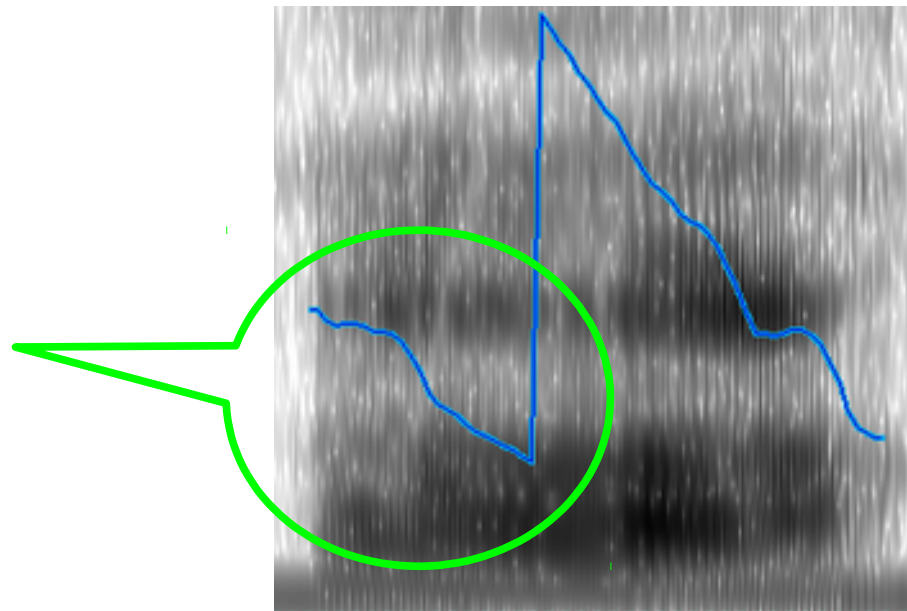
Ejercicio 2:

- Abrir `/home/gravano/ph/archivos-senales/a.wav`
- Escuchar (es un sonido `/a/` largo, subiendo y bajando el tono, primero de grave a agudo y luego al revés).
- Editarlo.
- Visualizar sólo el espectrograma.
 - ¿Cuál es el eje de referencia del espectrograma?
 - En *Spectrogram settings* poner rango = 0-5000 Hz.
 - ¿Qué ocurre con el espectrograma cuando el tono sube/baja?
- Visualizar sólo el Pitch track.
 - ¿Cuál es el eje de referencia del pitch track?
 - En *Pitch settings* poner rango = 150-500 Hz. ¿Qué ocurre con el pitch track?
 - Repetir con 50-150 Hz.

Errores de pitch halving y doubling

- ¿Qué ocurre con el pitch track?
 - Si el tono debería ser 150, pero marca 75: **pitch halving**.
 - Si el tono debería ser 150, pero marca 300: **pitch doubling**.
 - Errores en la estimación del tono (método de auto-correlación).
 - Usar un rango tonal adecuado ayuda a prevenir estos errores.

Ejemplo de
pitch halving



Scripting

- Desde la historia de la sesión:
 - *Praat* → *New Praat script* → *Edit* → *Paste history*
 - Se puede ejecutar todo o parte del script.
- Escribir scripts puede ser complicado.
- Modificar scripts existentes.
 - <https://lennes.github.io/spect/>
 - <http://uk.groups.yahoo.com/group/praat-users/>
 - <http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/praat.html>

Praat scripts

- En una terminal, ejecutar los siguientes comandos:

```
cd /home/gravano/ph/archivos-senales/  
praat duration.praat lamparita.wav
```

Devuelve la duración del archivo made1.wav, en segundos.

```
less duration.praat
```

Muestra el archivo duration.praat.

Praat scripts: duration.praat

```
# Praat script que toma como input un archivo de audio (.wav)  
# y devuelve su longitud en segundos.
```

```
# Argumento: archivo de audio.  
form Input parameters for sound length  
  word file .wav  
endform
```

```
# Los objetos 'long sound' no se levantan a memoria.  
Open long sound file... 'file$'
```

```
# Calcula la duracion.  
dur = Get duration
```

```
# La imprime y termina.  
echo 'dur:4'
```

Praat scripts: acoustics.praat

- En la terminal de Linux:

```
pwd
```

```
Directorio actual: /home/gravano/ph/archivos-senales/
```

```
praat acoustics.praat lamparita.wav 0.5 1.0 75 500
```

Computa un conjunto de mediciones acústicas para made1.wav entre 0.5 y 1.0 segundos, usando rango tonal 75-500Hz.

SECONDS:0.500	duración
F0_MAX:341.629	máxima f0 (Hz)
F0_MIN:247.602	mínima f0
F0_MEAN:311.274	media f0
F0_MEDIAN:317.807	mediana f0
F0_STDV:22.470	desvío estándar f0
ENG_MAX:83.361	máxima intensidad (dB)
ENG_MIN:46.801	mínima intensidad
ENG_MEAN:69.706	media intensidad
ENG_STDV:11.355	desvío estándar intensidad
VCD2TOT_FRAMES:0.532	proporción frames sonoros

Intensidad y nivel tonal en Praat

Ejercicio 3:

- Abrir el archivo **hola.wav** en Praat y tomar nota del comienzo y final de la /o/ en cada instancia de la palabra “hola”. ¡Sean precisos!
- Correr `acoustics.praat` sobre cada /o/:

```
praat acoustics.praat a.wav comienzo fin minpch maxpch
```

donde (minpch, maxpch) = (50, 300) para hombres, o
(75, 500) para mujeres.
- Comparar la intensidad en cada caso.

Intensidad y nivel tonal en Praat

Ejercicio 4:

- Crear una onda periódica compleja formada por dos ondas simples de 400 y 500 Hz.
 - *New > Sound > Create sound from formula...*
 - Formula: $\sin(2 * \pi * 400 * x) + \sin(2 * \pi * 500 * x)$
- Visualizar la forma de onda y computar F0 a mano.
- Visualizar el pitch track: Pitch > Show pitch, y comparar con el resultado del punto anterior.

Más ejercicios:

- ¿Dónde producimos el tono?
Grabarse diciendo las notas musicales *do, re, mi, fa, sol, la, si, do* en el tono correspondiente (aprox), pero **susurrando**. Editar el archivo y analizar el pitch track.
- Grabar las 5 vocales, como en el archivo “aeiou.wav”.
 - Grabar como aeiou-apellido.wav y traer la próxima clase.

Procesamiento Digital de Señales - Resumen

- Conversión analógica-digital, tasa de muestreo, precisión, teorema Nyquist-Shannon, aliasing, cuantización, saturación, filtros.
- Variables acústicas: intensidad (dB), nivel tonal o *pitch* (Hz).
- Herramientas: Praat y sox.