Instituto de Telecomunicaciones y Aplicaciones Multimedia

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Aplicaciones de audio en la nube para el Museo de la Telecomunicación "Vicente Miralles" (Audio on the cloud)

Vicent Molés Cases

Tabla de contenido

1 - Contexto 3

2 - Características generales 4

3 - Integración en el servidor web de la ETSIT 5

4 - Efectos utilizados 7

5 - Dispositivos simulados 10

- **Fonógrafo** 10
- **Gramófono** 11
- **Magnetófono de hilo** 13
- **Teléfono de baquelita** 15
- **Equipos radio** 16

1 - Contexto

El Museo de la Telecomunicación *"Vicente Miralles Segarra"* se encuentra situado en la planta baja y primera planta del edificio *4D* de la *Universidad Politécnica de Valencia* (*UPV*), en la sede de la *Escuela Técnica Superior de Ingenieros de Telecomunicación* (*ETSIT*). Contiene numerosos equipos de diferentes épocas, la mayoría de ellos son equipos de comunicación de voz y reproducción de sonido: Terminales telefónicos de telefonía fija, terminales telefónicos de comunicaciones militares, centralitas telefónicas de conmutación manual y automática, aparatos de radiodifusión de distintas épocas, gramófonos, etc.

El proyecto desarrollado, denominado *Audio-on-the-Cloud*, ha sido realizado en el seno del *Grupo de Tratamiento de Audio y Comunicaciones* (*GTAC*) del *Instituto de Telecomunicaciones y Aplicaciones Multimedia* (*iTEAM*) de la *UPV*. Éste, tiene como principal objetivo contextualizar los equipos y hacerse una idea de cómo funcionaban, mediante una serie de algoritmos que digitalmente producen una sensación sonora similar a la que producían los dispositivos del museo.

La aplicación está accesible en la propia página web del museo (link). La mencionada aplicación permite al usuario aplicar los efectos desarrollados a un audio escogido por el mismo. El funcionamiento básico de la aplicación consiste en que el usuario selecciona el audio que quiere procesar, éste audio es subido al servidor web de la *ETSIT* donde es procesado mediante los algoritmos desarrollados (*Audio-on-the-Cloud)*. Finalmente, el usuario puede descargar el resultado del procesado y escuchar el audio que había seleccionado previamente como si lo estuviera oyendo a través de uno de los dispositivos del museo.

En este proyecto se han desarrollado algoritmos para simular la sensación sonora producida por los dispositivos de audio y telefonía más representativos del museo, que son:

- Fonógrafo

- Gramófono

- Magnetófono de hilo

- Teléfono de baquelita

- Equipos de transmisión y recepción radio (tanto voz como morse)

MUSEO DE LA
TELECOMUNICACIÓN
VICENTE MIRALLES SEGARRA

2 - Características generales

En una primera fase del proyecto se utilizó *MATLAB* para determinar qué efectos y procesado (distorsiones, filtrados, etc) se tenían que aplicar para poder conseguir simular el comportamiento de cada uno de los dispositivos. Una vez realizada ésta tarea se decidió implementar mediante el lenguaje de programación C (uno de los más eficientes para el procesado de audio) los algoritmos correspondientes a cada uno de los dispositivos.

Los distintos programas desarrollados en C tienen ciertas características comunes, como:

- Formatos de audio de entrada:
  - **wav** y **flac**, mediante la librería *libsndfile 1.0.20* (para C).
    http://www.mega-nerd.com/libsndfile/
  - **mp3**, mediante el códec *lame 3.99* (línea de comandos*).
    http://lame.sourceforge.net/
  - **ogg/vorbis**, mediante el códec *oggdec 1.4.099* (línea de comandos).
    http://www.vorbis.com/
  - **aac**, mediante *faad 1.28* (línea de comandos).
    http://www.audiocoding.com/faad2.html

- Formatos de audio de salida:
  - **wav** y **flac**, mediante la librería *libsndfile 1.0.20* (para C).
  - **ogg/vorbis**, mediante el códec *oggenc 1.4.099* (línea de comandos).
    http://www.vorbis.com/
  - **aac**, mediante *faac 1.28* (línea de comandos).
    http://www.audiocoding.com/faac.html

(Para hacer las llamadas a los códecs de línea de comandos se ha hecho uso de la función *system()* de C que permite llamar a la ventana de comandos desde dentro de la ejecución de un programa)

- Para realizar el proceso de filtrado se ha utilizado el algoritmo *oversave* en todos los casos. Una de las características principales de dicho algoritmo es que para realizar el filtrado hace la *FFT* de la señal a filtrar y se realiza el producto de la señal y el filtro en el dominio frecuencial (en lugar de hacer la convolución en el dominio temporal). Para realizar el la *FFT* y la *IFFT* de una forma eficiente en C, se ha hecho uso de la librería fftw 3.3.3 (http://www.fftw.org/).

3 - Integración en el servidor web de la ETSIT

El primer paso para la integración de los programas diseñados en un servidor web es la compilación de los programas en el propio servidor. Para ello, el servidor debe tener instaladas las librerías y los códecs mencionados en el apartado anterior. Para compilar el código basta con situarse en la carpeta *Debug* del proyecto desde una ventana de línea de comandos:

```
$ cd /home/vicent/workspace/baquelita/Debug
```

Y ejecutar las dos siguientes líneas:

```
~/workspace/baquelita/Debug$ make clean
```

```
~/workspace/baquelita/Debug$ make all
```

Una vez compilado el código, el archivo resultante de la compilación se debe situar en el directorio adecuado (un directorio en el cual además del ejecutable, se encuentre la carpeta *src* que a su vez contenga la carpeta *effects*). Con la distribución de archivos correcta, se debe ejecutar el programa con los parámetros adecuados. A continuación se presentan los diferentes argumentos que se tiene que pasar a cada programa (cada programa tiene un *HELP* que indica los argumentos y su orden) y un ejemplo de llamada a cada programa:

- **Fonógrafo:**

HELP:

```
HELP:
./fonografo [path_in]
```

EJEMPLO:

```
./fonografo /home/vicent/Desktop/amelie.ogg
```

- **Gramófono:**

HELP:

```
HELP:
./gramofono [path_in]
```

EJEMPLO:

```
./gramofono /home/vicent/Desktop/amelie.ogg
```

- **Magnetófono de hilo:**

HELP:

```
HELP:
./magnetofono [path_in]
```

EJEMPLO:

```
./magnetofono /home/vicent/Desktop/amelie.ogg
```

- **Teléfono de baquelita:**

HELP:

```
HELP:
./baquelita [path_in]
```

EJEMPLO:

```
/baquelita /home/vicent/Desktop/amelie.ogg
```

- **Radio:**

HELP:

```
HELP:
./radio [path_in]
```

EJEMPLO:

```
./radio /home/vicent/Desktop/amelie.ogg
```

- **Radio morse:**

HELP:

```
HELP:
./radio_morse [text] -p [path_out]
```

EJEMPLO:

```
./radio_morse Hola como estas -p /home/vicent/Desktop/morse.ogg
```

En el servidor, el proceso que se sigue para realizar el procesado es el siguiente:

- El servidor recibe el archivo de audio subido por el usuario, le asigna un nombre aleatorio que identifica al archivo inequívocamente.

- Desde el código *php* del servidor se ejecuta uno de los programas presentados anteriormente, pasándole como parámetro de entrada el *path* del audio del usuario. Para realizar la llamada al programa desde *php* se utiliza el comando *exec()*.

- El audio de salida ya procesado y codificado con *ogg/vorbis* se guarda en el mismo directorio en el que se estaba almacenando el audio de entrada. El archivo de salida se guarda con el nombre: *[nombre_audio_entrada]_proc.ogg*. El path donde se encuentra el archivo de salida, se enlaza con un *tag* *<audio>* de *html 5*, de tal forma que el usuario puede reproducir y descargar el audio de salida.

4 - Efectos utilizados

Aunque cada dispositivo tiene unas características sonoras distintas, los efectos que se aplican en cada uno de los dispositivos tienen una base común (cambiando ciertos parámetros para cada caso). A continuación se presentan los efectos principales utilizados en los algoritmos:

- **Conversión estéreo - mono:** La mayoría de grabaciones digitales de hoy en día constan de 2 canales (*estéreo*), pero esto no ha sido siempre así. En los inicios de la grabación y reproducción analógica de audio (finales del siglo XIX) tan sólo se utilizaba un canal (*mono*). No fue hasta 1930, cuando se patentó la grabación y reproducción estéreo. Por ello, para conseguir un efecto sonoro realista mediante nuestros algoritmos hemos de hacer una conversión de estéreo a mono del audio digital de entrada. Así, realizamos un promediado de los dos canales digitales aplicando la siguiente ecuación:

$$\boldsymbol{x_{mono}} = \frac{\boldsymbol{x_{canal_1}} + \boldsymbol{x_{canal_2}}}{\mathbf{2}}$$

- **Filtrado digital:** En los equipos actuales de grabación y reproducción podemos llegar a tener una ancho de banda de grabación y reproducción muy grande (pueden llegar a tener un ancho de banda mayor a 20 *kHz*). En los antiguos equipos analógicos (como los que estamos tratando) esto no era posible. En ellos, la limitación en ancho de banda venía dada por el método de grabación, pero también por los soportes de reproducción que se utilizaban (normalmente magnéticos junto a bocinas). Cada uno de los dispositivos escogidos tiene una respuesta en frecuencia característica. Por ello, se analizaron grabaciones de los distintos dispositivos para obtener su respuesta en frecuencia y poder diseñar filtros digitales que consigan obtener una respuesta adecuada para el audio de salida del algoritmo.

- **Distorsión en amplitud:** La distorsión es uno de los efectos típicos introducidos por las grabaciones analógicas. En nuestro caso hemos utilizado una distorsión en amplitud que hace que la reproducción se asemeje más a la que se obtendría con los dispositivos del museo. Para obtener dicha distorsión, hemos aplicado la siguiente ecuación (que hace que los valores de amplitud de entrada que tengan valores cercanos, se junten más, teniendo de esta forma una menor riqueza de amplitudes en la salida distorsionada):

$$\boldsymbol{f(x)} = \boldsymbol{sign(-x)} \cdot \frac{\boldsymbol{x}}{\boldsymbol{|x|}} (\mathbf{1} - \boldsymbol{e}^{\boldsymbol{D} \cdot \frac{\boldsymbol{x}}{|\boldsymbol{x}|} \cdot \boldsymbol{sign(-x)}})$$

El parámetro *D* de la anterior ecuación determina el grado de distorsión que vamos a tener; en la figura 1 podemos ver la curva que representa la transformación de amplitud que provoca la distorsión (para un valor de *D = 3*):

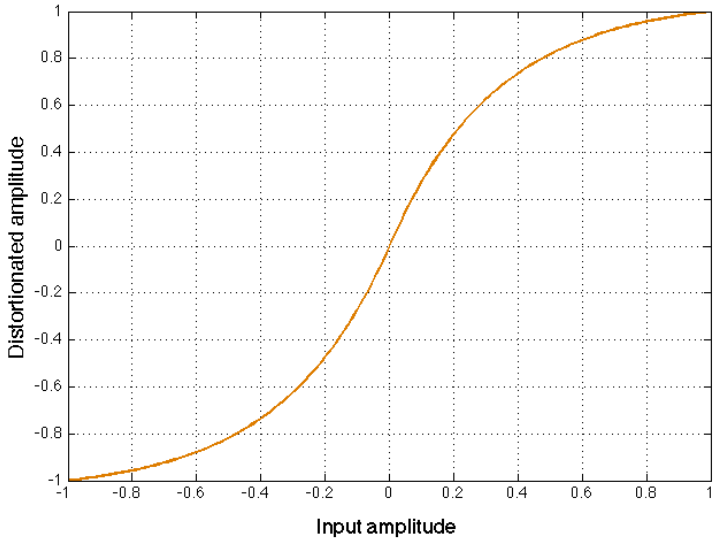

Figura 1. Curva de amplitudes de entrada y salida de la distorsión en amplitud.

Por último, en la figura 2 vemos el efecto que tendría la distorsión sobre un seno de frecuencia 100 *Hz*, fs = 44100 *Hz* y un D = 3:

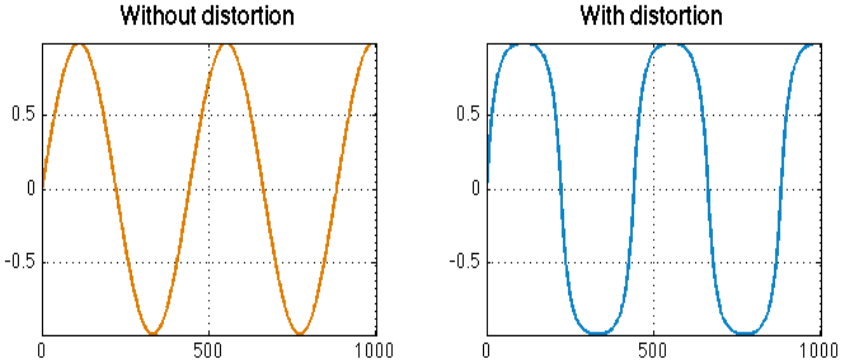

Figura 2. Efecto de la distorsión en amplitud en un seno de 100 Hz.

- ***Wow y Flutter*:** Son unos fenómenos que se dan en ciertos dispositivos de reproducción analógicos, debido a que la velocidad con la cual se recorre el soporte de grabación no es constante (por ejemplo, la velocidad con la que se recorre el cilindro en un fonógrafo o el disco en un gramófono). Esto provoca una pequeña modulación en frecuencia que hace que el espectro del audio de salida este ligeramente desplazado en frecuencia respecto al original (pensando en el dominio digital, se podría ver como una pequeña variación de la frecuencia de muestreo). La principal diferencia entre *wow* y *flutter* la encontramos en la frecuencia con la que se produce el desplazamiento en frecuencia:

    - *Wow*: se producen desplazamientos frecuenciales con una frecuencia de repetición baja (del orden de 5 a 10 Hz). Es consecuencia de cambios lentos en la velocidad con la que se recorre el soporte de reproducción

(por ejemplo en el gramófono se da este efecto si el disco no tiene una forma completamente circular, lo que provoca que la velocidad varíe ligeramente entre distintas posiciones del disco).

- *Flutter*: Se producen desplazamientos frecuenciales con una frecuencia de repetición alta (del orden de 100 a 500 Hz). Es consecuencia de variaciones muy rápidas de la velocidad con la que se recorre el soporte de producción debido a que los motores utilizados no tienen una velocidad constante (por ejemplo el motor de un magnetófono de hilo que puede producir pequeñas y rápidas variaciones de la velocidad).

Para simular los efectos anteriores, se ha desarrollado el algoritmo *vibrato.c*. Ésta función se basa en el algoritmo *vibrato* propuesto en la sección 3.3.1 del libro *DAFX – Digital Audio Effects*. La función desarrollada tiene los siguientes parámetros de entrada:

- La señal de entrada (x).
- La frecuencia de muestreo de la señal digital de entrada (f_s).
- La frecuencia con la que se repite el escalado $(Modfreq)$.
- Parámetro $Width$ (del que depende el grado de desviación en frecuencia).

El factor de desviación en frecuencia que produce la mencionada función en cada instante sigue la siguiente expresión:

$$\alpha[n] = 1 - Width \cdot 2 \cdot \pi \cdot Modfreq \cdot \sin\left(2\pi \cdot \frac{Modfreq}{f_s} \cdot n\right)$$

- ***LPC* (*Linear predective coding*):** La codificación *LPC*, nos proporciona los coeficientes del filtro *IIR* que tenemos que utilizar para filtrar ruido *AWGN* y obtener un distribución espectral muy similar a la del fragmento original (fragmento de entrada). Así, si disponemos de un fragmento del ruido de un dispositivo, sólo tenemos que aplicar la codificación *LPC* a dicho fragmento, para obtener un filtro *IIR* que al filtrar el ruido *AWGN* haga que éste deje de ser blanco y tenga una distribución espectral muy similar a la del fragmento de entrada. Con esta herramienta podemos simular los ruidos característicos de cada uno de los dispositivos.

5 - Dispositivos simulados

Fonógrafo

El primer prototipo de fonógrafo data 1877 y fue diseñado por Thomas Alva Edison. Este dispositivo supuso una revolución en el campo del audio, ya que fue el primer dispositivo capaz de grabar y reproducir audio.

Figura 3. Fonógrafo del museo de la ETSIT.

El funcionamiento básico del fonógrafo para la grabación de audio se basa en la utilización de un transductor acústico-mecánico, que se encarga de transformar las ondas sonoras en vibraciones mecánicas. Dichas vibraciones son utilizadas por un estilete, el cual graba sobre un cilindro metálico y por lo tanto recrea las vibraciones mecánicas en el cilindro. Para la reproducción del sonido el proceso es el inverso, es decir el estilete obtiene del cilindro las vibraciones mecánicas y las transmite a un sensor mecánico-acústico que transforma las vibraciones mecánicas en sonido.

A continuación se van a presentar los distintos efectos utilizados para la simulación del sonido de un fonógrafo:

- **Conversión estéreo – mono.**

- ***Flutter*:** Debido a que la velocidad de giro del cilindro sobre el estilete del fonógrafo no es perfectamente constante, provocando una pequeña modulación en frecuencia que da lugar al *Flutter*. En los algoritmos se ha establecido 100 Hz como la frecuencia de repetición para este efecto. Este efecto provoca cierta aleatoriedad y una leve metalización del sonido.

- **Distorsión en amplitud:** Se aplica la anteriormente mencionada distorsión en amplitud para intentar reducir el número de amplitudes (la riqueza en amplitudes en el intervalo [-1,1]). El parámetro de compresión D utilizado va variando aleatoriamente en cada instante de tiempo, siendo su valor máximo 30.

- **Filtrado paso banda:** La limitación en ancho de banda principal del fonógrafo viene dada por el soporte de grabación, pero también por la respuesta de las bocinas que se utilizaban junto al fonógrafo. De esta forma, la respuesta solía ser buena en el rango de 400 a 2000 Hz. Por ello, hemos aplicado al audio de entrada el filtro cuya respuesta en frecuencia podemos ver en la figura 4.

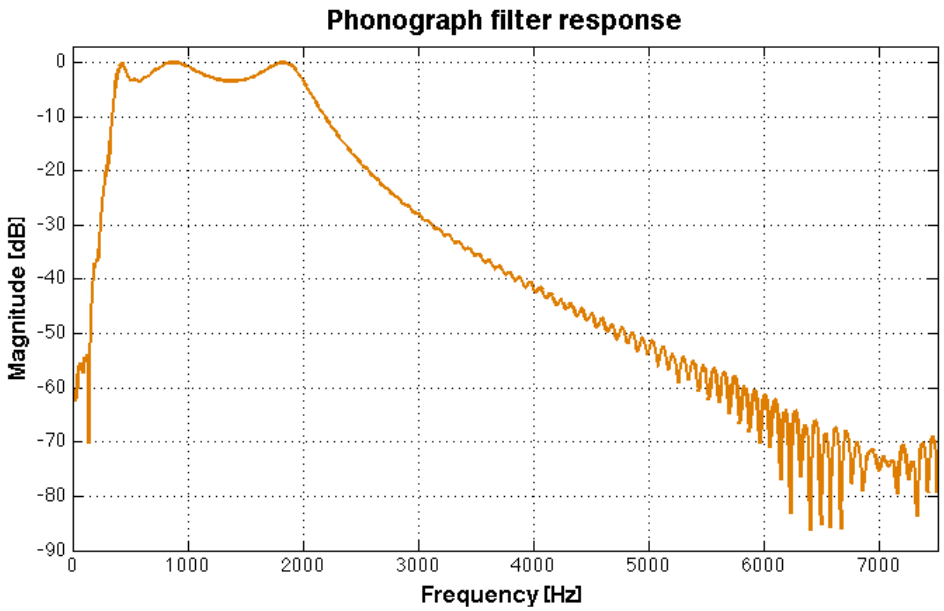

Figura 4. Respuesta del filtro utilizado para simular el fonógrafo.

- **Ruido periódico:** El mecanismo del que dispone el fonógrafo para rotar el cilindro y que de esta forma el estilete recorra su superficie está formado por engranajes y cintas. Dicho mecanismo produce un ruido periódico muy difícil de simular mediante procedimientos de tratamiento digital de la señal, por ello se determinó que la mejor solución era extraer un periodo completo de dicho ruido de una de las grabaciones de las que se dispone y repetirlo periódicamente durante toda la reproducción, multiplicando por una amplitud aleatoria para cada periodo.

Gramófono

El gramófono es un dispositivo de grabación y reproducción de audio analógico que fue patentado en 1888 por Emile Berliner. Su funcionamiento es bastante similar al del fonógrafo (utiliza un sensor acústico-mecánico y una aguja para aplicar las vibraciones mecánicas al soporte de grabación) pero con la característica de que en este caso el soporte de grabación y reproducción es un disco plano. El factor que hizo que el gramófono se impusiera al fonógrafo fue que a partir de un solo disco plano se podían hacer tantas copias como se quisieran, cosa que con el fonógrafo no era posible. Como contrapartida, con el gramófono los usuarios no podían realizar sus propias grabaciones ya que los discos planos, una vez grabados, necesitan la aplicación de un proceso químico para poder ser reproducidos en el gramófono.

Figura 5. Gramófono del museo de la ETSIT.

A continuación se van a presentar los distintos efectos utilizados para la simulación del sonido de un gramófono:

- **Conversión estéreo – mono.**

- ***wow*:** Este efecto se suele producir porque la aguja no se encuentra completamente centrada en la ranura del disco y por lo tanto la velocidad con la que la aguja pasa por la ranura no es constante dentro de una misma vuelta. Como los gramófonos tenían habitualmente una velocidad de giro de 33 rpm, para hacer más realista el efecto hemos hecho que el periodo de repetición del wow coincida con dicha velocidad. Así, tendremos una frecuencia de repetición de:

$$f_{rep} = \frac{33\ r.p.m}{60\ s/min} = 0.55\ Hz$$

- **Distorsión en amplitud:** En este caso, hemos utilizado la distorsión en amplitud para reducir el rango dinámico del audio de entrada, buscando que la relación entre la potencia de pico y la potencia media del audio procesado no sea mayor que un factor 2.5.

- **Filtrado paso banda:** El ancho de banda del gramófono va de 350 Hz a 2900 Hz. Así, hemos aplicado al audio de entrada el filtro cuya respuesta en frecuencia podemos ver en la figura 6.

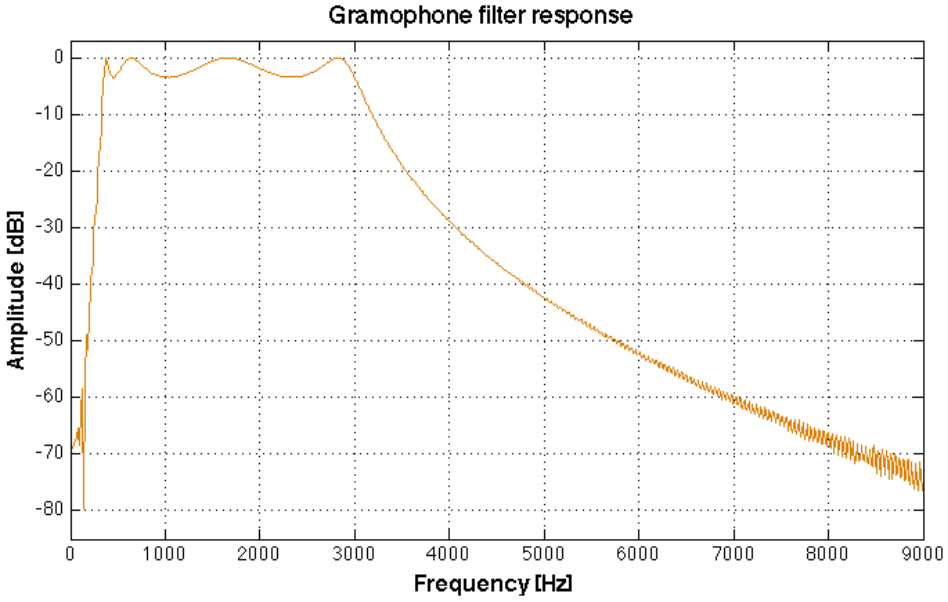

Figura 6. Respuesta del filtro utilizado para simular el gramófono.

- **Ruido:** Para simular el ruido introducido por el gramófono utilizamos la codificación *LPC*. Con ella obtuvimos el filtro cuya respuesta podemos ver en la figura 7. Este filtro aplicado a ruido *AWGN* nos proporciona un ruido con la misma

distribución espectral que el original. Para obtener un ruido lo más realista posible, se decidió fragmentar el ruido en tramas de 100 ms aplicando una amplitud aleatoria a cada una de estas tramas.

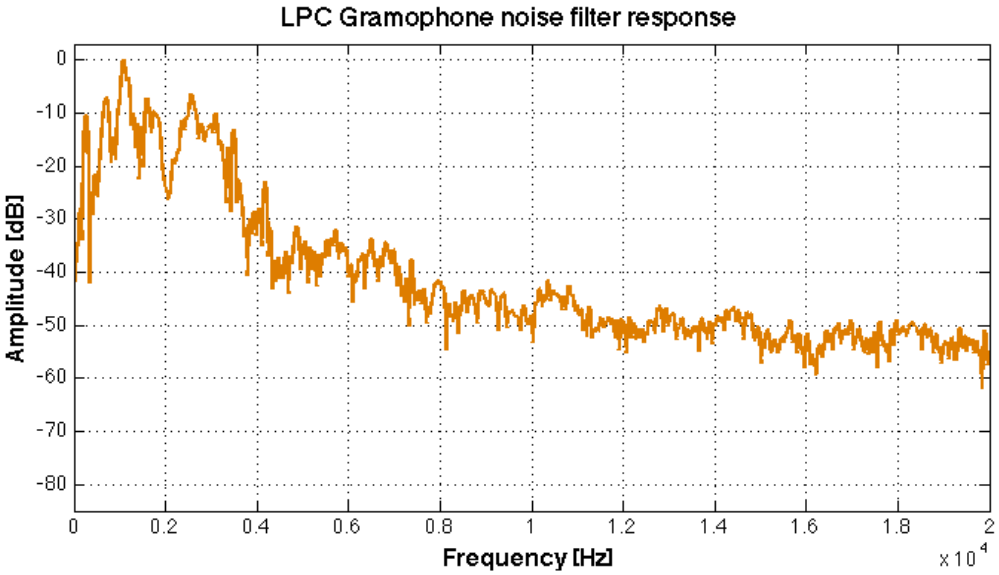

Figura 7. Respuesta del filtro mediante la LPC del ruido del gramófono.

Magnetófono de hilo

El magnetófono de hilo fue el primer dispositivo de grabación magnética de sonido que se comercializó. El funcionamiento del dispositivo se basaba en un hilo metálico que se magnetizaba a través de un magnetófono eléctrico de cabeza circular. La fuerza magnética aplicada al hilo dependía de la intensidad del sonido de entrada, quedando almacenado magnéticamente el audio en el hilo metálico. Para la reproducción, se convertía el magnetismo almacenado en el hilo en corriente eléctrica, la cual alimentaba a un altavoz.

Figura 8. Magnetófono de hilo del museo de la ETSIT.

Seguidamente se muestran los distintos efectos utilizados en la simulación del magnetófono de hilo:

- **Conversión estéreo – mono.**

- ***Flutter.***: Debido a que la velocidad de giro del motor del magnetófono de hilo no es constante, produciéndose pequeñas variaciones de velocidad, tenemos que tener en cuenta el efecto *Flutter* para la simulación de este dispositivo. En este caso, se ha seleccionado una frecuencia de repetición de 80 Hz.

- **Distorsión en amplitud:** La grabación magnética sobre un hilo presentaba muchas limitaciones. Una de ellas la de precisión en la amplitud del audio grabado, por ello utilizamos el efecto de distorsión en amplitud con un parámetro *D* igual a 8.

- **Filtrado paso banda:** Se ha aproximado la respuesta del magnetófono de hilo con un filtro paso banda que tiene como frecuencia de corte inferior 200 Hz y superior 5 kHz. En la figura 9 se puede ver la mencionada respuesta frecuencial.

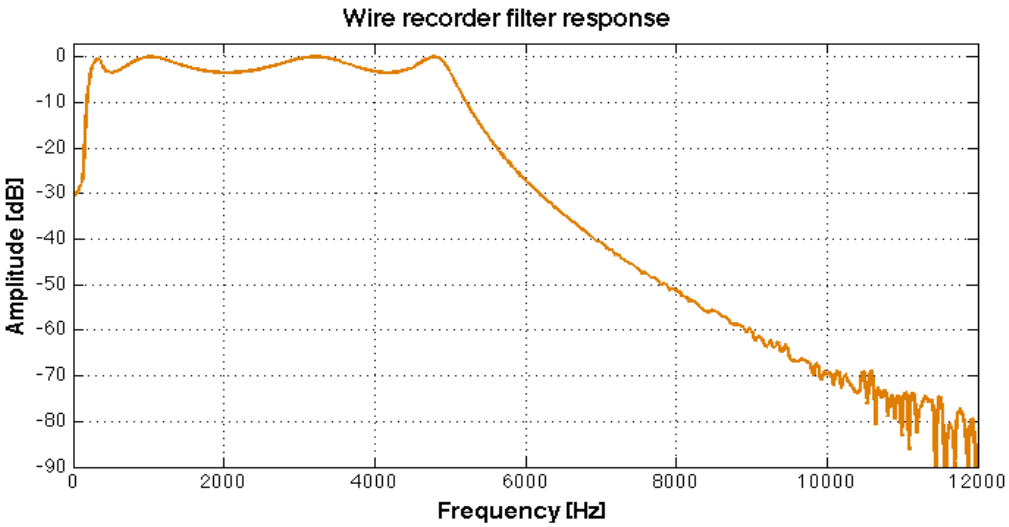

Figura 9. Respuesta del filtro utilizado para simular el magnetófono de hilo.

- **Ruido:** Para la simulación del ruido del magnetófono de hilo, se ha aplicado una codificación *LPC*, obteniendo el filtro cuya respuesta podemos ver en la figura 10.

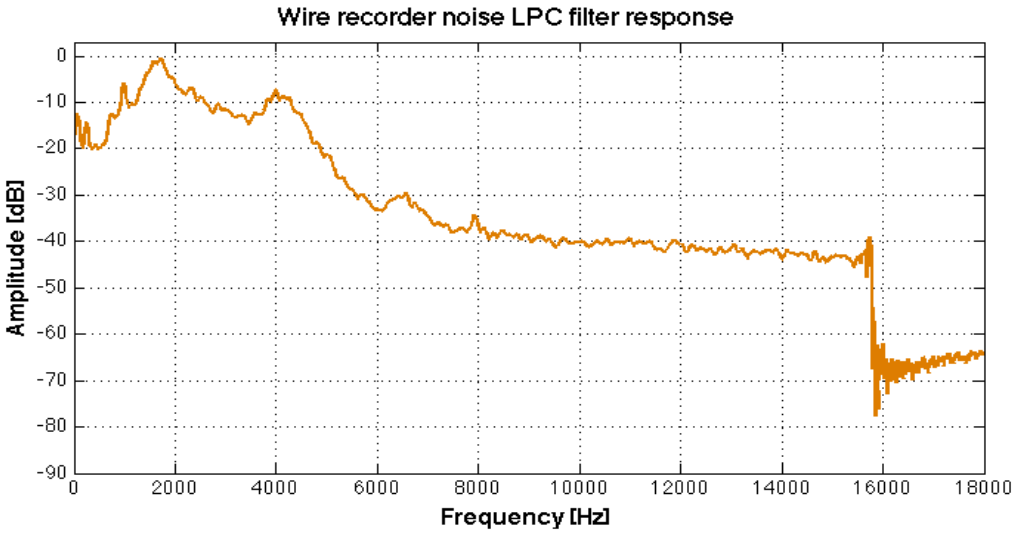

Figura 10. Respuesta del filtro obtenido con la LPC del ruido del magnetófono de hilo.

Teléfono de baquelita

Los primeros teléfonos de baquelita empezaron a fabricados por la compañía Sueca Ericsson en 1931. Este tipo de dispositivos supusieron un gran avance en la telefonía fija y fueron los teléfonos más comunes y usados durante más de medio siglo. La respuesta de estos dispositivos (tanto de su altavoz como su micrófono) está pensada sólo para voz y no para música, así, la banda eliminada de su respuesta empieza en 3000 *Hz*.

Figura 11. Teléfono de baquelita del museo de la ETSIT.

A continuación se van a presentar los distintos efectos utilizados para la simulación del sonido de un teléfono de baquelita:

- **Conversión estéreo – mono.**

- **Ruido y filtro:** Como se disponía del audio generado por el teléfono de baquelita cuando el audio de entrada era ruido *AWGN*, aplicando una simple codificación *LPC* pudimos obtener el filtro que modela el comportamiento del teléfono, cuya respuesta podemos ver en la figura 12. Una vez conocida la respuesta del teléfono, tan solo tenemos que sumar la señal mono de entrada con ruido *AWGN* (con el nivel de *SNR* adecuado, para nuestro caso 60 dB) y filtrar dicha suma con el filtro obtenido, obteniendo la señal deseada (como representa la siguiente ecuación).

$$y[n] = (x[n] + n_{AWGN}[n]) * h[n]$$

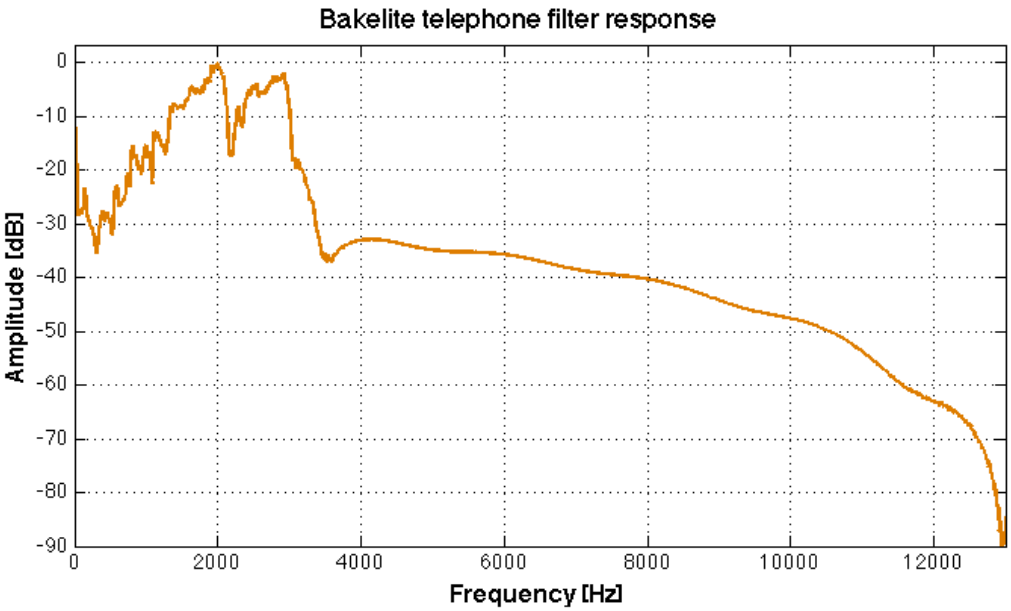

Figura 12. Respuesta del filtro utilizado para simular el teléfono de baquelita.

Equipos radio

En este programa se ha simulado el efecto sonoro introducido por un transceptor de Marconi como el que podemos ver en la figura 13. En transceptor de Marconi era un equipo de campaña para uso militar datado de 1949 cuya función principal fue transmitir telegrafía y telefonía. En el modo de telefonía se podía operar de dos formas concretas: transmitir o transmitir y recibir al mismo tiempo. Trabajaba en las bandas de frecuencia de 2,7 MHz a 6MHz.

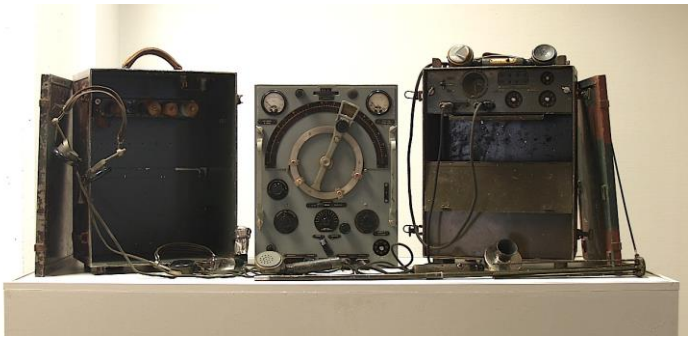

Figura 13. Transceptor Marconi del museo de la ETSIT.

Seguidamente se muestran los distintos efectos utilizados en la simulación del transceptor Marconi:

- **Conversión estéreo – mono.**

- **Filtrado paso banda:** Se ha aproximado la respuesta del transceptor Marconi con un filtro paso banda que tiene como frecuencia de corte inferior 200 Hz y superior 2 kHz. Dicho filtro se ha obtenido al realizar la codificación *LPC* de una muestra de ruido del transceptor Marconi En la figura 14 se puede ver la mencionada respuesta frecuencial.

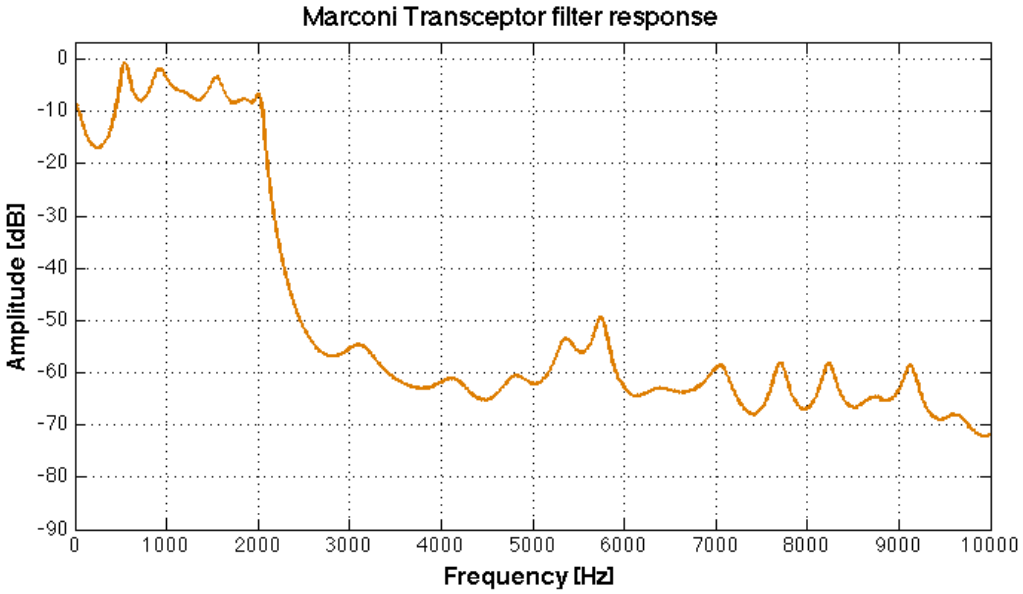

Figura 14. Respuesta del filtro utilizado para simular el transceptor Marconi.

- **Ruido característico:** El ruido introducido por el transceptor Marconi tiene unas características muy concretas que con la codificación *LPC* no se conseguían completamente. Por ello, se decidió filtrar el ruido *AWGN* directamente con una muestra representativa del ruido, consiguiendo así el ruido deseado. En la figura 15 se presenta la respuesta frecuencial de las muestras utilizadas para realizar el filtrado.

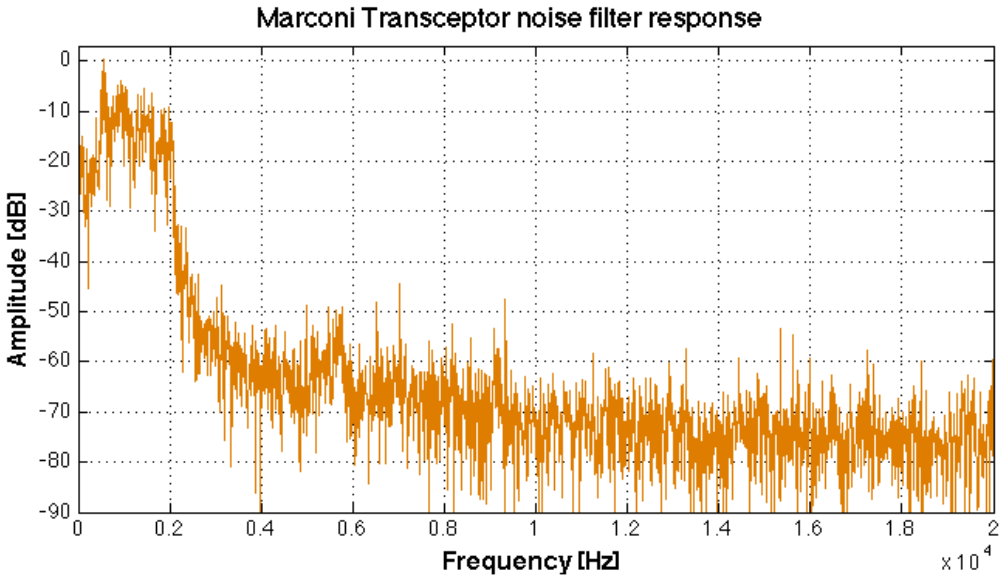

Figura 15. Respuesta del filtro utilizado para simular el ruido del transceptor Marconi.

- **Morse:** Para simular la transmisión en morse, se ha utilizado la tabla de codificación que se puede ver en la figura 16. El código morse codifica cada una de las letras del alfabeto con intervalos largos y cortos como se ve en la tabla. En la figura 17 se puede ver la respuesta temporal del intervalo largo y corto que se ha utilizado para la simulación de la transmisión. Como se puede ver la duración del intervalo corto es de 140 *ms*, mientras que la duración del intervalo largo es de 375 *ms*.

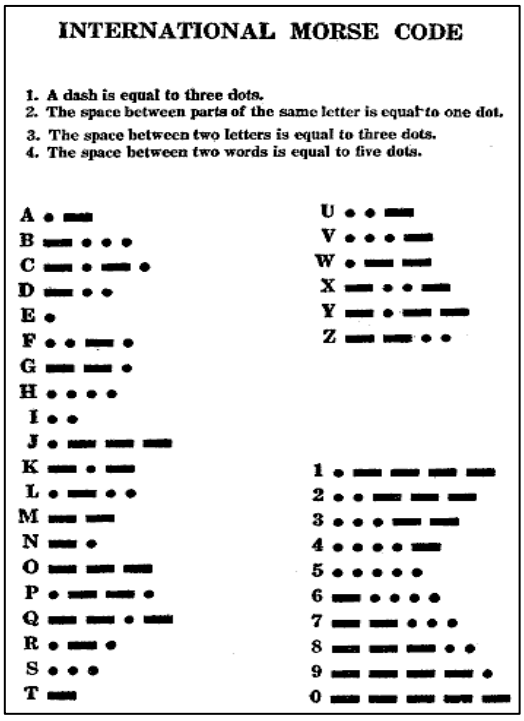

Figura 16. Tabla de codificación Morse.

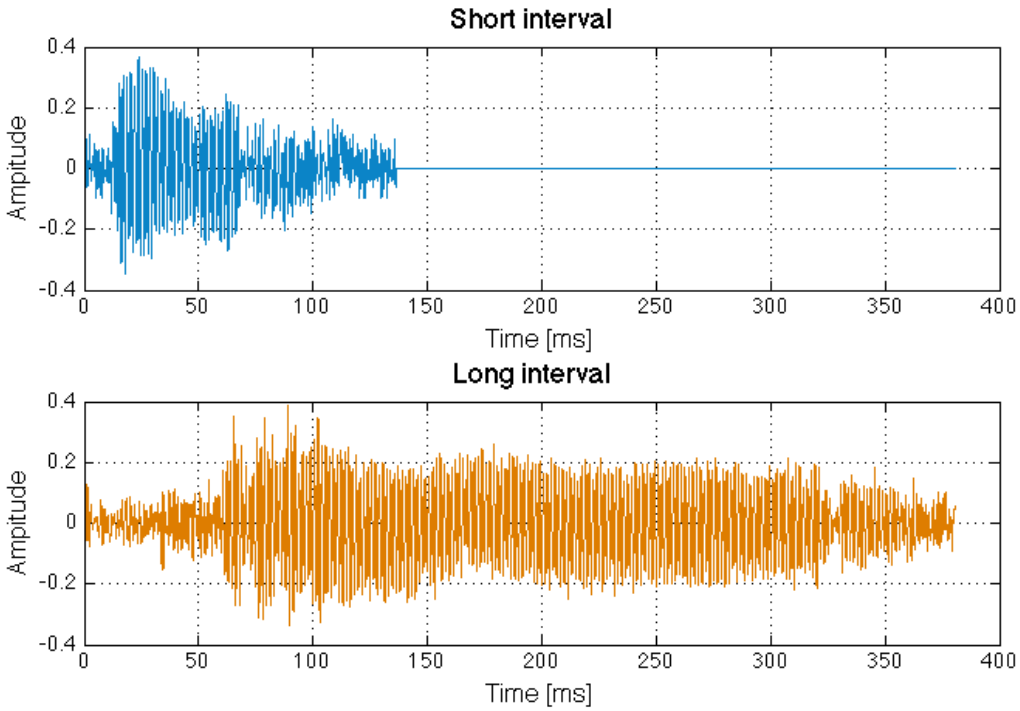

Figura 17. Intervalos de la transmisión Morse.