

Informe Pràctica Àudio

Ernest Anguera Aixalà

Mariona Farré Tapias

Naïm Barba Morilla

Adrián Romera González

Índex

| | |
|------------------------------------------------------------------|----|
| Introducció | 4 |
| Exercici 1 – Fitxers PCM en cru (“raw”) | 5 |
| Apartat a | 5 |
| Apartat b | 8 |
| Fonemes sonors | 8 |
| Fonemes sords | 9 |
| Apartat c..... | 11 |
| Signed 16 bit, Little-endian, 1 canal, 22050 Hz..... | 11 |
| Unsigned 8 bit, Little-endian, 2 canals (Stereo), 22050 Hz | 11 |
| Unsigned 8 bit, Little-endian, 1 canal, 44100 Hz..... | 11 |
| Exercici 2 – Codificant amb FLAC i vocoders..... | 12 |
| Apartat a | 12 |
| Apartat b | 13 |
| Voice..... | 13 |
| Music..... | 14 |
| Voice-tel | 15 |
| Music-tel | 16 |
| Apartat c..... | 17 |
| Apartat d | 18 |
| Apartat e | 18 |
| Apartat f | 19 |
| Apartat g | 19 |
| Nivells de configuració de FLAC..... | 19 |
| Diferent bitrate per AMR..... | 19 |
| Exercici 3 – Comprimint àudio amb MPEG | 20 |
| Apartat a | 20 |
| Apartat b | 20 |
| Apartat c..... | 20 |
| Exercici 4 – Efectes de MPEG sobre la forma d'ona | 21 |
| MP3..... | 22 |
| Conclusió | 22 |
| Recomanacions | 22 |
| Observacions addicionals | 22 |

| | |
|--------------------------------------|----|
| Exercici 5 – MP3 VBR | 23 |
| Apartat a | 23 |
| Apartat b | 25 |
| Apartat c | 25 |
| Apartat d | 28 |
| Exercici 6 - Streaming d'àudio | 29 |
| Apartat a | 29 |
| Apartat b | 32 |
| Conclusió..... | 35 |

Introducció

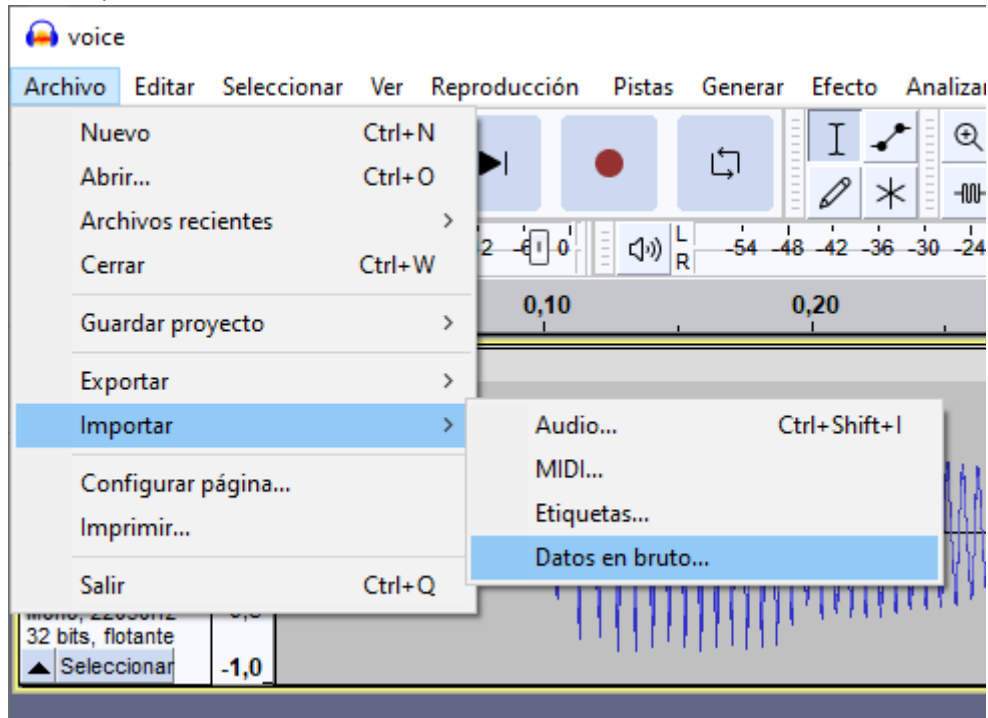
En aquesta unitat pràctica, ens endinsarem en el món de la compressió i el streaming d'àudio, explorant les tècniques i tecnologies que permeten la transmissió eficient de so a través de diferents mitjans. Des de l'anàlisi de formes d'ona fins a l'ús de còdecs avançats, aquest document proporciona una comprensió profunda dels principis fonamentals que regeixen la compressió d'àudio i les seves aplicacions pràctiques. Amb exercicis detallats i anàlisis exhaustius, els estudiants obtindran una visió clara de com els diferents formats d'àudio i els seus paràmetres afecten la qualitat i la mida dels fitxers, així com la seva transmissió en temps real.

Exercici 1 – Fitxers PCM en cru (“raw”)

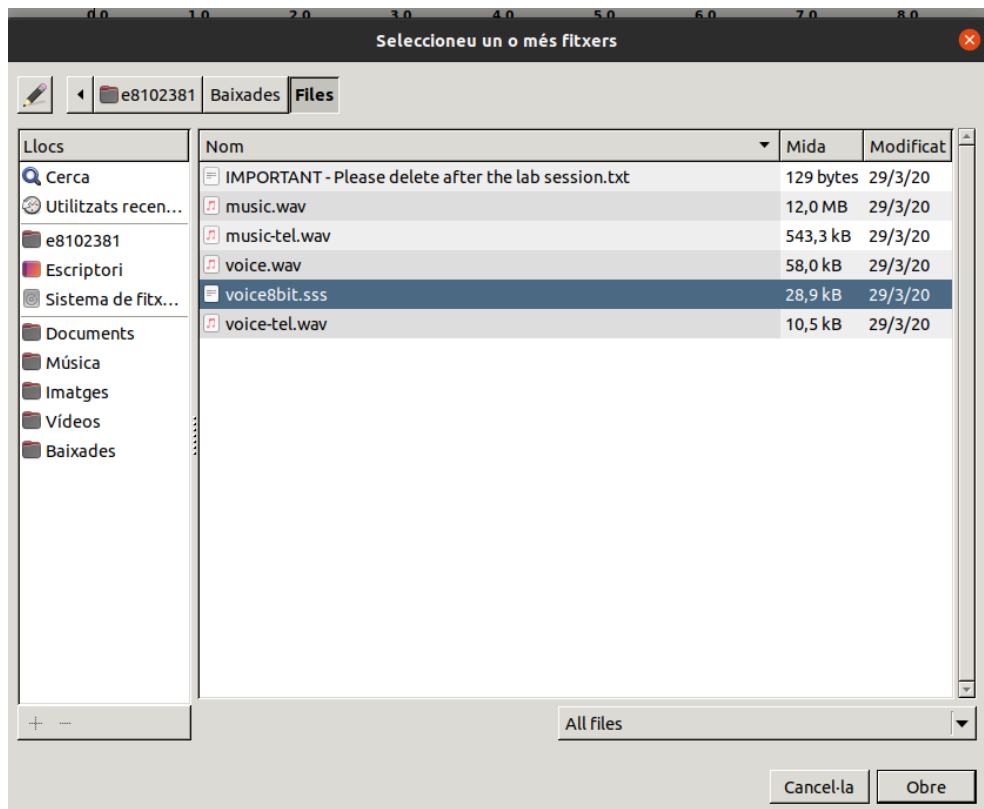
Apartat a

Per a poder obrir l'arxiu en cru amb Audacity, hem d'importar aquest arxiu amb les dades correctes, per a fer-ho, seguim els següents passos:

1. Clicar a importar -> Dades en cru:



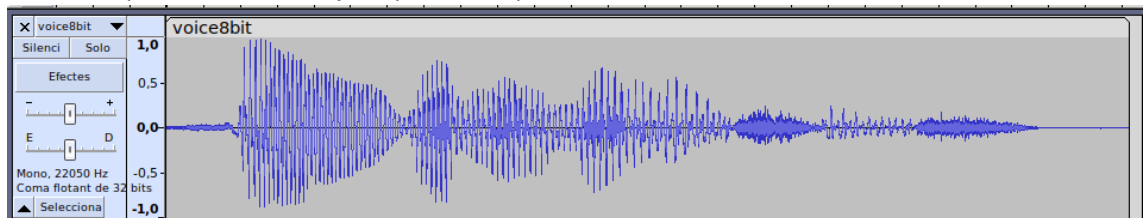
2. Seleccionem l'arxiu voice8bit.sss:



3. Un cop seleccionat, hem d'escollir les opcions més adequades:

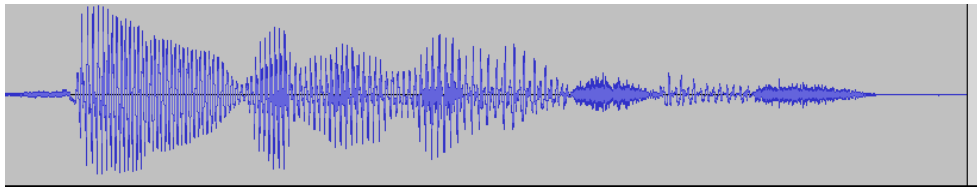


4. Amb les opcions escollides, ja el podem importar:

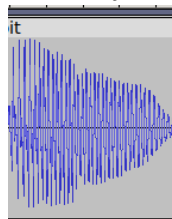


5. Forma d'ona:

- a. Observant la forma d'ona de l'arxiu, podem veure que hi ha diferents seccions clarament diferenciades.

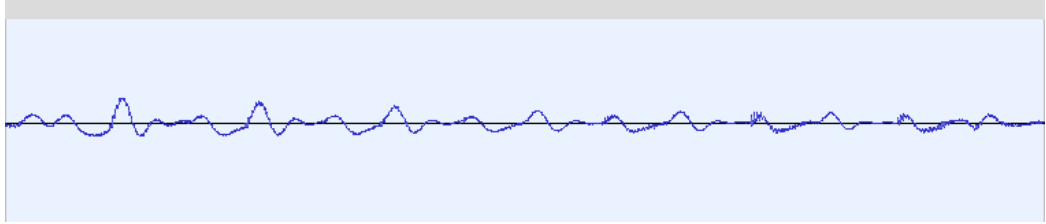


En aquesta imatge anterior, es pot observar que l'ona té una freqüència d'ona major en les zones en les que hi ha fonemes molt junts, com per exemple en la secció:

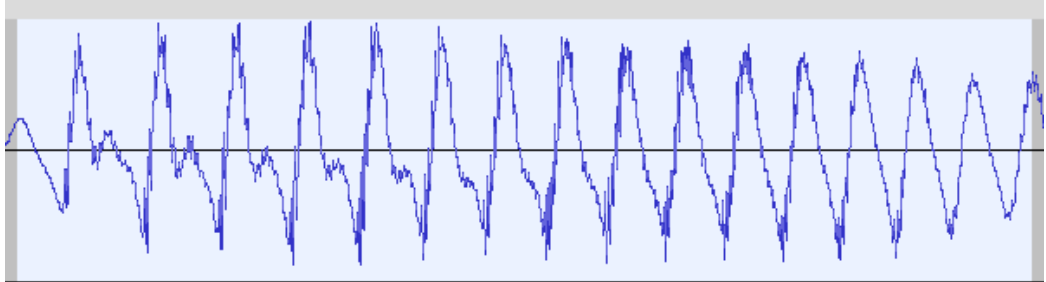


Conté els fonemes in que sonen pràcticament junts.

Si observem l'ona mentre reproduïm l'àudio, notarem que cadascuna d'aquestes seccions representa un fonema. Per exemple:



Aquesta part de l'ona, representa el fonema i de la síl·laba sis de la paraula anàlisis. I es pot observar que la seva forma és molt diferent a la:



Aquesta part representa el fonema i de la paraula fin.

Apartat b

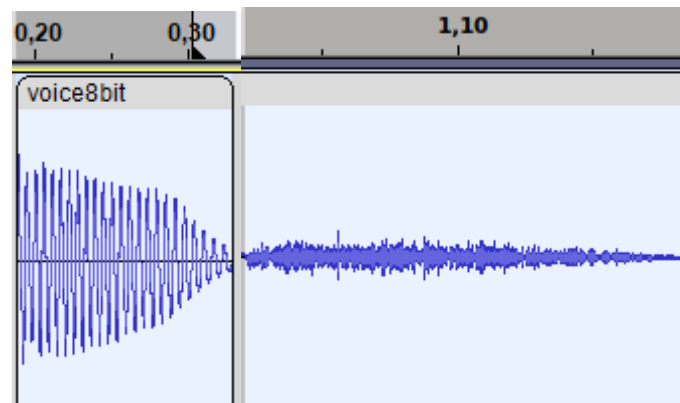
Fonemes sonors

A la pista en cru que hem d'analitzar, podem escoltar que hi ha els següents fonemes sonors:

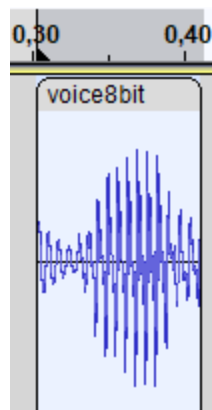
/n/, /d/, /l/

Analitzant-los, podem observar que cada fonema té les següents formes d'ona:

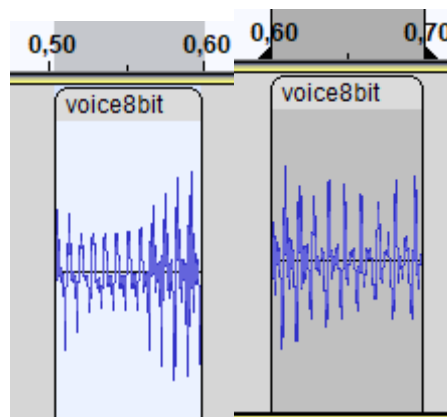
/n/



/d/



/l/

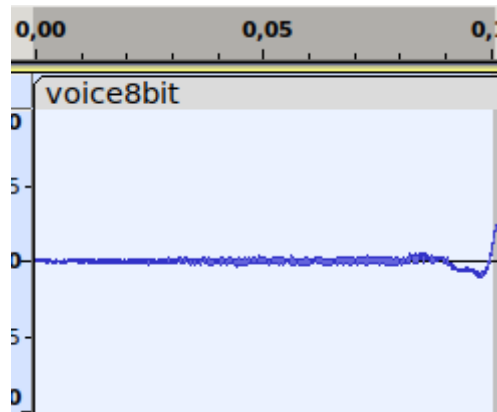


Fonemes sords

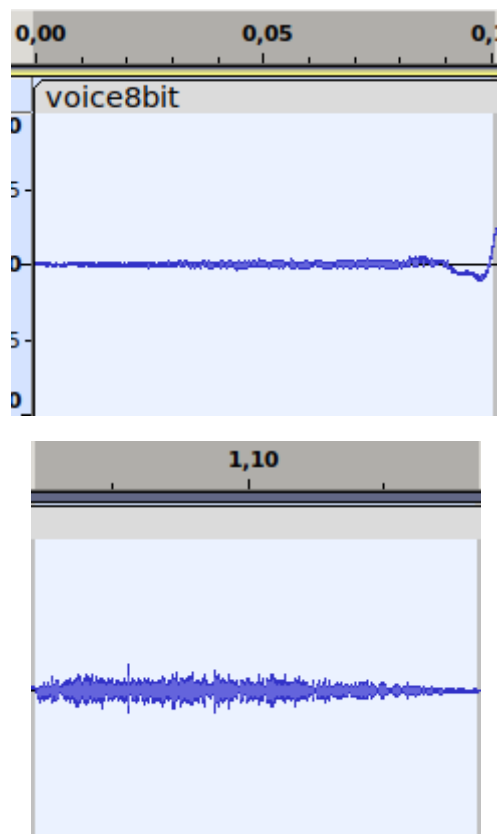
A la pista en cru que hem d'analitzar, podem escoltar que hi ha els següents fonemes sords: /f/ i el fonema /s/

Analitzant-los, podem observar que tenen les següents parts:

/f/



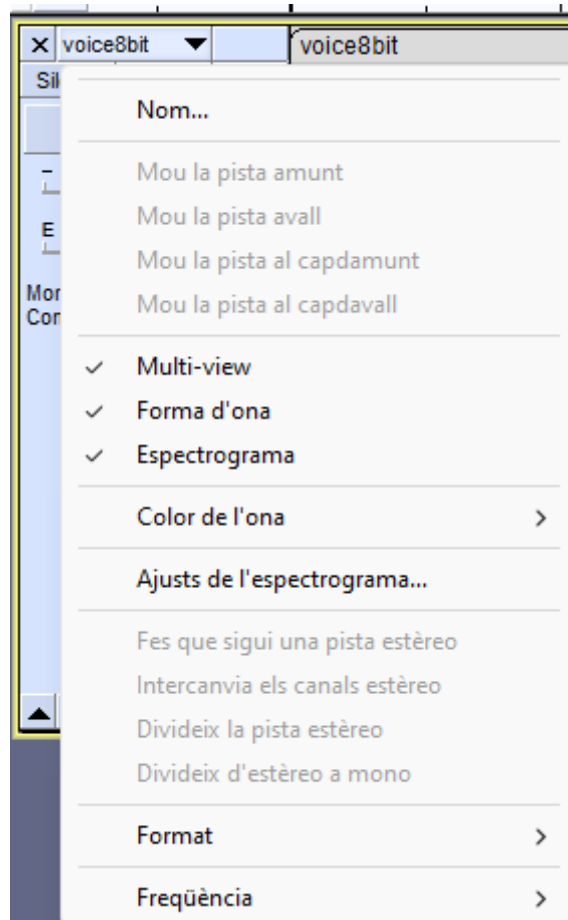
/s/



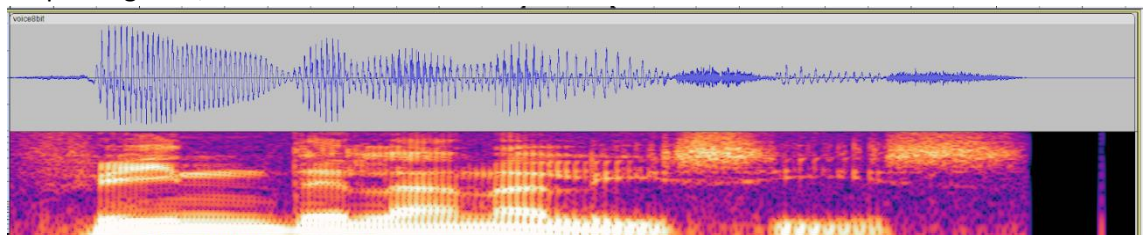
Com es pot observar, cada fonema ocupa aproximadament 1 mil·lisegon.

Ara procedim a analitzar l'espectrograma d'aquesta pista, per a veure'l, només cal seguir aquests passos:

1. Clica sobre el nom de la pista al menú de l'esquerra:



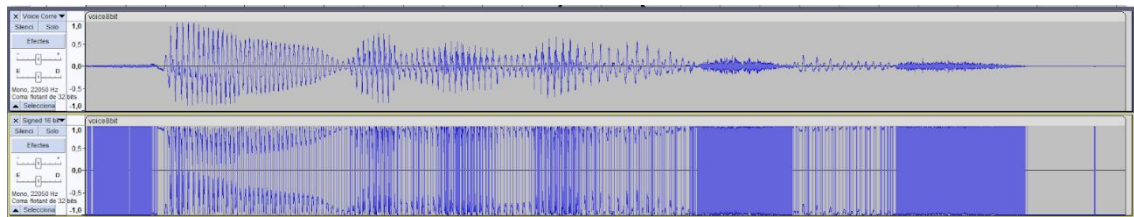
2. Ara seleccionem Espectrograma o bé si volem fer la comparativa entre la forma d'ona i l'espectrograma, seleccionem Multi-view



Es pot veure clarament en el espectrograma que quan l'ona és més gran, el color és més clar, semblant al blanc. Això correspon a zones on hi ha soroll, tal com es veu en la forma de l'ona. A mida que la longitud d'ona disminueix, el color es torna més blau, indicant menys soroll. A més, es noten dos punts negres en l'espectrograma, que assenyalen silenci.

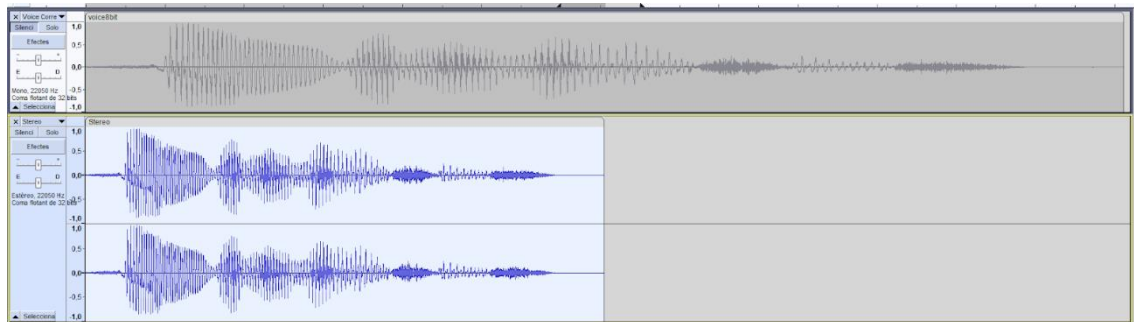
Apartat c

Signed 16 bit, Little-endian, 1 canal, 22050 Hz



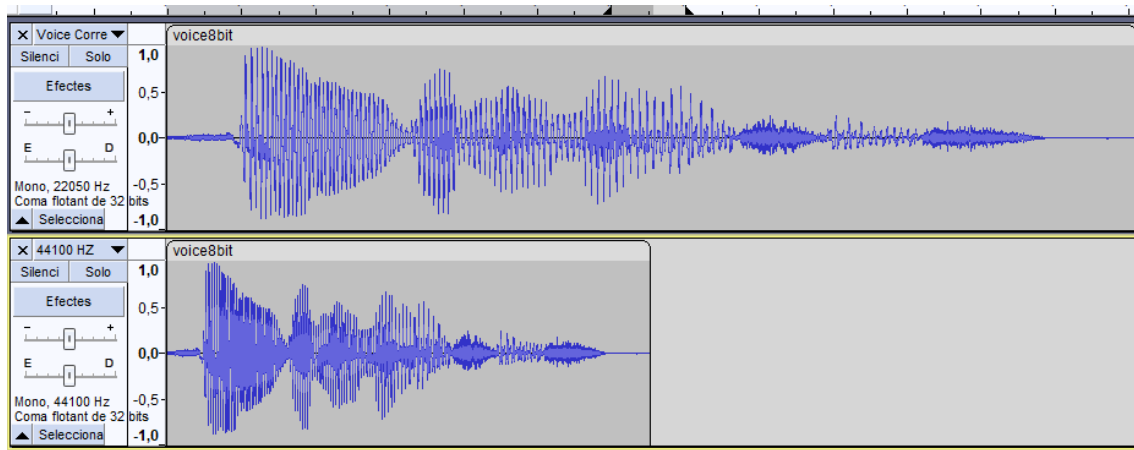
Com es pot observar la forma d'ona ha canviat dràsticament, això es deu a que amb l'opció signed, permet tenir mostres negatives, però també fa que la qualitat del so sigui molt inferior i que costi molt diferenciar l'àudio de la pista.

Unsigned 8 bit, Little-endian, 2 canals (Stereo), 22050 Hz



En aquest cas, la pista 1 (mono) dura més perquè només té un canal, mentre que la pista 2 (estèreo) és més curta perquè conté dues versions de la mateixa informació. Això també provoca que la veu soni més aguda i accelerada.

Unsigned 8 bit, Little-endian, 1 canal, 44100 Hz



Es pot veure que en aquest cas, la mostra amb una freqüència de mostreig més alta, dura menys temps. Tot i que hi ha la mateixa quantitat de mostres en ambdós casos, en el cas de 44100 Hz, les mostres es prenen molt més ràpidament, cosa que fa que la forma d'ona sigui igual que l'original, però més compacta. Això també provoca que la veu soni més aguda i accelerada.

Exercici 2 – Codificant amb FLAC i vocoders

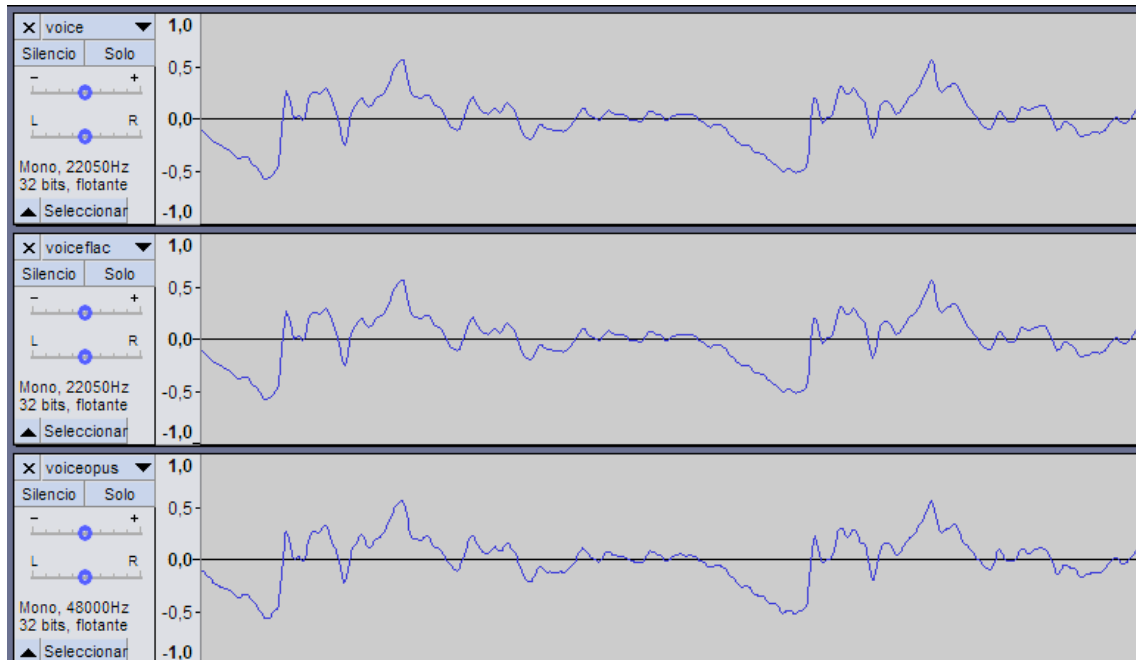
Apartat a

| Codec | Fichero WAV | Frec. muestreo (Khz) | Mono / Stereo | Bits / muestra | Bitrate (Kbits/s) | Tamaño (Kbytes) | Factor de compresión | Calidad subjetiva |
|--------------------------------------------------------------------------------------------------------------------------------------|-------------|----------------------|---------------|----------------|-------------------|-----------------|----------------------|-------------------|
| Análisis de ficheros de voz y música de calidad fullband (16-22 Khz) | | | | | | | | |
| Original WAV (PCM) – Calidad CD | voice | 22,05 | Mono | 16 | 352,8 | 56,5 | 1 | 4.33 |
| | music | 44,1 | Estéreo | 16 | 1411,2 | 11400 | 1 | 4.33 |
| FLAC | voice | 22,05 | Mono | N/A | 71,1 | 11,4 | 4,96 | 4 |
| | music | 44,1 | Estéreo | N/A | 841 | 6810 | 1,67 | 4.33 |
| Opus | voice | 24 | Mono | N/A | 77,2 | 12,4 | 4,56 | 4 |
| | music | 48 | Estéreo | N/A | 64,6 | 5352 | 2,13 | 4.33 |
| Análisis de ficheros de voz y música de calidad narrowband (calidad telefónica, 4 KHz). Notad que se usan los ficheros –tel . | | | | | | | | |
| Original WAV (PCM) – calidad telefónica | voice-tel | 8 | Mono | 8 | 64 | 10,3 | 1 | 1.67 |
| | music-tel | 8 | Mono | 8 | 64 | 531 | 1 | 2 |
| ADPCM a 32 Kbit/s (G.721) | voice-tel | 8 | Mono | N/A | 32 | 5,16 | 2 | 2 |
| | music-tel | 8 | Mono | N/A | 32 | 265 | 2 | 1.67 |
| AMR-NB al menor bitrate (4.75 Kbit/s) GSM 6.10 | voice-tel | 8 | Mono | N/A | 13,1 | 2,09 | 4,93 | 1.33 |
| | music-tel | 8 | Mono | N/A | 13 | 108 | 4,92 | 1.67 |
| Opus | voice-tel | 8 | Mono | N/A | 78 | 12,5 | 0,82 | 2.33 |
| | music-tel | 8 | Mono | N/A | 64 | 530 | 1 | 1.67 |

Les codificacions i descodificacions s'han realitzat amb els valors predeterminats de les opcions d'exportació per a cada còdec a Audacity. En el cas del còdec AMR-NB, com que no era disponible com a format d'exportació, s'ha optat per utilitzar el còdec GSM 6.10. Aquest còdec presenta característiques similars al còdec AMR-NB en el seu bitrate més baix (4,75 kbit/s).

Apartat b

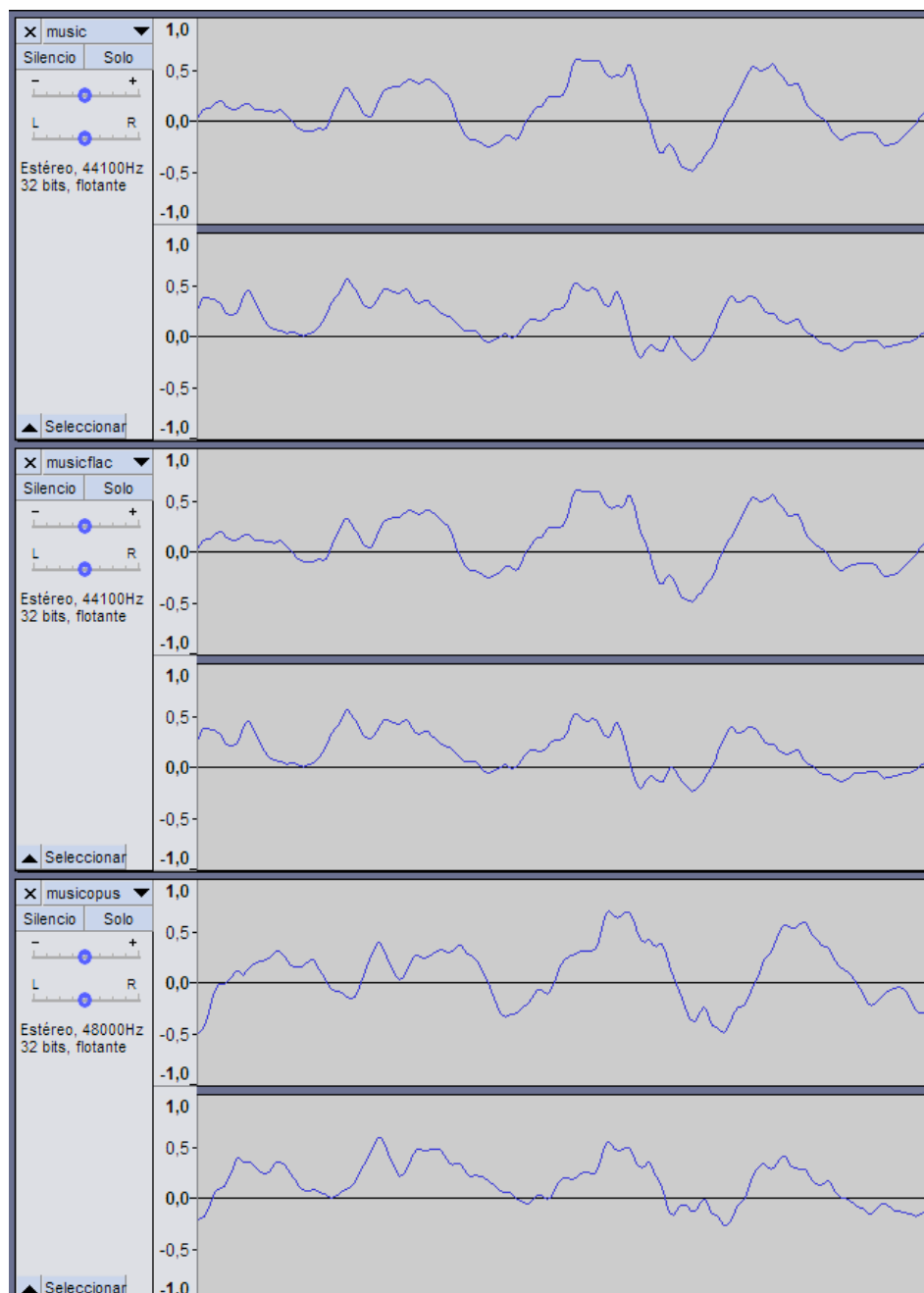
Voice



A la part superior, es mostra la forma d'ona del fitxer original voice.wav. Al centre, es pot veure la forma d'ona del fitxer voice.wav després de la compressió FLAC. A la part inferior, es troba la forma d'ona del fitxer voice.wav després de la compressió Opus.

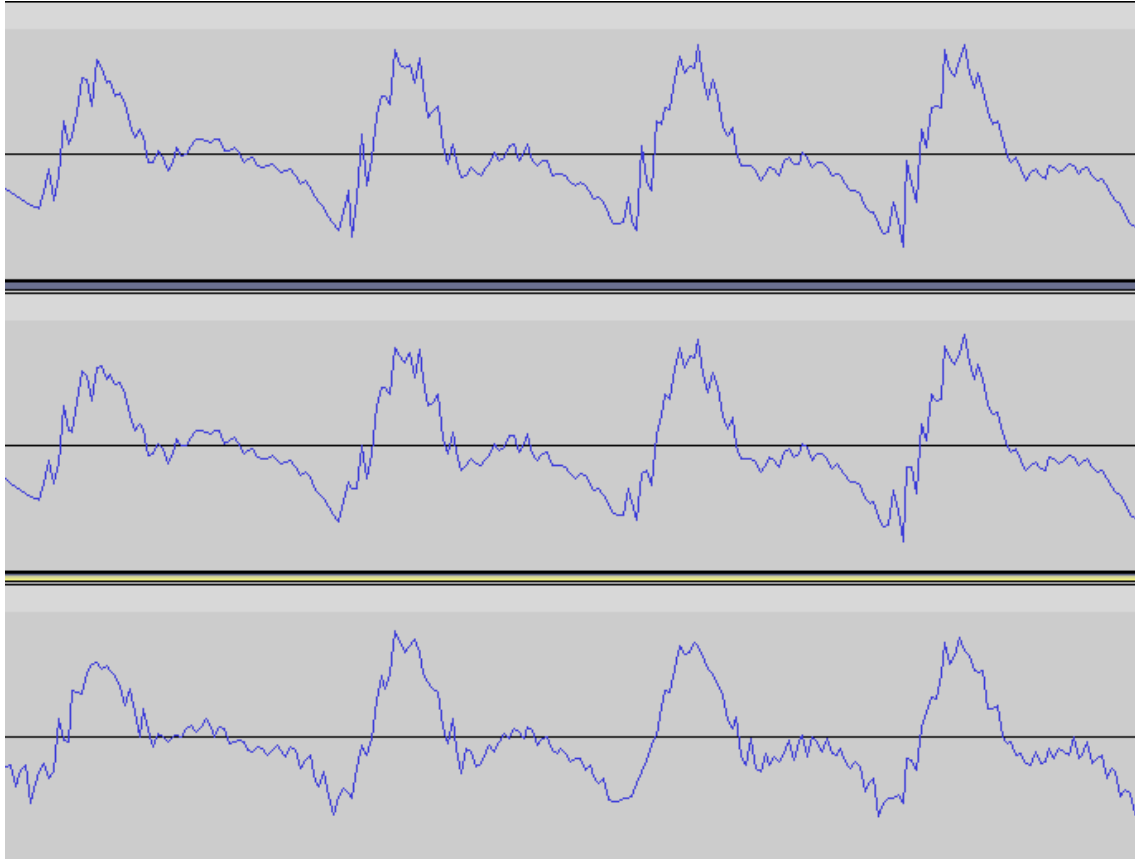
Com podem observar, la forma d'ona del fitxer original voice.wav i la del fitxer comprimit amb FLAC, que és un códec sense pèrdues, presenten un nivell de detall més alt. Això es deu al fet que FLAC conserva tota la informació original del so, sense perdre cap detall durant el procés de compressió. En canvi, la forma d'ona dels fitxers comprimits amb Opus, que és un códec amb pèrdues, mostra una petita diferència. Això ocorre perquè Opus optimitza l'espai de l'emmagatzematge reduint la mida del fitxer, però aquesta compressió comporta una pèrdua de certa informació sonora, la qual cosa afecta la qualitat i el detall de la forma d'ona visualitzada.

Music



A la part superior, es mostra la forma d'ona del fitxer original music.wav. Al centre, es pot veure la forma d'ona del fitxer music.wav després de la compressió FLAC. A la part inferior, es troba la forma d'ona del fitxer music.wav després de la compressió Opus.

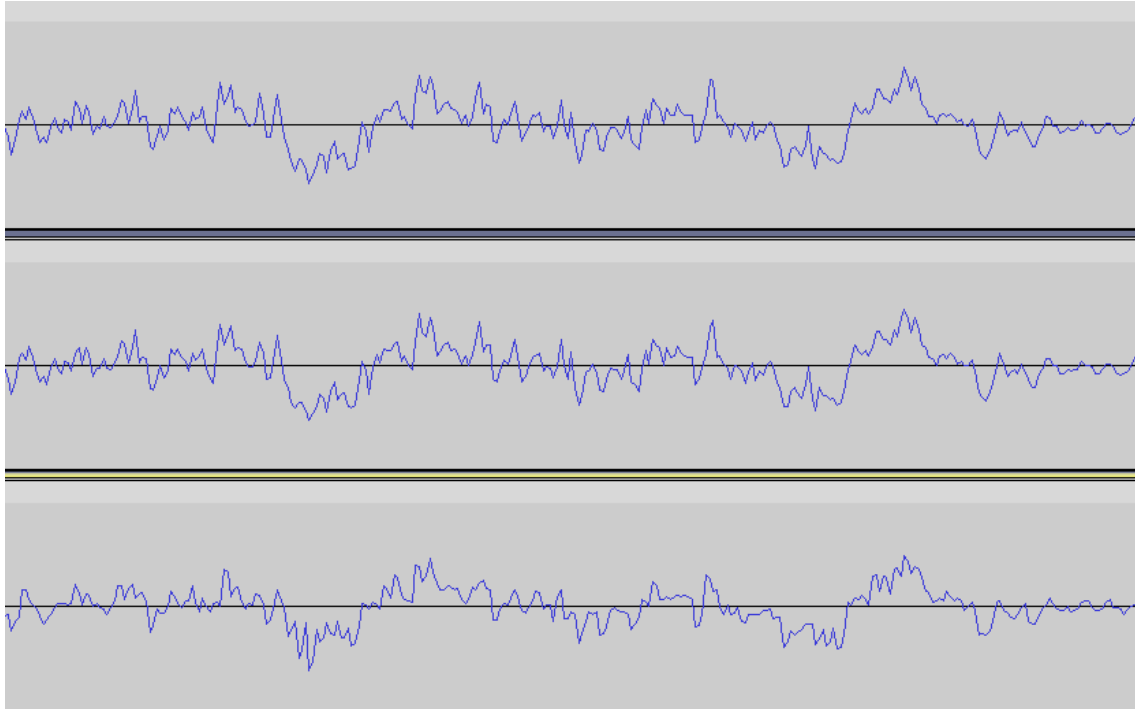
En aquest cas, observem un patró similar al que hem vist abans. La forma d'ona del fitxer original music.wav i la del fitxer comprimit amb FLAC, un codec sense pèrdues, exhibeixen un nivell de detall molt superior. FLAC manté tota la informació original sense cap pèrdua durant la compressió. En contrast, els fitxers comprimits amb Opus, un codec amb pèrdues, mostren una diferència notable. Opus busca optimitzar l'espai de l'emmagatzematge reduint la grandària del fitxer, però a costa de perdre certa informació sonora. Això afecta la qualitat i el detall de la forma d'ona visualitzada, esperant-se una certa pèrdua especialment en passatges complexos amb molts instruments o vocals superposats.



A la part superior, es mostra la forma d'ona del fitxer original voice-tel.wav. Al centre, es pot veure la forma d'ona del fitxer voice-tel.wav després de la compressió ADPCM. A la part inferior, es troba la forma d'ona del fitxer voice-tel.wav després de la compressió GSM.

Com podem observar, la forma d'ona del fitxer original voice-tel.wav presenta un nivell de detall molt més alt que les formes d'ona dels fitxers comprimits amb ADPCM i GSM. Això es deu a la naturalesa de la compressió d'àudio, la qual implica una pèrdua d'informació per tal de reduir la mida del fitxer.

En el cas dels formats ADPCM i GSM, la pèrdua d'informació se centra principalment en les freqüències altes. La raó d'això és que l'oïda humana és menys sensible a aquestes freqüències, per la qual cosa els compressors les eliminen o les representen amb menys precisió per aconseguir una major reducció de la mida del fitxer.



A la part superior, es mostra la forma d'ona del fitxer original music-tel.wav. Al centre, es pot veure la forma d'ona del fitxer music-tel.wav després de la compressió ADPCM. A la part inferior, es troba la forma d'ona del fitxer music-tel.wav després de la compressió GSM.

Com s'ha assenyalat anteriorment, la compressió d'àudio mitjançant tècniques com ADPCM i GSM comporta una inevitable pèrdua d'informació, la qual es concentra principalment en les freqüències altes. Aquesta pèrdua d'informació es manifesta de manera més accentuada en fitxers de música en comparació amb fitxers de veu.

La causa d'aquesta diferència radica en la naturalesa dels vocoders, els algorismes emprats en ADPCM i GSM. Els vocoders estan dissenyats específicament per a la codificació de fonemes, les unitats bàsiques del so en la parla humana. En el context de la parla humana, la pèrdua de detall en les freqüències altes no és tan crítica, ja que aquestes freqüències no són tan rellevants per a la intel·ligibilitat de la veu.

En canvi, la música conté un espectre de freqüències molt més ampli, que inclou harmonies, melodies i timbres. La pèrdua de detall en les freqüències altes en aquest tipus d'àudio pot afectar significativament la qualitat percebuda, resultant en una reproducció menys fidedigna i rica.

Apartat c

Per a l'avaluació de la qualitat subjectiva dels fitxers de so, s'ha optat per una prova d'escolta a cegues i aleatòria. En aquesta prova, els membres del grup van escoltar els fitxers de so en un ordre aleatori sense saber quina codificació s'havia aplicat a cada un. Després de cada escolta, van omplir una enquesta de qualitat, assignant una puntuació entre 1 i 5 a cada fitxer. La puntuació mitjana de cada fitxer s'ha calculat per a obtenir la qualitat subjectiva mitjana.

| Codec | Fichero WAV | Qualitat Subjectiva Ernest | Qualitat Subjectiva Naim | Qualitat Subjectiva Mariona | Qualitat Subjectiva Adrián | Qualitat Subjectiva Mitja |
|--------------------------------------------------------------------------------------------------------------------------------------|-------------|----------------------------|--------------------------|-----------------------------|----------------------------|---------------------------|
| Análisis de ficheros de voz y música de calidad fullband (16-22 Khz) | | | | | | |
| Original WAV (PCM) – Calidad CD | voice | 5 | 3 | 4 | 5 | 4.25 |
| | music | 4 | 4 | 4 | 5 | 4.25 |
| FLAC | voice | 4 | 3 | 4 | 5 | 4 |
| | music | 4 | 4 | 5 | 5 | 4.5 |
| Opus | voice | 4 | 3 | 4 | 5 | 4 |
| | music | 5 | 3 | 5 | 5 | 4.5 |
| Análisis de ficheros de voz y música de calidad narrowband (calidad telefónica, 4 KHz). Notad que se usan los ficheros –tel . | | | | | | |
| Original WAV (PCM) – calidad telefónica | voice-tel | 1 | 2 | 2 | 2 | 1.75 |
| | music-tel | 2 | 2 | 1 | 2 | 1.75 |
| ADPCM a 32 Kbit/s (G.721) | voice-tel | 3 | 1 | 2 | 2 | 2 |
| | music-tel | 2 | 1 | 2 | 2 | 1.75 |
| AMR-NB al menor bitrate (4.75 Kbit/s) GSM 6.10 | voice-tel | 2 | 1 | 2 | 1 | 1.5 |
| | music-tel | 1 | 2 | 1 | 1 | 1.25 |
| Opus | voice-tel | 3 | 2 | 3 | 2 | 2.5 |
| | music-tel | 1 | 2 | 2 | 2 | 1.75 |

En el cas dels vocoders (ADPCM i GSM), s'observa una millor qualitat subjectiva en el fitxer de veu en comparació amb el fitxer de música. Això es deu al fet que els vocoders estan dissenyats específicament per a codificar fonemes, les unitats bàsiques del so en la veu humana. En aquest context, la pèrdua de detall en les freqüències altes no és tan crítica, ja que aquestes freqüències no són tan rellevants per a la intel·ligibilitat del parla. En canvi, per al fitxer de música, els vocoders poden introduir una pèrdua de fidelitat notable, especialment en les melodies, harmonies i timbres. Això es deu al fet que la música conté un espectre de freqüències molt més ampli que la veu, i la pèrdua de detall en les freqüències altes pot afectar negativament la riquesa i la nitidesa del so.

Cal destacar que, tot i així, els altres codificadors (WAV, FLAC i Opus) semblen proporcionar una bona qualitat subjectiva tant en el fitxer de música com en el de veu. Això és a causa que el codificador FLAC no introdueix cap pèrdua respecte al fitxer original WAV. A més, tot i que Opus introdueix pèrdues, no s'ha observat cap afectació en la qualitat subjectiva.

Apartat d

No s'ha trobat cap inconsistència en les dades i conclusions obtingudes en els apartats anteriors respecte als obtinguts en els estudis previs.

Apartat e

Còdecs sense pèrdua:

- WAV original (PCM): Tant en qualitat de CD com telefònica, aquest format no introdueix pèrdues. Manté la qualitat original del senyal d'àudio, però sense compressió, resultant en arxius de gran grandària.
- FLAC: Un còdec sense pèrdua que comprimeix l'àudio sense sacrificar la qualitat. Els arxius FLAC són generalment la meitat de la mida que els WAV sense comprimir.

Còdecs amb pèrdua:

- Opus: Un còdec versàtil que ofereix una compressió molt eficient mantenint pràcticament la qualitat original.
- ADPCM a 32 Kbit/s (G.721): Aquest còdec introdueix pèrdues significatives per aconseguir una alta compressió. S'utilitza comunament en telefonia, on la intel·ligibilitat de la parla és prioritària sobre la qualitat de la música.
- AMR-NB al bitrate més baix (4.75 Kbit/s): Similar a l'ADPCM G.721, aquest còdec amb pèrdua comprimeix encara més, sacrificant la qualitat d'àudio per obtenir una mida d'arxiu extremadament petit. S'utilitza en trucades de baixa qualitat o on l'amplada de banda és molt limitada.

Els còdecs sense pèrdua com WAV original (PCM) i FLAC mantenen la qualitat original de l'àudio, però varien en termes de factor de compressió. FLAC pot comprimir arxius de veu 5 vegades i de música 1.67 vegades respecte al WAV original (PCM), però manté la mateixa qualitat. Opus, d'altra banda, ofereix una compressió molt eficient en el seu mode sense pèrdua, però també manté la qualitat original de l'àudio.

D'altra banda, els còdecs amb pèrdua com Opus, ADPCM a 32 Kbit/s (G.721) i GSM 6.10, comprimeixen l'àudio a costa de la qualitat, malgrat que amb l'Opus la qualitat subjectiva no es veu afectada. Per exemple, el còdec ADPCM a 32 Kbit/s (G.721) introdueix pèrdues per aconseguir una compressió de 2 vegades respecte al WAV original (PCM). De manera similar, GSM 6.10 sacrifica la qualitat d'àudio empitjorant la qualitat subjectiva per obtenir una mida d'arxiu extremadament petita amb una compressió de 5 vegades respecte al WAV original (PCM).

Els còdecs sense pèrdua ofereixen una qualitat d'àudio excel·lent utilitzant algorismes de compressió molt eficients. Per exemple, el còdec FLAC té un factor de compressió de 4.96 per a la veu i 1.67 per a la música, però manté una alta qualitat subjectiva de 4 per a la veu i 4.33 per a la música. Això significa que, tot i la compressió, la qualitat de l'àudio es manté intacta. D'altra banda, els còdecs amb pèrdua poden oferir una qualitat d'àudio útil o deficient, però amb un factor de compressió més gran. Per exemple, el còdec GSM 6.10 té un factor de compressió de 5 per a la veu i 1.67 per a la música, però la qualitat subjectiva és de 4 per a la veu i 4.33 per a la música. Això indica que, tot i que la mida de l'arxiu es redueix significativament, la qualitat de l'àudio es veu afectada. Per tant, la elecció del còdec depèn de l'equilibri desitjat entre la qualitat de l'àudio i la mida de l'arxiu. Si es prioritza la qualitat de l'àudio, es poden preferir els còdecs sense pèrdua. Si es prioritza la reducció de la mida de l'arxiu, es poden preferir els còdecs amb pèrdua.

Apartat f

En vocoders i compressors perceptuals, el concepte de bits per mostra no s'aplica, ja que aquests mètodes de processament d'àudio es basen en la percepció humana del so en comptes de la representació digital precisa del senyal d'àudio.

Els vocoders analitzen el senyal d'àudio en diferents bandes de freqüència i després sintetitzen un nou senyal d'àudio utilitzant només un nombre limitat de bits per banda de freqüència. Això significa que la quantitat total de bits utilitzats per representar el senyal d'àudio complet és menor que la quantitat de bits que s'utilitzarien en una representació digital sense comprimir. No obstant això, la qualitat del so resultant encara pot ser alta, ja que el vocoder conserva les característiques més importants del so que són percebudes per l'oïda humana.

Els compressors perceptuals ajusten el nivell de volum del senyal d'àudio en funció de la sensibilitat de l'oïda humana a diferents freqüències. Això significa que les parts del senyal d'àudio que són menys perceptibles per a l'oïda humana es comprimeixen més que les parts que són més perceptibles. Això permet reduir la quantitat total de bits utilitzats per representar el senyal d'àudio sense afectar significativament la qualitat del so percebuda.

En resum, el concepte de bits per mostra no s'aplica en vocoders i compressors perceptuals perquè aquests mètodes de processament d'àudio es basen en la percepció humana del so en comptes de la representació digital precisa del senyal d'àudio. Això permet reduir la quantitat total de bits utilitzats per representar el senyal d'àudio sense afectar significativament la qualitat del so percebuda.

Apartat g

Nivells de configuració de FLAC

La quantitat de compressió de la mida de l'arxiu utilitzada per empaquetar les dades d'àudio sense pèrdues, similar als nivells de compressió en els arxius ZIP. La configuració del nivell varia de 0 a 8.

- Nivell 0: està optimitzat per codificar el més ràpid possible.
- Nivell 8: està optimitzat per empaquetar de la manera més eficient possible, produint arxius lleugerament més petits que el nivell 0 a costa de tardar més en codificar.

Els arxius codificats en el nivell 8 rarament són més d'un petit percentatge més petits que els codificats en el nivell 0, però poden tardar tres o quatre vegades més en codificar-se.

Diferent bitrate per AMR

Com que el còdec AMR-NB no estava disponible com a format d'exportació, no s'ha pogut comparar la qualitat entre els dos extrems, és a dir, entre la qualitat més alta i la més baixa.

Exercici 3 – Compriment àudio amb MPEG

Mida del fitxer original music.wav = 12.0MB

| Codec | Bitrate(Kbit/s) | Libreria | Tamaño del fichero comprimido (Mbytes) | Tamaño de cada trama (bytes) | Factor de compresión | Calidad subjetiva MOS |
|-------|-----------------|----------|----------------------------------------|------------------------------|----------------------|-----------------------|
| MP2 | 32 | Audacity | 0.26 | 104 | 46.15 | 1 |
| MP2 | 128 | Audacity | 1.1 | 416 | 10.9 | 4 |
| MP3 | 32 | Audacity | 0.26 | 104 | 49.6 | 1 |
| MP3 | 32 | FFmpeg | 0.26 | 104 | 46.15 | 1 |
| MP3 | 128 | Audacity | 1.1 | 416 | 10.9 | 4 |
| MP3 | 128 | FFmpeg | 1.1 | 416 | 10.9 | 4 |

Apartat a

S'observa una millora substancial en la qualitat de l'àudio en augmentar el bitrate de 32 kbps a 128 kbps. Aquesta diferència és especialment notable en la nitidesa i la definició del so. Tant FFmpeg com Audacity ofereixen resultats de compressió molt semblants, pràcticament idèntics.

Pel que fa a la mida del fitxer, s'observen petites desviacions en comparació dels càlculs de l'exercici previ, però no són significatives.

Apartat b

A partir de les proves realitzades, es conclou que la qualitat de l'àudio depèn principalment del bitrate i no tant del layer utilitzat. Les diferències entre mp2 i mp3 no són tan notables com les que es produeixen en canviar el bitrate, per exemple, de 32 kbps a 128 kbps.

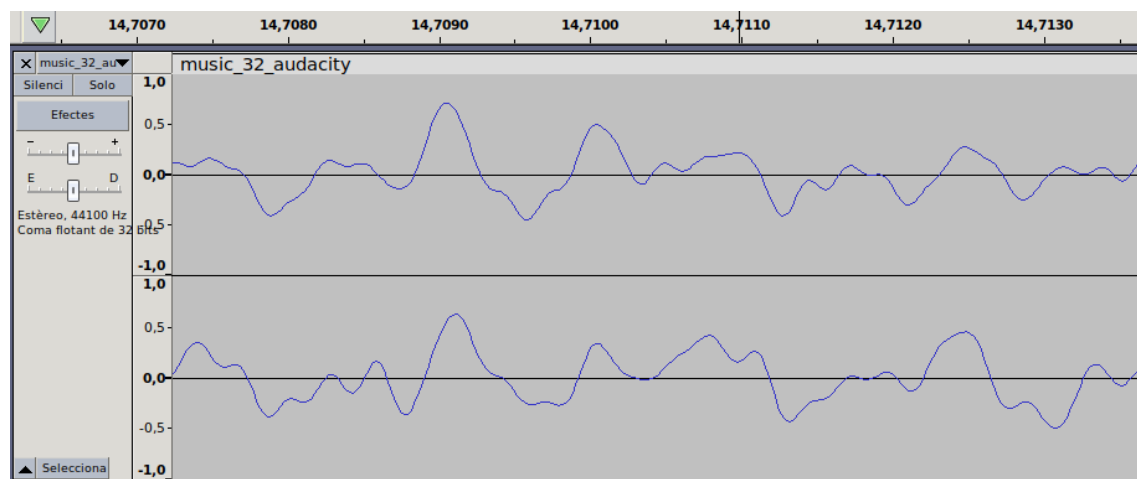
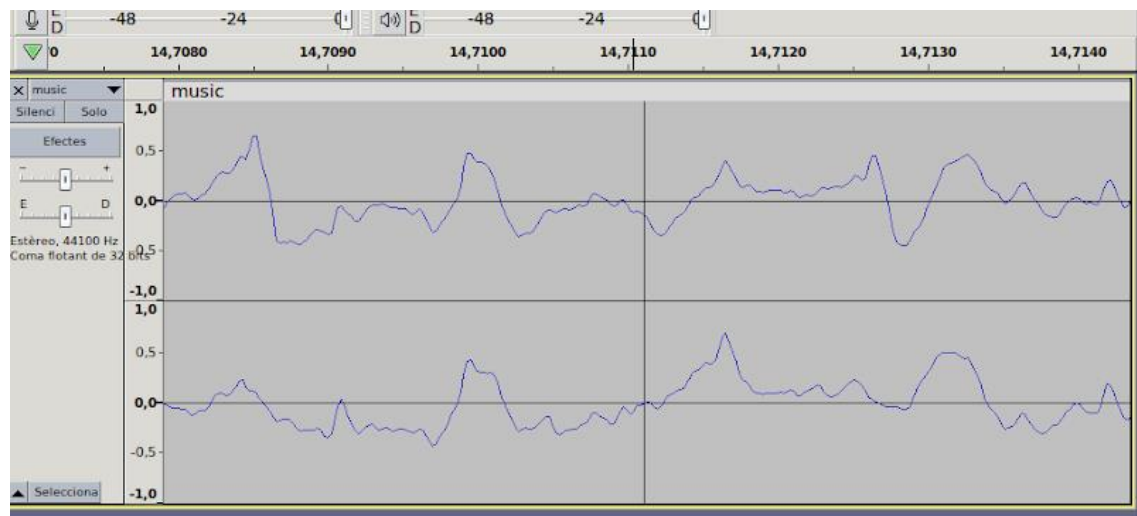
Apartat c

No s'han trobat diferències significatives en la qualitat de l'àudio en comprimir fitxers amb FFmpeg o Audacity. Els resultats obtinguts amb tots dos programes són molt similars, tant en termes de dades com en la percepció auditiva.

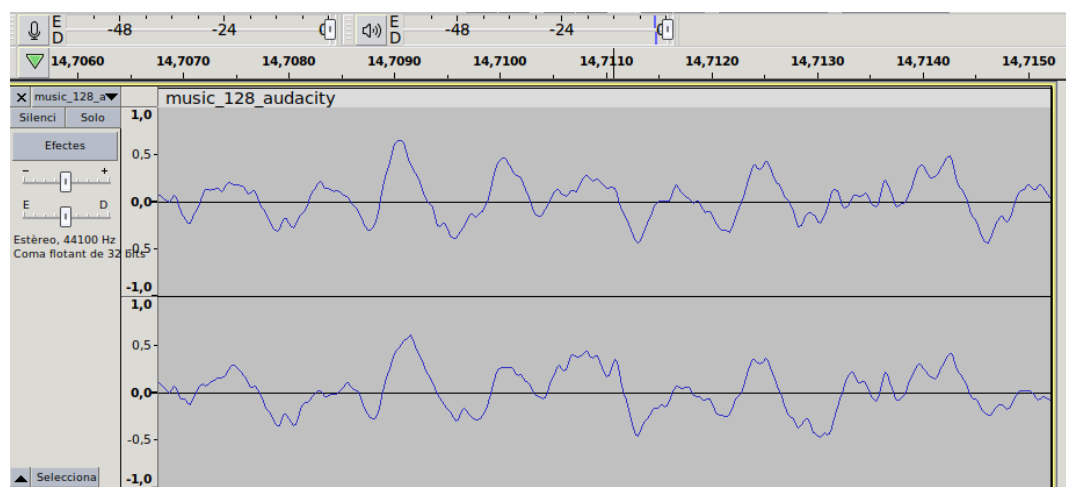
Les petites diferències que hi podrien haver s'atribueixen als diferents algorismes de compressió utilitzats per cada programa.

Exercici 4 – Efectes de MPEG sobre la forma d'ona

.wav



128 kbps mp3



En augmentar el zoom i comparar els formats WAV i MP3 amb diferents bitrats (32 kbps i 128 kbps), s'observen canvis significatius en la representació de les ones sonores:

MP3

32 kbps: Les ones es veuen encara més arrodonides i amb una definició encara menor que a WAV. La compressió provoca una pèrdua considerable de detall, fent que les ones siguin quasi uniformes.

128 kbps: Les ones presenten més definició que en 32 kbps, encara que segueixen sent menys nítides que a WAV. S'aprecien millor els canvis de pujades i baixades, indicant una presència més gran de detall.

Conclusió

El bitrate té un impacte significatiu en la representació visual de les ones sonores.

A més bitrate, les ones es mostren amb més definició i detall, tant en format WAV com MP3.

A WAV, la diferència de qualitat entre 32 kbps i 128 kbps és menys pronunciada que en MP3, ja que WAV conserva més informació original fins i tot amb compressió.

A MP3, la diferència de qualitat entre 32 kbps i 128 kbps és més evident, ja que la compressió en aquest format afecta més la definició de les ones.

Recomanacions

Per a més fidelitat i detall de l'àudio, es recomana utilitzar formats sense comprimir com WAV o FLAC, especialment en aplicacions professionals o per a fitxers d'alta qualitat.

Si l'objectiu és reduir la mida del fitxer i la portabilitat, el format MP3 amb un bitrate de 128 kbps o superior ofereix un bon equilibri entre qualitat i compressió, i és adequat per a la majoria d'usuaris.

És important considerar el tipus de contingut (música, veu, etc.) i l'ús previst (reproducció personal, streaming, etc.) en triar el format i el bitrate adequats.

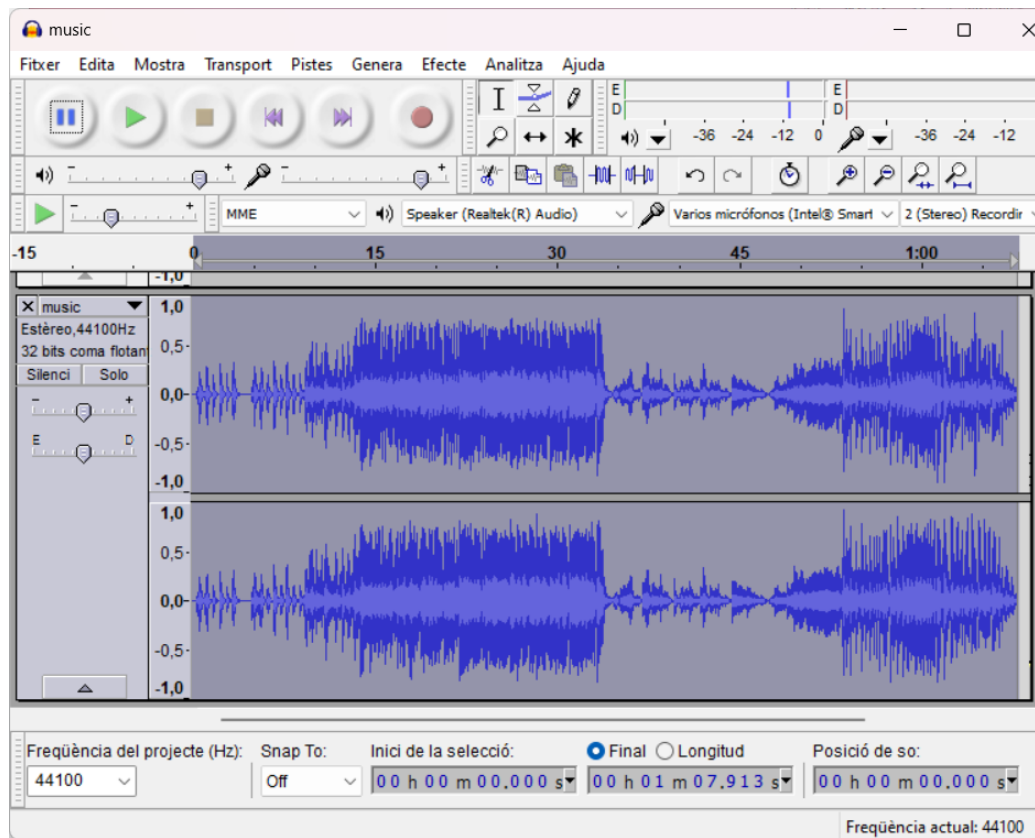
Observacions addicionals

La percepció de la qualitat de l'àudio és subjectiva i depèn de diversos factors, com ara l'equip de reproducció, l'experiència auditiva de l'usuari i el tipus de contingut.

És recomanable fer proves auditives amb diferents formats i bitrate per tal de determinar l'opció que millor s'adapta a les preferències individuals.

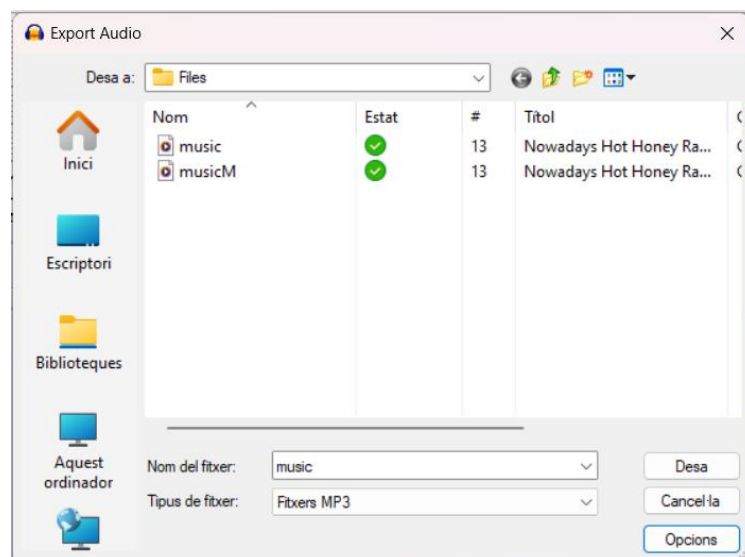
Exercici 5 – MP3 VBR

Entrar en Audacity i obrir el fitxer musica.wav donat pel laboratori.



Apartat a

Per veure les diferents opcions, anar a Fitxer -> Export Audio , per modificar-les entrar en les -> opcions de L'Export:



Amb les següents opcions a especificar:

Els modes de resolució:

- Configuració: Permet personalitzar les característiques del bitrate (la taxa de bits) segons les necessitats específiques d'un usuari. Es pot triar entre diferents valors de bitrate dependents de la qualitat desitjada.
- Variable: En el mode de bitrate variable (VBR), la compressió s'ajusta dinàmicament depenent de la complexitat de l'àudio en cada moment. Això pot millorar la qualitat general sense incrementar excessivament la mida de l'arxiu. En aquest mode, s'ha de seleccionar una qualitat de VBR, que va de 0 (millor qualitat, arxius més grans) a 9 (pitjor qualitat, arxius més petits).
- Mitjana: El bitrate mitjà (ABR) és un entremig entre el mode constant i variable. Estableix un bitrate mitjà, però permet certes variacions per mantenir una qualitat més consistent.
- Constant: En el mode de bitrate constant (CBR), el bitrate es manté igual durant tota la durada de l'arxiu MP3. Que facilita el càlcul de la mida de l'arxiu final, però pot resultar en una qualitat inferior quan es tracta de segments d'àudio complexos.

Qualitat:

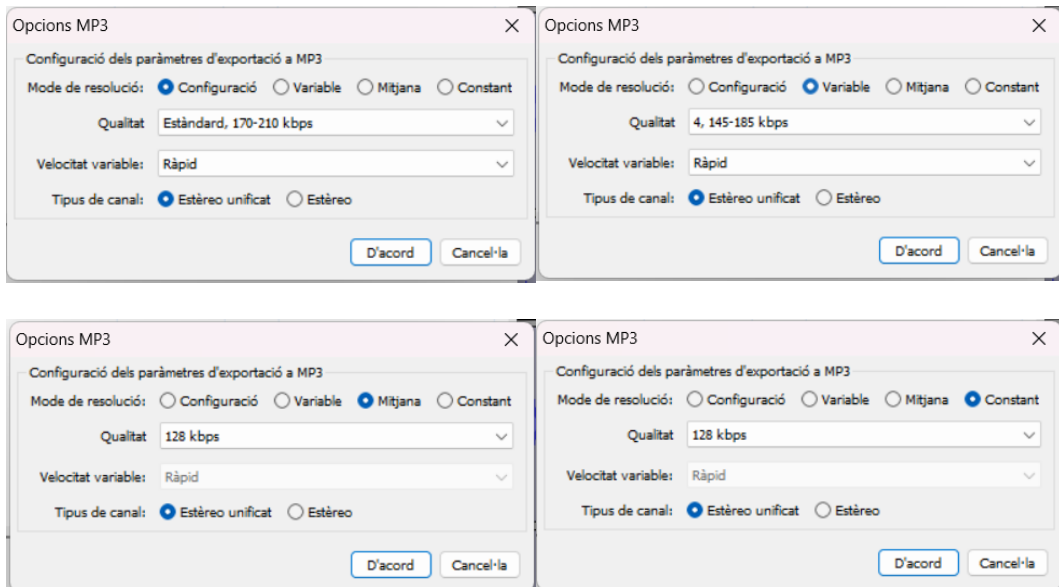
- Mode Configuració: En aquest mode, tens control total sobre el bitrate, el que significa que es pot ajustar completament segons les preferències del usuari.
- Una regla general és que a mesura que augmentes el bitrate, millora la qualitat de l'àudio. Un bitrate més alt pot capturar millor les nuances de la música o la veu, especialment en passatges complicats, però un bitrate més baix pot conduir a una pèrdua de qualitat perceptible, especialment en passatges amb molta informació.
- Mode Variable: En aquest mode, s'ajusta dinàmicament el bitrate en funció de la complexitat de l'àudio. La qualitat en el mode VBR pot ser superior per al mateix bitrate mitjà comparat amb el mode de resolució constant.
- Els ajustos de qualitat (d'entre el 0 a 9) et permeten triar entre una millor qualitat i arxius més grans o una qualitat inferior i arxius més petits.
- Mode Mitjana: En aquest mode, es selecciona un bitrate mitjà, i Audacity ajusta el bitrate dinàmicament per mantenir aquesta mitjana.
- Pot proporcionar una qualitat més consistent en tot l'arxiu en comparació amb el VBR, ja que no hi ha fluctuacions tan grans en el bitrate.
- Mode Constant: En aquest mode, es selecciona un bitrate i es manté constant en tot l'arxiu. Això proporciona una qualitat més previsible, ja que tot l'arxiu utilitza el mateix bitrate, independentment de la complexitat de l'àudio. Però pot ser menys eficient en el cas d'àudio amb zones amb diferent complexitat, ja que no es pot ajustar dinàmicament.

Velocitat Variable:

En els modes de resolució variable(VBR) i mitjana(ABR), es pot seleccionar la velocitat de l'encoder a "Ràpid" o "Estàndar". La velocitat "Ràpid" prioritza una major velocitat de codificació, mentre que "Estàndar" pot millorar la qualitat a costa de prendre més temps. Aquesta opció no està disponible en el mode constant (CBR) perquè el bitrate és fix i no necessita ajustaments dinàmics en la velocitat.

Tipus de Canal:

- Estéreo Unificat: es processen els canals esquerre i dret junts, mantenint la relació entre ells durant la compressió. Això pot ser útil per a mantenir la imatge estèreo correcta.
- Estéreo: els canals esquerre i dret es codifiquen de manera independent. Això pot ser útil si els canals tenen diferències significatives i vols maximitzar la qualitat de cada canal individualment.

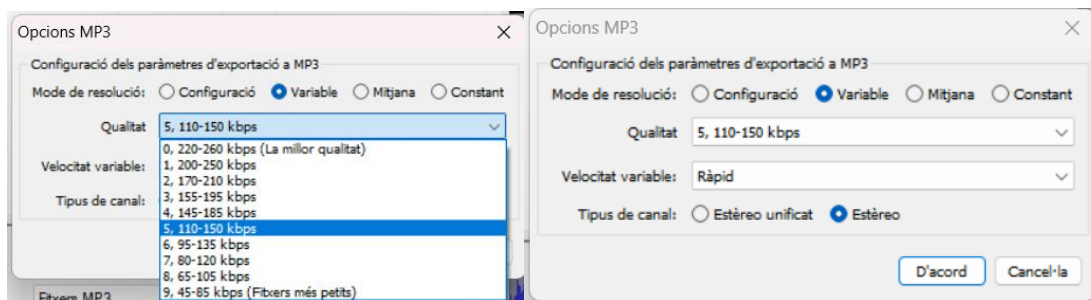


Apartat b

Els paràmetres triats per l'exportació a MP3 és una qualitat de 5 i la tasa de bits mitjana entre els 110-150 kbps.

La tria d'una tasa de bits mitjana baixa, juntament amb la velocitat variable ràpida i el tipus de canal estèreo, és una elecció equilibrada per observar els efectes de la compressió MPEG mentre es compara amb l'original en format WAV. Aquesta configuració proporciona una bona combinació entre qualitat d'àudio, temps de compressió i fidelitat a l'original per a fins d'avaluació i comparació.

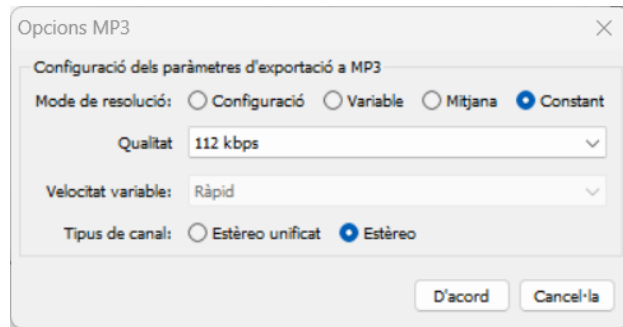
Opcions triades:



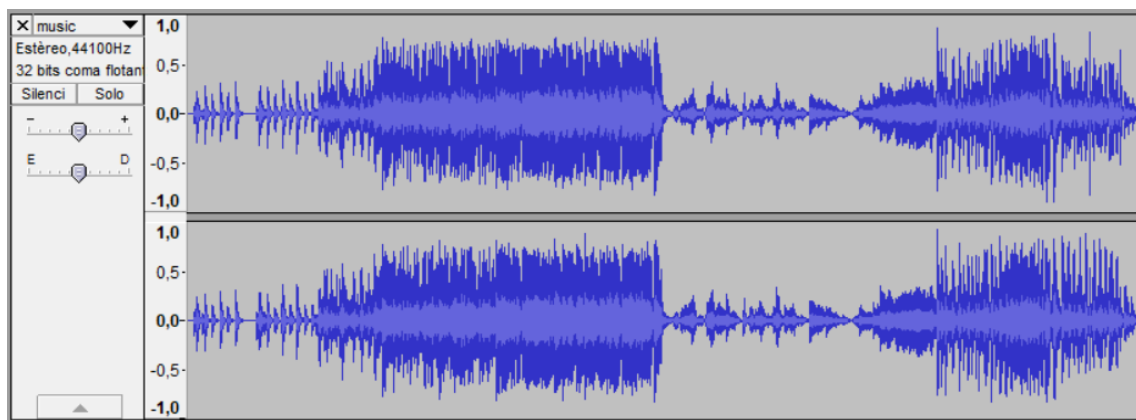
Apartat c

Una vegada ja tenim l'àudio codificat en VBR amb una tasa de bits en entre 110 i 150 kbps, fer la una versió codificada constant de CBR la tasa de bits de 112 kbps.

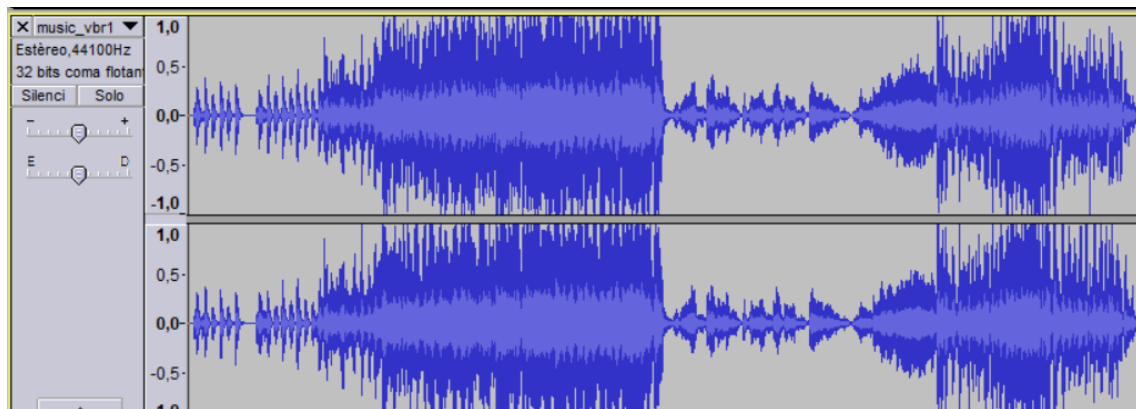
Triar una tasa de bits constant de 112 kbps després de codificar en VBR permet una comparació directa, assegura una qualitat constant en tot l'arxiu, pot estalviar espai i millorar la compatibilitat amb alguns dispositius o plataformes.



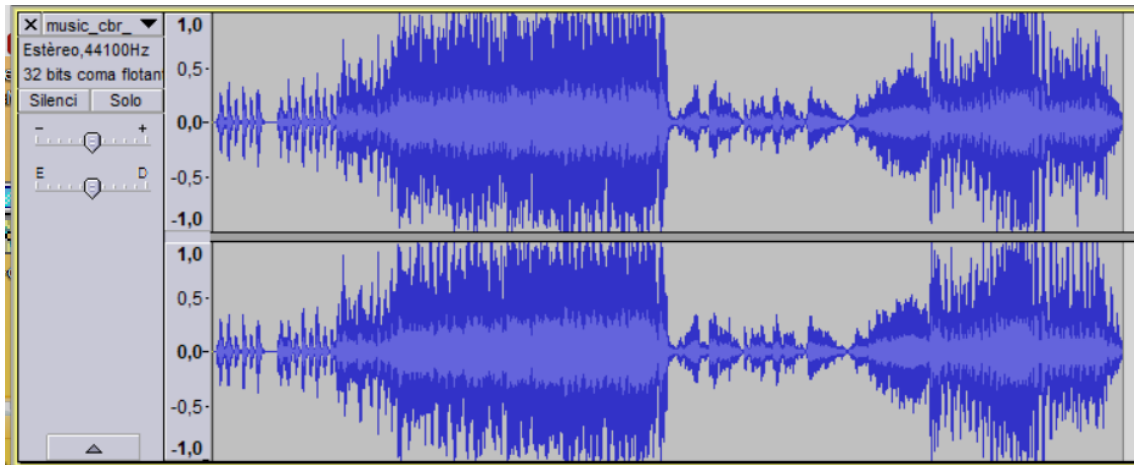
Original music.wav:



Fitxer music.mp3 Mode VBR 110-150:



Fitxer music.mp3 Mode CBR 112:



Les diferències entre aquests tres fitxers es poden apreciar escoltant-los seguidament i amb molta atenció. El fitxer original d'àudio .wav és d'alta qualitat, ja que guarda les dades d'àudio sense compressió o amb una compressió molt lleu, el que permet mantenir tots els detalls sonors de l'enregistrament original. Això resulta en una millor fidelitat del so de la cançó amb una reproducció clara i sense efectes addicionals.

En canvi, amb els fitxers CBR i VBR es poden notar els canvis produïts per la compressió de l'àudio original. En el cas del fitxer CBR, la taxa de bits constant de 112 kbps és relativament baixa per a l'estàndard MP3, cosa que comporta una pèrdua notable de qualitat comparada amb el WAV original. Aquesta baixa taxa de bits pot no és suficient per codificar acuradament tots els detalls de l'àudio, particularment en les freqüències altes, com els cops de bateria inicials, que sonen més metal·litzats. També apareix el fenomen de l'efecte de doble àudio, que es manifesta com a ecos o repeticions lleugeres.

Pel que fa al fitxer VBR, la taxa de bits variable de 110-150 kbps es pot ajustar a la complexitat de l'àudio en diferents moments de la pista. Aquest mode ofereix una millor qualitat que CBR a taxa de bits similars perquè pot augmentar la taxa de bits durant sons més complexos i reduir-la quan l'àudio és més simple. Encara que VBR tendeix a tenir una millor qualitat de so que CBR, es poden detectar diferències en el so, especialment en les freqüències altes dels cops de la bateria, presentant també un so de tambor metal·litzat, indicant una pèrdua de detalls en les freqüències més altes i l'efecte de doble àudio en algunes parts de l'àudio.

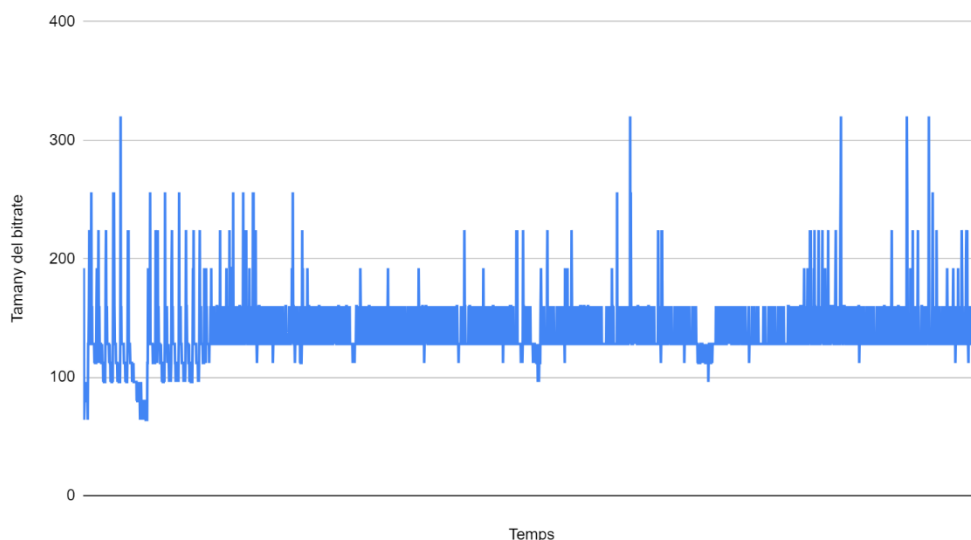
Apartat d

Fet el gràfic de la taxa de bits en el fitxer music.mp3 comprimit amb el mode VBR i un bitrate de 110-150 kbps.

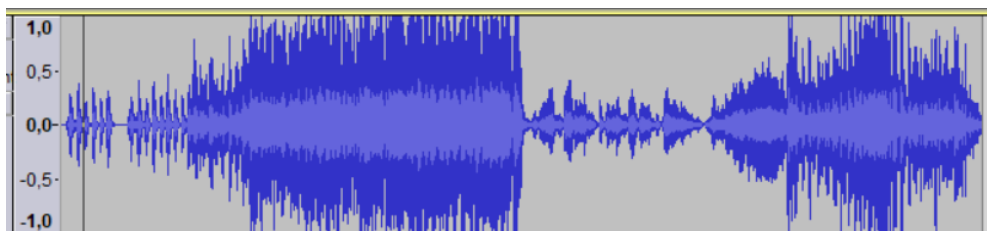
Podem veure les diferents mides del paquets utilitzant l'aplicació: AnaMP3, que retorna el bitrate de cada paquet.

Després amb excel es pot fer una gràfica de tots aquests:

Evolució temporal bitrate VBR



Àudio music.mp3 VBR:



Al gràfic superior és la representació de la variació de la taxa de bits en un fitxer MP3 codificat amb VBR i al gràfic inferior, és l'ona sonora del fitxer d'àudio corresponent. Si s'observa detalladament, hi ha una correlació entre els pics de la taxa de bits i els moments de major intensitat en l'ona sonora, que sovint representen els sons de l'àudio amb més informació sonora o complexitat.

Freqüències altes, com ara els detalls de la bateria o altres efectes sonors amb components de freqüència elevada, requereixen més bits per ser codificats eficaçment per preservar la seva claredat i detall. Això justifica per que en les seccions del gràfic VBR on veiem els pics més alts, corresponen a les parts de la música on les freqüències altes són les més predominants.

El que podem afirmar és que el VBR s'adapta correctament a la necessitat d'àudio. Quan hi ha un augment de complexitat o d'energia en el senyal d'àudio, com aquestes freqüències altes, el VBR incrementa la taxa de bits per preservar la qualitat. Aquesta adaptabilitat és un dels avantatges del VBR sobre el CBR, ja que permet una millor gestió de l'espai d'emmagatzematge que al mateix temps que manté la qualitat d'àudio en les seccions més crítiques.

Exercici 6- Streaming d'àudio

Apartat a

Mp3CBR112.pcappng

Los timestamps incrementan de 7053 , el payload es de mpeg1 layer 3

El marker bit el trobem a 0 al final de la transmissió ja que indica que és el final d'un marc d'àudio.

timestamps:

```
Time=1775175238
```

```
Time=1775182291
```

payload type:

```
Payload type: MPEG-I/II Audio (14)
```

marker bit:

```
10... .. = Version: RFC 1889 Version (2)
..0. .... = Padding: False
...0 .... = Extension: False
.... 0000 = Contributing source identifiers count: 0
1... .... = Marker: True
```

mp3VBR110.pcappng

timestamps:

- Las diferencias entre els timestamps son de 7053

```
Time=3062467373  
Time=3062474426  
Time=3062481479  
Time=3062486181  
Time=3062493234
```

payload type:

- El payload type es MPEG1 layer 2

Payload type: MPEG-I/II Audio (14)

marker bit:

```
10... .. = Version: RFC 1889 Version (2)  
..0. .... = Padding: False  
...0 .... = Extension: False  
.... 0000 = Contributing source identifiers count: 0  
1... .... = Marker: True
```

musicPCMwav.pcapng

timestamps:

```
Time=665391590
Time=665391939
Time=665392289
Time=665392639
Time=665392989
```

payload type:

Payload type: 16-bit uncompressed audio, stereo (10)

maker bit:

```
10... .. = Version: RFC 1889 Version (2)
..0. .... = Padding: False
...0 .... = Extension: False
.... 0000 = Contributing source identifiers count: 0
1... .... = Marker: True
```

Conclusions:

- Comentari sobre les dades:

Format MP3:

- Tipus de càrrega: MPEG1 Layer 3. Aquesta informació indica que l'àudio es codifica utilitzant l'estàndard MPEG1 Layer 3, un algorisme de compressió amb pèrdua que redueix la mida del fitxer en eliminar algunes dades d'àudio.
- Dades d'àudio: Comprimits amb pèrdua d'informació. Com es va esmentar anteriorment, la compressió amb pèrdua significa que s'elimina una part de les dades d'àudio originals, cosa que pot afectar la qualitat del so.
- Marcador de bits: 1 per indicar el final de la transmissió. Aquest bit indica al descodificador que el fitxer d'àudio s'ha acabat.

Format PCM:

- Tipus de càrrega: PCM (Premeu Code Modulation). PCM és un format d'àudio sense comprimir que representa el senyal d'àudio original com una sèrie de valors discrets. Això vol dir que es conserven totes les dades d'àudio originals, el que resulta en la millor qualitat de so possible.
- Dades d'àudio: Sense comprimir. Com es va esmentar anteriorment, la manca de compressió significa que es conserven totes les dades d'àudio originals, cosa que garanteix la màxima fidelitat.
- Marcador de bits: 1 per indicar el final de la transmissió. Similar al MP3, aquest bit indica al descodificador que el fitxer d'àudio ha acabat.

Apartat b

Totes aquestes dades han estat extretes de les capçaleres MPEG (per als .MP3)

Formato MP3CBR112

00 00 00 00 ff fb 82 04

Resum:

- La pila de protocols per a la transmissió d'àudio és Ethernet -> IP -> UDP -> RTP.
- El bitrate és de 112 kbps.
- La freqüència de mostreig és de 44.1 kHz.
- Cada paquet RTP conté 3 trames d'àudio de 26 ms, cosa que equival a 78 ms d'àudio.
- La mida dels paquets RTP és de 1155 o 1156 bytes.
- El bitrate a nivell IP és de 117.025 kbps.

Detalls:

- Pila de protocols:

La pila de protocols utilitzada per a la transmissió d'àudio és la següent:

- Ethernet: S'utilitza per transmetre dades a nivell d'enllaç de dades.
- IP: S'utilitza per a la transmissió de dades a nivell de xarxa.
- UDP: S'utilitza per a la transmissió de dades sense connexió a nivell de transport.
- RTP: S'utilitza per a la transmissió de dades d'àudio i vídeo a temps real.

Bitrate i freqüència de mostreig:

- El bitrate és de 112 kbps, cosa que significa que es transmeten 112.000 bits per segon.
- La freqüència de mostreig és de 44.1 kHz, cosa que significa que es prenen 44.100 mostres d'àudio per segon.

Detalls del paquet RTP:

- Cada paquet RTP conté 3 trames d'àudio de 26 ms, cosa que equival a 78 ms d'àudio.
- La mida dels paquets RTP és de 1155 o 1156 bytes.
- La base de temps de rellotge RTP per a àudio MPEG és de 90.000 Hz.

Càlcul del bitrate a nivell IP:

- Per calcular el bitrate a nivell IP, cal restar els 14 bytes de la capçalera Ethernet de la mida del paquet RTP.

$$\text{Mida del paquet RTP} - \text{Capçalera Ethernet} = \text{Dades IP}$$

$$1155 \text{ bytes} - 14 \text{ bytes} = 1141 \text{ bytes}$$

$$1141 \text{ bytes} * 8 \text{ bits/byte} = 9128 \text{ bits}$$

Dividint el nombre total de bits per la durada del paquet en segons, s'obté el bitrate a nivell IP:

$$\text{Bitrate IP} = 9128 \text{ bits} / 0.078 \text{ s} = 117.025 \text{ kbps}$$

Conclusions:

- La pila de protocols utilitzada per a la transmissió d'àudio és Ethernet -> IP -> UDP -> RTP.
- El bitrate és de 112 kbps a nivell RTP i de 117.025 kbps a nivell IP.
- La freqüència de mostreig és de 44.1 kHz.
- Cada paquet RTP conté 3 trames d'àudio de 26 ms.

Formato MP3VBR110

Per a l'anàlisi del següent fitxer agafarem 3 paquets de mides diferents per fer-los servir com a mostres ja que en ser VBR cada paquet pot ser diferent.

Pila de protocols: Ethernet -> IP -> UDP -> RTP

Paquet

```
RTP 1414 PT=MPEG-I/II Audio, SSRC=0x6A320007, Seq=12578, Time=3061327148
```

Primera trama:

ff fb 90 04

- Mpeg v1 layer 3
- Bitrate: per a v1 l3 i 9 = 1001 tenim un bitrate de 128 kbps
- Fm: en ser 0 la freqüència serà 44100hz

Una altra trama:

ff fb a0 04

- Mpeg v1 layer 3
- Bitrate: per a v1 l3 ia = 1010 tenim un bitrate de 160 kbps
- Fm: en ser 0 la freqüència serà 44100hz

Cada paquet transporta una quantitat diferent de trames d'àudio per paquet ja que el bitrate és variable, podem veure quantes trames porta un paquet mirant la diferència del timestamps amb el paquet anterior, per exemple:

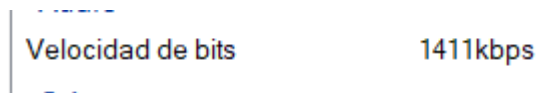
```
789 PT=MPEG-I/II Audio, SSRC=0x6A320007, Seq=12577, Time=3061324797
1414 PT=MPEG-I/II Audio, SSRC=0x6A320007, Seq=12578, Time=3061327148
```

La diferència entre timestamps és de 2351 ticks llavors $2351/90000 \text{ ticks/s} = 26 \text{ms}$ com una trama d'àudio mpeg són 26 ms podem concloure que porta una trama d'àudio i 26ms d'àudio.

FORMATO musicPCMwav

16 bit uncompressed àudio per 2 canals ja que és estèreo són 4 bytes d'àudio, tots els paquets són de 1454 bytes als quals cal restar-los 54 bytes de capçaleres, per tant el paquet sense capçaleres són 1400 bytes que entre 4 ens dóna que tenim 350 trames d'àudio de 4 bytes per paquet.

Analitzant el fitxer .wav i veient les seves propietats veiem que la velocitat de bits és de 1411 kbps.



llavors si cada mostra són $16 \times 2 \text{ bits} = 32 \text{ bits}$, i sabem que cada segon s'envien 1411000 bits per tant a partir d'aquestes dades si són 141100 bits cada segon si dividim entre allò que val cada mostra ens dóna que tenim una freqüència de mostreig de 44.1hz

Si dividim la freqüència de mostreig entre la diferència de timestamps, que és de 350, obtenim que per cada paquet són 8 ms d'àudio.

El temps que reporta Wireshark que hi ha entre paquets és de 20 ms per al paquet que hem seleccionat per tant el bitrate serà de:

$$1440 \times 8 / 0.02s = 576 kbps$$

1454 PT=16-bit uncompressed audio, stereo, SSRC=0xA2D39547, Seq=55708, Time=664668536

0.020858

Conclusió

Les nostres conclusions generals de la pràctica són:

- **Anàlisi d'Ondes Sonores:** L'estudi detallat de les ones sonores mostra com la compressió afecta la qualitat de l'àudio. Els diferents paràmetres de codificació, com el bitrate i el format, tenen un impacte directe en la nitidesa i la fidelitat del so.
- **Comparació de Còdecs:** La comparació entre còdecs amb pèrdua i sense pèrdua revela que mentre els primers redueixen significativament la mida del fitxer, també poden comprometre detalls importants de l'àudio. Els còdecs sense pèrdua, com FLAC, preserven la qualitat original, però resulten en fitxers més grans.
- **Impacte del Bitrate:** S'ha observat que un bitrate més elevat millora la qualitat de l'àudio, especialment en formats comprimits. Això es tradueix en una major definició i menys distorsió, essent crucial per a aplicacions on la qualitat del so és primordial.
- **Percepció Subjectiva:** La qualitat percebuda de l'àudio pot variar segons l'usuari, depenent de factors com el tipus de contingut (veu o música), el context d'ús (streaming o arxius locals) i les preferències personals. Aquesta subjectivitat suggereix la necessitat de considerar l'audiència objectiva quan es tria un còdec i una taxa de bits.