



M Ű E G Y E T E M 1 7 8 2

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Hálózati Rendszerek és Szolgáltatások Tanszék

Páll Tünde Anna

**PSZICHOAKUSZTIKAI
MODELLEK AZ
AUDIOTÖMÖRÍTÉSBEN**

KONZULENS

Dr. Rucz Péter

BUDAPEST, 2018

Tartalomjegyzék

Kivonat.....	5
Abstract.....	6
1 Bevezetés	7
2 Az emberi hallás.....	8
2.1 Az emberi hallás tulajdonságai	9
2.1.1 Frekvenciafüggő tulajdonságok	9
2.1.2 Frekvenciatartománybeli elfedés	11
2.1.3 Időtartománybeli elfedés.....	12
3 Veszteséges audiotömörítés	14
3.1 A pszichoakusztikus kódoló működése	16
4 MODEL-1 megvalósítása Matlab környezetben.....	18
4.1 Spektrum-analízis és a hangnyomásszint normálása	18
4.2 Tonális- és zaj maszkolók meghatározása	19
4.3 Decimálás és a maszkolók újrendezése	20
4.4 Egyenkénti maszkoló küszöbök számítása	21
4.5 A globális maszkoló görbe számítása	22
4.6 Érzékelésalapú bit-kiosztás	22
5 Audiojelek tömörítése a MODEL-1 felhasználásával	24
5.1 Szinusz-jel elemzése	24
5.2 A kódoló bemutatása zenei hangminta egy szeletére	26
5.3 Több időszelvény kódolása.....	27
5.3.1 Hann-ablak.....	27
5.3.2 Hosszabb hangminták kódolása.....	28
5.3.3 Hirtelen tranziensek hatása	30
6 Eredeti és kódolt hangminták objektív összehasonlítása PEAQ modell segítségével.....	33
6.1 PEAQ kialakulásának története	33
6.1.1 Érzeti mérési elvek.....	35
6.1.2 Fül-modell.....	37
6.1.3 Észlelési modell	37
6.2 PEAQ.....	38

6.2.1	Analízisre való előkészítés.....	38
6.2.2	PEAQ verziók.....	38
6.3	PEAQ Basic verzió megvalósítása Matlab környezetben.....	41
6.3.1	PEAQ működése Matlab-ban	41
6.4	Saját audiokódoló vizsgálata a PEAQ basic modellel	42
7	Változó ablakméretek.....	44
7.1	Változó ablakméret megvalósítása	44
7.2	Változó ablakméretek alkalmazó kódoló eredményeinek vizsgálata PEAQ- kel	47
8	Összegzés.....	49
	Irodalomjegyzék.....	50

HALLGATÓI NYILATKOZAT

Alulírott **Páll Tünde Anna**, szigorló hallgató kijelentem, hogy ezt a diplomatervet meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy hitelesített felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Kelt: Budapest, 2018. 12. 06.

.....
Páll Tünde Anna

Kivonat

A diplomatervezés feladatom egy veszteséges audiotömörítő létrehozása volt Matlab környezetben. Ehhez először a szükséges háttérismereteket kellett elsajátítanom.

Jelen dokumentumban megtalálható az emberi hallás jellemzése, és annak bemutatása, hogy a hallás tulajdonságaiból kifolyólag milyen hiányosságokat használhatunk ki audiotömörítés során.

Dolgozatom következő lépése a kódoló implementálása Matlab-ban. Ehhez a szükséges irodalomkutatás elvégzése után, egy megfelelő dokumentációt használva alapul elkészült egy veszteséges kódoló. Ennek a kódolónak a segítségével eleinte egyszerűbb mintajeleken, majd zenei időszeleteken keresztül bemutatom a hallás korábban megismert tulajdonságait és az ezeket kihasználó alkalmazott módszereket.

A kódolón átalakításokat és kiegészítéseket végeztem, így nem csak rövidebb időszeletek, hanem tetszőlegesen hosszú hangminták kódolására is képes.

A meghallgatásos tesztekben kapott szubjektív eredmények nem bizonyultak elég informatívnak a kódoló minősítéséhez, ezért egy objektív minőség mérőszám számítására volt szükség. Erre alkalmas a PEAQ audiominőség elemző. A dokumentum tartalmazza a PEAQ kialakulásának történetét és egy létező implementációjának bemutatását. Kisebbségi átalakítások után ez a Matlab-kód alkalmas az általam készített kódoló vizsgálatára. Több különböző hangmintát azonos kódolási paraméterek mellett kódolva a kódoló képességeit a PEAQ által számított minőség mérőszámok segítségével vizsgálhatjuk.

A referenciaként választott kódolónak egy hiányosságával szembesültem, melynek javítására javaslatot teszek, majd a javasolt megoldást meg is megvalósítom. A módosított kódolóval így újra kódolhatjuk ugyanazokat a hangmintákat, és a PEAQ segítségével objektív mérőszámokat is kapunk eredményeink vizsgálatára.

Az elkészült audiotömörítő és az azt bemutató jelen dokumentum hasznos lehet további oktatási célokra is, mivel a kódolás minden lépése nyomonkövethető, a számítás részeredményei külön is megjeleníthetőek, elemezhetőek.

Abstract

The objective of this thesis is to create a lossy audio encoder in Matlab environment. To achieve this, I had to acquire new knowledge about the functionality and implementation of such encoders.

This document first describes the characterization of human hearing and focuses on the deficiencies that can be exploited in psychoacoustics-based lossy audio compression methods.

The next step is to implement the encoder in Matlab. After reviewing the corresponding literature, I found a proper documentation, based on which the lossy encoder was created. Using this encoder I introduce the aforementioned properties of human hearing and the applied psychoacoustical signal processing methods through simpler sample patterns and a music time slice.

I also made changes and improvements on the encoder, so it can not only code shorter time slices, but is able to code arbitrarily long sound samples.

The subjective results obtained by listening tests did not prove to be sufficiently informative to rate the encoder algorithm, thus, an objective quality measurement method needed to be established. The PEAQ, which is an audio quality analyzer model is suitable for this need. The thesis contains the history of the development of PEAQ and discusses the operation of an existing implementation. After minor modifications, this Matlab code became suitable for testing my encoder. Multiple voice samples encoded with the same encoding parameters can be tested using the quality measurements calculated by PEAQ in order to examine the capabilities of the encoder algorithm.

I realized a significant deficiency of the reference encoder, I proposed an improvement to overcome the limitations, and then implemented my solution. Using the modified encoder, the same sound samples can be re-encoded objective measurements can be used to validate the test results.

The completed audio encoder and this document that discusses it in detail can also be useful for further educational purposes, as all steps of the encoding are clearly followable, and the intermediate results of each step can be visualized and examined separately.

1 Bevezetés

Ebben a dokumentumban a Diplomaterv témám kereteiben elsajátított ismereteket és elért eredményeket mutatom be. Először az emberi hallás tulajdonságait ismertetem, majd rátérek, hogy ezek a tulajdonságok milyen módon használhatóak ki a veszteséges audiotömörítés során. Ezután bemutatok egy lehetséges pszichoakusztikai modellt, mely már tömörítésre használható. Ezt a modellt elkészítettem Matlab környezetben, és ennek segítségével ábrákon keresztül bemutatom, a kódolás menetét, lépéseit, illetve esetleges hiányosságait.

Objektív értékelésre alkalmas PEAQ modell bemutatása és működésének megismerése után, értékelem vele az elkészült audiotömörítőt, így szubjektív meghallgatásos tesztek mellett egy objektív mérőszámmal is jellemezhetem az újrakvantált hangmintát. Ennek segítségével megismert hiányosságok kijavítására javaslatot teszek, majd elméleti háttérrel alátámasztva ezt a javaslatot megvalósítom. Így elkészül egy kiegészített, korábban ilyen formában nem létező, audiotömörítő. Végül ennek a módosított kódolónak is megvizsgálom az eredményeit és ezeket is bemutatom.

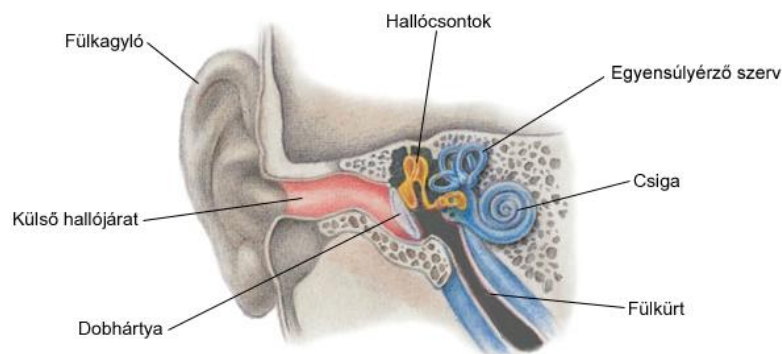
A mai világban körülvesznek minket a veszteséges audiotömörítést alkalmazó hangminták, és a tömörítő módszerek működését mégsem feltétlenül ismerjük. Ennek a dokumentumnak a segítségével bárki jobban megértheti, hogy miért használjuk és hogyan működnek az ilyen audiotömörítő eljárások. Mivel az elkészített kódoló alapja egy régebbi modell, így ennek segítségével nem a legjobban tömörített, és adott tömörítési arányhoz tartozó legjobb hangminőséget kapjuk, viszont az egyszerűbb implementáció miatt könnyebben vizsgálható és megérthető a működése.

Számomra ez a téma több okból is érdekes. Egyrészt mivel a hallással foglalkozik, azért sokkal kézzelfoghatóbb téma, mint sok másik. Sok érdekességet tanulhattam meg munkám során, amelyek később hasznos információk lehetnek, elmondhatom, hogy erről is vannak ismereteim. Még emlékszem, hogy általános iskolában, zenét megosztani csak úgy tudtunk, hogy CD-t írtunk. Később már számítógépen tároltuk zene file-okat, de nehezen lehetett megszerezni egy-egy új albumot. Így számomra érdekes volt megismerni, hogy mégis hogyan terjedt el ennyire a digitális zenemegosztás.

2 Az emberi hallás

A veszteséges audiotömörítő eljárások az emberi hallás sajátosságait és hiányosságait veszik figyelembe, így érdemes ezeket megismerni először.

Hallásnak azt a folyamatot nevezzük, amikor hallószervünkön keresztül feldolgozzuk a beérkező hanghullámokat. A hanghullámok levegőben mechanikai úton terjedő rezgések. Az alábbi 2.1. ábrán látható az emberi hallószerv felépítése, mely három észre osztható.



2.1. ábra. Az emberi hallószerv felépítése

Külső fülnek nevezzük a fülünk látható részét, a fülkagylót. Ez segíti a különböző irányból érkező hanghullámok irányítását a hallójárat felé.

Középső fülnek nevezzük a hallójáratot. A dobhártya választja el a belső fület a külvilágtól. A dobhártya két oldalán különböző nyomás uralkodik. A hanghullámok pedig a nyomásváltozással rezgésbe hozzák a dobhártyát.

A dobhártya és a belső hallószerv közötti kapcsolatot a hallócsontok teremtik meg. Nevük, rendre a dobhártya felől, kalapács, üllő és kengyel. Ezek felelősek, a dobhártya rezgésén keresztül, a hangnyomás-változás mechanikai energiává való átalakításért. Továbbá lehetővé teszik, hogy a hallószervünknek több mint 120 dB-es dinamikataromány legyen.

A fülünk halláson kívül felelős a megfelelő egyensúlyozásért is, ezt a feladatot látja el az egyensúlyérző szerv.

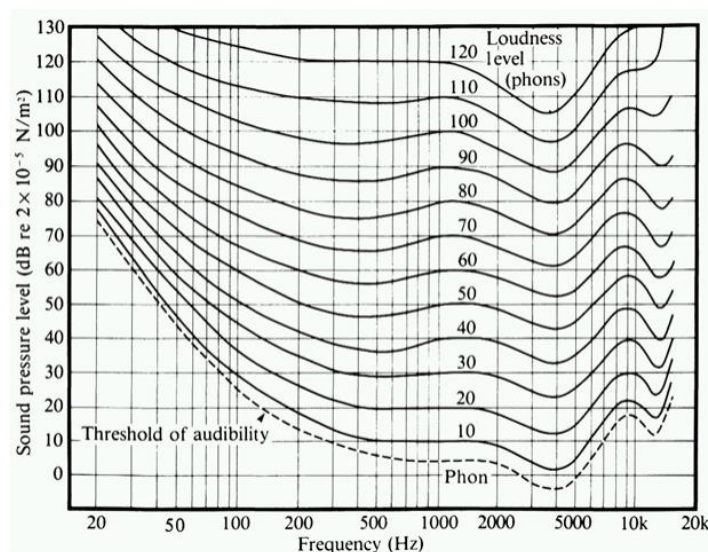
Szintén a belső fülben található a csiga, az úgynevezett corti szerv, ez egy két és félszer feltekert cső. Ennek bejáratához kapcsolódik az ovális ablakon keresztül a kengyel. A csigában található tér két részre osztott hosszában, ezeket választja el a baziláris membrán. Amikor a kengyel rezgésbe hozza az ovális ablakot, akkor a csigában lévő folyadékban hullámok keletkeznek. A baziláris membránon találhatóak a halló sejtek. Ezek elhelyezkedése határozza meg, hogy milyen frekvencia érzékeléséért felelősek. Az ovális ablakhoz közelebb eső hallósejtek a magas hangok (nagy frekvenciák) hallásában játszanak szerepet, a csiga belsejében lévők pedig értelem szerűen a mély hangokért (alacsony frekvenciák) felelősek. Tehát a csiga és a benne található érzékelők határozzák meg az emberi hallás frekvencia tartományát. Természetesen más tényezők is közrejátszanak. A felnőtt emberi hallást a 20 Hz-es tartománytól a 20 kHz-es tartományig szoktuk értelmezni. Azonban ez meglehetősen függ az életkortól, és a hallószervet ért korábbi megterhelő hatásoktól.

A hallósejtekhez kapcsolódó idegpályák felelősek az inger agyba való eljuttatásához, ahol feldolgozásra kerül az információ.

2.1 Az emberi hallás tulajdonságai

2.1.1 Frekvenciafüggő tulajdonságok

A fülünk által érzett hangosság szint a mérhető hangnyomásszintnek nemlineáris függvénye, ezért dB skálával szokás jellemezni. A hangosság érzet frekvenciafüggő, ezt jellemzik a Fletcher-Munson görbék, melyek a 2.2. ábrán láthatóak.



2.2. ábra. Fletcher-Munson görbék

A vízszintes tengelyen a frekvencia logaritmikus skálázásban látható, a függőlegesen pedig a hangnyomás szint, szintén logaritmikus skálán. Az alsó szaggatott görbe a hallás küszöb. A küszöb alatti hangnyomásszinteket nem képes érzékelni a fülünk, többek között ezt is használják ki a veszteséges audiotömörítő eljárások. A hangnyomásszint határozza meg a hangosságot, azonban a hangosság érzet a frekvenciától is függ. Ezt jellemzik a görbék. Az 1 kHz-en érzékelt hangnyomásszinthez tartozó hangosságot véve alapul, ehhez hasonlítjuk a további frekvenciákon érzékelt hangosságokat. Ugyanolyan hangosság érzékeléséhez alacsony és magas frekvenciákon nagyobb hangnyomásszint szükséges, mint az 1 kHz középfrekvenciájú sávban. Ezeket a hangnyomásszint értékeket kötik össze az ábrán látható görbék. Tehát egy görbe mentén azonos hangossággal érzékeljük az adott frekvenciájú hangot. A görbékre írt mennyiségek a phon értékek, melyek a referenciaként vett 1 kHz-es hang hangnyomásszintjével egyeznek meg. A görbék mentén így, a különböző frekvenciákhoz tartozó azonos hangosság érzetet keltő hangnyomásszint olvasható le. Tehát ha referenciának vesszük egy 1 kHz-es hang 10 phon-os hangosságát (így egyben 10 dB hangnyomásszintet), akkor egy 100 Hz-es hang szintén 10 phon-os nagyságához 30 dB hangnyomásszintre van szükség.

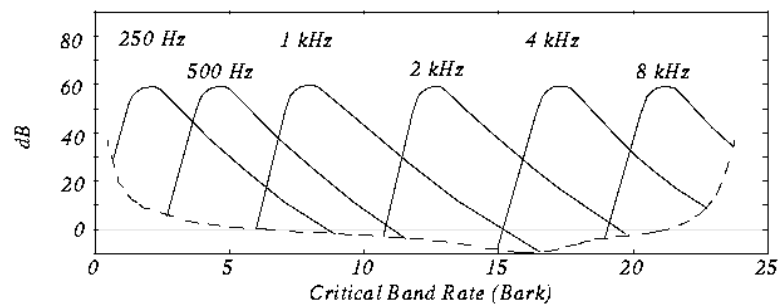
A görbéken szintén megfigyelhetőek az emberi hallás frekvencia korlátai. Az egyre erősebb hangok esetén a görbék fokozatosan laposodnak. A felső határ a fájdalom küszöb. A fájdalom határa nem egyezik meg a halláskárosodás határával, mely alacsonyabb. A halláskárosodás függ a hang erősségétől és az időtartamtól, ameddig a hangforrásnak ki van téve a hallószerv.

A görbék jól jellemzik az emberi hallás frekvenciafüggő érzékenységet. Az 1 kHz-et vettük alapul, de valójában a fülünk 3kHz környékén a legérzékenyebb. Ennek evolúciós okai lehetnek.

Egy oktáv hangmagasságbeli különbség 1 : 2 frekvencia arányt jelent. Mivel a hallásunk az azonos frekvencia arányokat érzékel azonos hangmagasságugrásoknak, ezért a pár száz Hz-es tartományban 100 Hz változás lényeges különbséget jelent. Míg a 10 kHz-es tartományban a 100 Hz-es változás a hangmagasságérzet szempontjából nem olyan meghatározó. Ebből következik az úgynevezett kritikus sávok szélességének frekvenciafüggése is, melyet szintén kihasználunk a veszteséges audiotömörítő eljárásokban.

A kritikus sávok kimérését Fletcher végezte el. Egy adott közép frekvenciához változó sáv szélességű zajt kevert. A sáv szélesség növekedésével a zavaró hatás nőtt. Az adott frekvenciához tartozó kritikus sáv határa ott definiálható, amelyen felüli sáv szélességű zaj már nem okoz változást. A kritikus sávok szélessége frekvencia függő, nagyobb frekvenciák esetén szélesebbek.

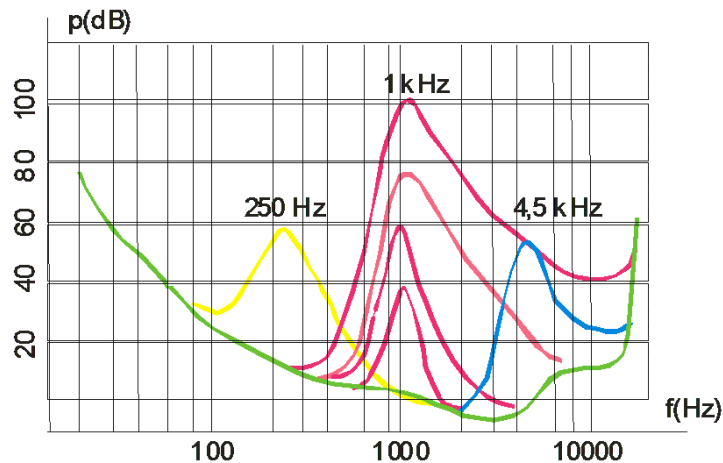
Ha két hang frekvenciabeli távolsága kritikus sávon belüli, akkor intenzitás alapján összegződik hangerejük. Ezeket a sávokat más néven Bark sávoknak is nevezzük, ebből alakult ki a Bark-skála, melynek segítségével hallásunk frekvenciafüggősége szemléletesebben ábrázolható. Összesen 24 db szabványos Bark-sáv definiált. Az alábbi 2.3. ábrán látható a hallásküszöb és néhány maszkológörbe Bark skálán ábrázolva, a maszkológörbéről pedig a következő alfejezetben részletesebben olvashatunk.



2.3. ábra. Bark-skálán ábrázolt hallásküszöb és kritikus sávok

2.1.2 Frekvenciatartománybeli elfedés

A pszichoakusztikai tulajdonságokat figyelembevevő audiotömörítő eljárások legmeghatározóbb eleme a frekvenciatartománybeli elfedés. Ha megszólal egy erős hang egy adott frekvencián, akkor a környező frekvenciákon történő kisebb hangnyomásszintű jeleket nem halljuk meg. Lényegében megemelkedik a hallásküszöb a megszólaló hang frekvenciájának környékén. Ez a baziláris membrán tehetetlenségének tudható be. Az alábbi 2.4. ábrán láthatóak példák különböző frekvenciájú és erősségű hangok elfedő hatására.



2.4. ábra. Frekvenciatartománybeli elfedési görbék

A rózsaszín görbéken megfigyelhető, hogy hogyan függ egy 1 kHz-es hang elfedése a hangnyomásszinttől. Minél nagyobb a hangnyomásszint, annál szélesebb a hozzá tartozó görbe. A görbék meredeksége a kisebb és nagyobb frekvenciák felé haladva különböző, nagyobb frekvenciák felé lankásabb. Továbbá a hangnyomásszint növelésével is változik a meredekség, a nagyobb amplitúdóval bíró hangok nagyobb maszkoló hatással vannak távolabbi nagyfrekvenciás hangokra, mint ugyanazon frekvenciájú, kisebb amplitúdójú hang jelek.

A sárga és kék görbéken a különböző frekvenciákra jellemző görbék vizsgálhatóak meg. Mindkettőre igaz, hogy a nagyobb frekvenciák irányába a görbe kiterjedése nagyobb. Viszont a kék görbe, mely lényegesen nagyobb frekvenciájú hanghoz tartozik, sokkal szélesebb. Az ábrán ez elsőre nem látható, de a logaritmus skálázásból kifolyólag megfigyelhető.

A későbbiekben ezt a tulajdonságot használjuk ki kódoláskor. Hiszen a megemelkedett küszöbszintből kifolyólag, nagyobb a megengedett zaj is. Tehát egy kevesebb biten kódolt szegmens nagyobb kvantálási zaja nem lesz hallható az elfedési jelenségből kifolyólag.

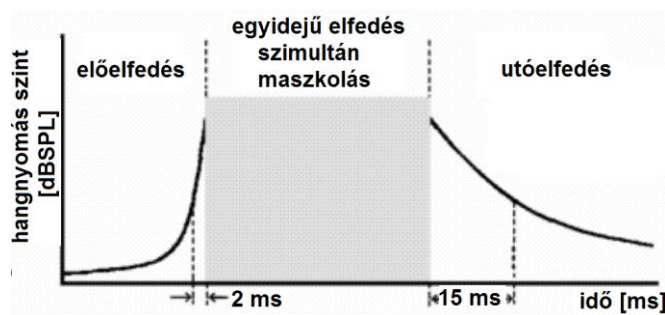
2.1.3 Időtartománybeli elfedés

Frekvenciatartománybeli elfedésen kívül létezik az úgynevezett időtartománybeli elfedés is. Egy erős hanghatás hatására, ez lehet hirtelen tranziens is, vagy egy hosszabb hang, a hang megszűnése után, hasonlóan a frekvenciatartománybeli párjához, kevésbé érzékelünk halkabb hangokat. Lényegében itt is ideiglenesen

megemelkedik a hallásküszöb. Ez az időtartomány változik a hang erősségétől és hosszától függően. Ezt a jelenséget nevezzük utóelfedésnek, melynek több oka lehet.

Ennél meglepőbb, az előelfedés jelensége. Pár milliszekundummal a hang fülünkbe való megérkezése előtt már megemelkedik a hallásküszöb. Legalább is így tűnik, pedig a jelenség valójában a hang feldolgozásával van kapcsolatban. A halkabb korábban megszólaló hangot a fülünk elkezd feldolgozni, és az idegpályán küldi az információt. Rövidesen utána érkezik egy sokkal erősebb, hirtelen hanghatás, ezt is feldolgozza fülünk és küldi az üzenetet az agynak. Jelenlegi tudásunk alapján úgy tűnik, hogy ez erősebb, hangosabb hanghatáshoz tartozó ingerek gyorsabban terjednek az idegpályákon, így az agyunkba előbb érkeznek meg, és elfedhetik a halkabb hatásokhoz tartozó, korábbi ingereket.

Ez a két elfedési jelenség együtt az időtartománybeli elfedés, és az alábbi 2.5. ábrán látható.



2.5. ábra. Időtartománybeli elfedési jelenségek

3 Veszteséges audiotömörítés

A digitális hanganyagok napjainkban való nagymértékű elterjedését a veszteséges audiotömörítő eljárások tették lehetővé, többek között a jól ismert MP3 is. Annak ellenére, hogy veszteséges tömörítési eljárás, mégis megmarad az élvezhető hangminőség. Egy átlagos minőségű MP3 bitfolyam bitsebessége 128 kbit/s. Ha 48 kHz-es mintavételi frekvenciát használunk, akkor mintánként mindössze 1,33 bit-et kapunk. Az audio CD-k 16 bites mintáihoz képest, ez 12-szeres tömörítési arányt jelent közel azonos minőségen. Így létrehozhatunk nagyon kis tárhelyet igénylő audio file-okat. Ezzel a digitális hanganyagok otthoni tárolása is lényegesen egyszerűbbé vált, bárki tárolhat a saját számítógépein, vagy más eszközein több tucat zenei albumot. Ezen felül a file-ok megosztása és terjesztése is kényelmessé vált. Már 10 évvel ezelőtt is, sokkal lassabb internet-hozzáféréseken keresztül is, percek alatt meg lehetett szerezni bármilyen zenei albumot.

Ennek következményeként megnehezedett a szerzői jogok védelme. Digitális zene file-ok megvételére iTunes és más hasonló programok használhatóak. Az ilyen módon letöltött zene file-ok lényegesen olcsóbbak, mint az audio CD-n megvett párjaik, de mégsem ingyenesek, mint az illegálisan letöltöttek. Az iTunes azzal is bíztat a szolgáltatásaik használatára, hogy amit egyszer digitálisan megvásároltunk az nem veszik el. Ezt úgy teszik lehetővé, hogy a felhasználó belépéshez használt adatai mellett, a már megvásárolt file-ok listáját is tárolják. Így az illegálisan letöltött és iTunes-hoz hozzáadott zenék mind elvesznek, ha valaki új PC-t vesz. A legálisan megvásárolt file-ok automatikusan letölthetőek az új PC-re miután bejelentkeztünk az iTunes-ba, és így nem szükséges mappák százait menteni és felmásolni az új eszközre. Továbbá ezzel a módszerrel a digitális tartalmak iránt preferenciát alakítanak ki. Hiába létezik az iTunes úgynevezett „rip-off” funkciója, amellyel saját audio lemezeinkről készíthetünk digitális, iTunes-ban használható másolatot, ezeket nem menti el a rendszer a felhasználó adataival együtt. Tehát egy új számítógép beszerzése esetén, újra végre kéne hajtani a műveletet az összes lemezen, ennél pedig sokkal egyszerűbb direkt digitálisan megvásárolni őket.

A legelterjedtebb audiotömörítő eljárás a már említett MP3. Ezen kívül más opciók is léteznek. Az iTunes által használt veszteséges audiotömörítő eljárás az AAC

(Advanced Audio Coding), mellyel azonos bitsebességen jobb hangminőség érhető el. Az AAC file név kiterjesztése a .m4a, ennek a container file-nak más kiterjesztései is léteznek, például a csengőhangokat megkülönböztető .m4r.

Egy harmadik elterjedt veszteséges audiotömörítő az Ogg Vorbis. Ogg a container formátumra utal, míg Vorbis maga az audiotömörítő formátum. Ez az eljárás a legeredményesebb az eddig tárgyaltak közül, jobb audiomínőséget elérhető vele azonos bitsebességnél.

Veszteségmentes audiotömörítő eljárások is léteznek. Megtartják a hangminőséget, de cserébe nem képesek olyan tömörítési arány elérése, mint a veszteséges eljárások. Legismertebb a FLAC, mely közel felezni képes a file méretét, mindeközben megőrizve a teljes tartalmat. Az Apple veszteségmentes tömörítő eljárása az ALAC (Apple Lossless elterjedtebb nevén).

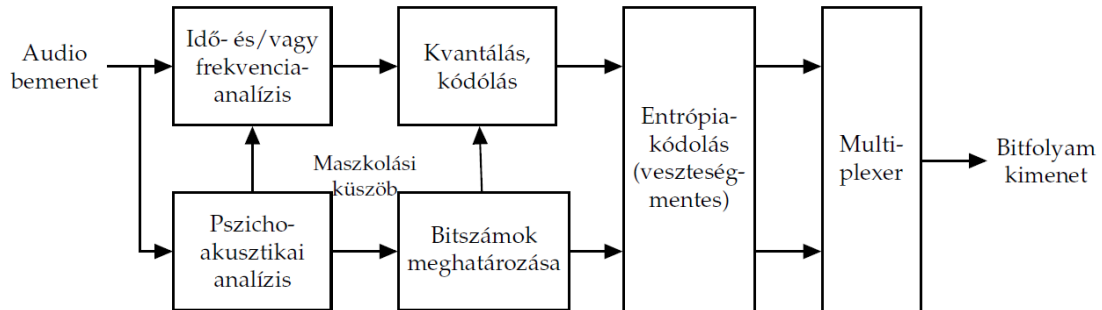
Zene file-ok terjesztése mellett számos felhasználási módja van az audiotömörítő eljárásoknak. Amikor filmet nézünk, nem az eredetileg felvett, vagy stúdióban kevert hangot halljuk vissza. Amikor egy videót hanganyag kíséri fontos annak szinkronizálása, és konstans bitsebessége. A tömörítő eljárások pedig nem feltétlenül konstans bitsebességgel kódolnának. Vannak olyan részek ahol kevés biten is lehet kódolni, de egy későbbi időpontban előfordulhat, hogy kevés bit nélkülözése lehetséges.

Konstans bitsebesség szükséges az online- zene és rádió hallgatásakor is. Így egyes időszakokban a minőség romlása elkerülhetetlen, de iteratív korrekció alkalmazható. Az MP3 fix méretű csomagokat használ, viszont egy nem teli csomagot egy másik csomag is kihasználhat, így nem romlik jelentősen az adott időszak minősége.

A konstans bitsebesség megkövetelésén felül egyszerűen dekódolhatónak is kell lennie a formátumnak. A kódolás lehet bonyolultabb, ha nem élő adásról beszélünk, de amikor lejátszani kívánjuk a hanganyagot, akkor elvárás a rövid dekódolási idő, például egy film esetében nem várhatunk arra, hogy a hang megszólaljon. Az egyszerű dekódolás egyúttal olcsó lejátszó eszközök előállítását és forgalmazását teszi lehetővé.

3.1 A pszichoakusztikus kódoló működése

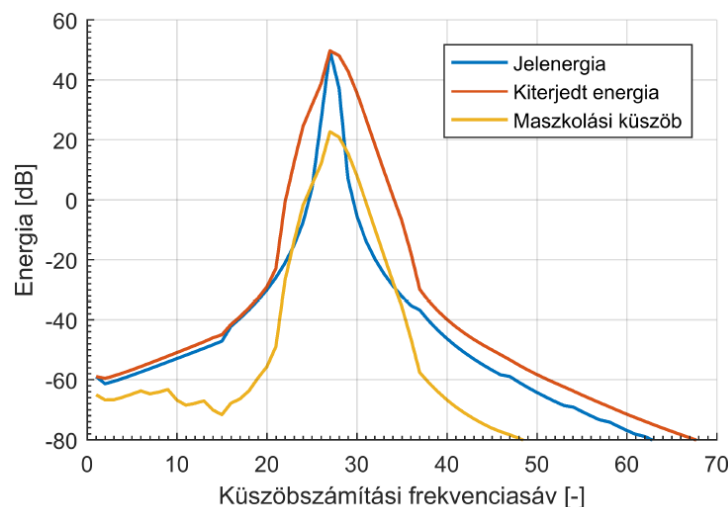
Az alábbi 3.1 ábrán egy általános blokkvázlat látható egy veszteséges audiokódoló működésére.



3.1. ábra. Veszteséges audiotömörítő általános blokkvázlata

A kódolandó audio bemenetet az idő- és frekvenciaanalízis mellett a pszichoakusztikai modul is elemzi (ezzel foglalkozunk részletesen a későbbiekben), amely az analízátor számára is szolgáltat információkat. A pszichoakusztikai analízis modul legfontosabb feladata a bitszámok meghatározásához szükséges globális maszkolási görbék kiszámítása. Ha ismert idő- és frekvenciaszeletenként a kódoláshoz szükséges bitszám, akkor elvégezhetjük a kvantálást. Ez a lépés a veszteséges kódolás, mert a bitszámot úgy választjuk meg, hogy a kódolásból adódó kvantálási zaj a maszkolási görbe alá essen, így az emberi fül nem hallja ezt a zajt, ha megfelelő volt az analízis.

Az alábbi 3.2 ábrán egy keskenysávú jel maszkolási görbéje látható. A globális maszkolási görbe ezek összegeként áll elő.



3.2. ábra. Keskenysávú jel maszkolási görbéje

Az analízis megkezdésekor a hanganyagot időszelletekre bontjuk. Ezek általában átlapolódó Hann-ablakok. Egy időszeletet FFT vagy DCT segítségével frekvenciatartományba transzformálunk, ahol megkezdődhet a frekvenciatartománybeli analízis. Detektáljuk a tonális (azonosítható hangmagasságú) és atonális (zajszerű) komponenseket, az előbbi komponensek erősebben maszkolnak, ezek felelősek a hallásküszöb megemelkedéséért. Kiválogatjuk ezek közül azokat, amelyek valószínűleg tényleg felelősek a környékbeli frekvenciák maszkolásáért, tehát egy sávon belül a legnagyobb csúcsoakat tartjuk meg további számításainkhoz. Mindegyik maszkoló komponenshez a saját frekvenciájához és nagyságához tartozó maszkoló-görbét rendeljük. Végül ezeket összegezve kapjuk meg a globális maszkolási görbét, melyet a kvantálási bitszámok meghatározásához használunk.

A kódolás másik kimenete a jel érzeti entrópiája, mely egy egyszerű mérőszám. Ha az érzeti entrópia magas, akkor a jelben hirtelen tranziens következik és érdemes az időablak méretét változtatni, kisebbre venni. Tehát az érzeti entrópia kiszámítása az időszeltek méretének meghatározásához szükséges.

A kvantált minták egy veszteségmentes entrópia-kódolással kerülnek a kimeneti bitfolyamba. Az entrópia-kódolás lehet például Huffman-kódolás. A mintákon kívül plusz adatokat is tartalmaz a bitfolyam, melyek a visszafejtéshez vagy éppen a csomagok szinkronizációjához lehetnek szükségesek.

4 MODEL-1 megvalósítása Matlab környezetben

Ebben a fejezetben, az előzőekben tárgyalt jelenségek felhasználásának megvalósítását követjük végig. Az alábbiakban az úgynevezett MODEL-1 megvalósítását részletezem [7]. A MODEL-1, melyet az MPEG-1, Layer 3 szabvány ajánlasként ismertet, meghatározza a maximálisan megengedhető kvantálási zajenergiát minden kritikus sávra, amely mellett a zaj nem hallható az emberi hallás számára. Az egyik módszer 512-pontos FFT-t használ a spektrum nagyfelbontású analiziséhez. Majd becslést ad az összes bemeneti keret maszkolóira, melyek egy időben maszkolnak az adott keretben. Ezek a jel spektrumából megállapítható tonális- és zaj maszkolóknak köszönhetőek. A következő lépés a globális maszkológörbe becslése additív összegzéssel a tonális és a-tonális egyedi maszkolókból az eredeti 256 frekvencia bin egy adott részhalmazára. A továbbiakban ezeknek a műveleteknek lépésről-lépésre való magyarázata következik. E lépések megvalósításával készült el a veszteséges audiotömörítést megvalósító végső Matlab kódom.

A Model-1 alkalmazható mind három MPEG-1 Layer-re, viszont a szabvány ajánlata szerint célszerű csak Layer I és Layer II-vel való használata. Míg a Model-2 javasolt Layer III számára, ez az általunk ismert MP3. A Layer-ek különböző tömörítési szinteket határoznak meg audiojelekre.

4.1 Spektrum-analízis és a hangnyomásszint normálása

Első lépésként a spektrum analizálását és a normalizálást végezzük el. Ennek célja, hogy a bemenetünket, spektrum komponenseket SPL-ben (hangnyomásszint) fejezhessünk ki, melyek az eredeti minta nagyfelbontású becslései. A normalizációt úgy végezzük el, hogy egy 4 kHz-es jel +/- 1 bit-es amplitúdója feleljen meg közel 0 dB-nek SPL-ben. Míg egy full-scale szinuszjel közel 90 dB-nek felel meg.

A spektrumanalízis lépései a következők. A bemeneti audiomintákat, $s(n)$, normalizáljuk az FFT hosszának, N , megfelelően. Az analízis az alábbi képlet alapján történik, b -vel jelölve a mintánkénti bitszámot [7].

$$x(n) = \frac{s(n)}{N(2^{b-1})}$$

A normalizált inputot, $x(n)$, 12 ms-os szegmensekre bontjuk fel, így kapunk 512 db mintát. Ezekből a spektrális teljesítménysűrűség (PSD) becslését, $P(k)$, 512 pontos FFT-vel kapjuk meg a következő képpen az alábbi egyenlettel.

$$P(k) = PN + 10 \log_{10} \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi kn}{N}} \right|^2, 0 \leq k \leq \frac{N}{2}$$

Ahol a PN , teljesítmény normalizáció indexe, értéke 90,302 dB. A Hann ablak, $w(n)$, pedig a következő:

$$w(n) = \frac{1}{2} \left[1 - \cos \left(\frac{2\pi n}{N} \right) \right]$$

Mivel a lejátszás hangereje nem ismert a jel pszichoakusztikai analízisének, ezért a normalizáció és a PN paraméter a bemenet SPL-jének óvatos becslésére alkalmasak.

4.2 Tonális- és zaj maszkolók meghatározása

A PSD becslése és a hangnyomásszint-normalizálás után a következő lépés a maszkoló jelek meghatározása.

A Model-I szerint tonális maszkolók azok, amelyek olyan lokális maximumai a PSD-nek, melyek adott Bark távolságon belüli szomszédos komponenseknél 7 dB-lel nagyobbak. Ennek megvalósítása az S_T változóban az alábbi egyenletben olvasható.

$$S_T = \left\{ P(k) \left| \begin{array}{l} P(k) > P(k \pm 1) \\ P(k) > P(k \pm \Delta_k) + 7dB \end{array} \right. \right\},$$

ahol

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & (0,17 - 5,5 \text{ kHz}) \\ [2,3] & 63 \leq k < 127 & (5,5 - 11 \text{ kHz}) \\ [2,6] & 127 \leq k \leq 256 & (11 - 20 \text{ kHz}) \end{cases}$$

Az S_T -ben felsorolt spektrális csúcsokból számíthatóak a tonális maszkolók, $P_{TM}(k)$ (dB-ben értendő) az alábbi módon:

$$P_{TM}(k) = 10 \log_{10} \sum_{j=-1}^1 10^{0,1P(k+j)}$$

Tehát egy tonális maszkoló, az adott maximumból és három szomszédos spektrális komponensből kerül számításra. A későbbiekben az ábrákon a tonális maszkolókat 'x' jelöli.

A zaj maszkolók kiszámítása pedig a következőképpen történik. Minden kritikus sávban egy zaj maszkolót, $P_{NM}(\bar{k})$ számítunk ki a többi spektrális vonalból, amelyek nincsenek a $\pm\Delta_k$ környezetében a tonális maszkolónak.

$$P_{NM}(\bar{k}) = 10 \log_{10} \sum_j 10^{0,1P(j)} \text{ (dB)}, \quad \forall P(j) \notin \{P_{TM}(k, k \pm 1, k \pm \Delta_k)\},$$

ahol \bar{k} geometriai átlaga a kritikus sávban lévő spektrális vonalaknak, a következőképpen számítható ki:

$$\bar{k} = \left(\prod_{j=l}^u j \right)^{1/(l-u+1)}$$

l az alsó, u pedig a felső határindexe a kritikus sávnak.

A zaj maszkolókra azért van szükség, mert a kritikus sávban lévő maradék energia nem tartozik egyértelműen a tonális maszkolóhoz. Tehát egy adott kritikus sávban, az összes olyan spektrális komponens energiájából, mely nem járul hozzá egyik tonális maszkolóhoz sem, egy átlag számítódik, és ez lesz a zaj maszkoló. A későbbiekben az ábrákon a zaj maszkolókat 'o' jelöli.

4.3 Decimálás és a maszkolók újrendezése

A maszkolók számát két kritérium alapján csökkentjük. Ha bármely tonális vagy zaj maszkoló a hallásküszöb alá esik, akkor továbbiakban nem vesszük figyelembe. Tehát csak azokat tartjuk meg, amelyek megfelelnek az alábbi egyenlőtlenségnek.

$$P_{TM,NM}(k) \geq T_q(k)$$

ahol $T_q(k)$ -val jelöljük a hallásküszöb SPL-ben kifejezett értékét az adott k spektrális vonalnál.

A második feltétel megvalósításához egy 0,5 Bark széles csúszó ablakkal vizsgáljuk meg a maszkolókat. Minden olyan maszkoló közül, amelyek 0,5 Bark-nál közelebb esnek egymáshoz, csak az erősebbet tartjuk meg. Ezek után, az alábbi egyenletekben látható módon, újrendezzük a maszkolókat.

$$P_{TM,NM}(i) = P_{TM,NM}(k)$$

$$P_{TM,NM}(k) = 0,$$

ahol

$$i = \begin{cases} k, & 1 \leq k \leq 48 \\ k + (k \bmod 2), & 49 \leq k \leq 96 \\ k + 3 - ((k - 1) \bmod 4), & 97 \leq k \leq 232 \end{cases}$$

Az utóbbinak hatása a maszkolók 2:1 decimálása a 18-22.-ik kritikus sávokban, illetve 4:1 decimálás a 22-25.-ik kritikus sávokban anélkül, hogy maszkoló komponensek elvesznének. Ez a művelet csökkenti a tonális és zaj maszkolók számát, így egyszerűsítve a számítás további lépéseit.

4.4 Egyenkénti maszkoló küszöbök számítása

A decimált tonális és zaj maszkolókat használva következő lépésként ezekből egyenként maszkoló küszöbököt számítunk. Minden egyes küszöb reprezentálja az adott j frekvencia bin-nél elhelyezkedő maszkolónak az adott i frekvencia bin-re való hatását.

A tonális maszkolók küszöbszintjei a következő módon számítandóak [8].

$$T_{TM}(i, j) = P_{TM}(j) - 0,275Z_b(j) + SF(i, j) - 6,025(\text{dB SPL}),$$

ahol $P_{TM}(j)$ jelenti a tonális maszkoló hangnyomásszintjét az adott j bin-ben. $Z_b(j)$ jelenti a j . frekvenciabin Barkban kifejezett frekvenciáját. A maszkolás kiterjedését j maszkoló bin-től i maszkolt bin-re $SF(i, j)$ (spreading function) jellemzi, az alábbi módon.

$$SF(i, j) = \begin{cases} 17\Delta_{Z_b} - 0,4P_{TM}(j) + 11, & -3 \leq \Delta_{Z_b} < -1 \\ (0,4P_{TM}(j) + 6)\Delta_{Z_b}, & -1 \leq \Delta_{Z_b} < 0 \\ -17\Delta_{Z_b}, & 0 \leq \Delta_{Z_b} < 1 \\ (0,15P_{TM}(j) - 17)\Delta_{Z_b} - 0,15P_{TM}(j), & 1 \leq \Delta_{Z_b} < 8 \end{cases}$$

$SF(i, j)$ valósítja meg a 2.1.2 alfejezetben tárgyalt maszkológörbékét a tonális típusú maszkolókra. Korábban a 2.4-es ábrán ábrázolt görbéken jelent meg. A hang erősségének függvényében a maszkológörbe meredeksége változik.

A zaj maszkolók maszkológörbéi az alábbi módon számíthatóak.

$$T_{NM}(i, j) = P_{NM}(j) - 0,175Z_b(j) + SF(i, j) - 2,025(\text{dB SPL}),$$

ahol $P_{NM}(j)$ jelzi a zaj maszkoló hangnyomásszintjét az adott j bin-ben. $Z_b(j)$ jelenti a j . bin Barkban kifejezett frekvenciáját. $SF(i, j)$ pedig az előző számítás alapján kapható, úgy hogy $P_{NM}(j)$ -ket helyettesítünk a $P_{TM}(j)$ -k helyére.

4.5 A globális maszkoló görbe számítása

Ebben a lépésben összegezzük az egyes maszkolókhöz tartozó maszkológörbét, hogy egy becsült globális maszkológörbét hozhassunk létre minden frekvencia bin-hez. A model azzal a feltételezéssel él, hogy a maszkoló hatások additívak. Így a $T_g(i)$ globális maszkoló az alábbi módon számítható.

$$T_g(i) = 10 \log_{10} \left(10^{0,1T_q(i)} + \sum_{l=1}^L 10^{0,1T_{TM}(i,l)} + \sum_{m=1}^M 10^{0,1T_{NM}(i,m)} \right) (dB SPL)$$

ahol $T_q(i)$ a korábban kiszámított hallásiküszöb a adott i frekvencia bin-re. $T_{TM}(i, l)$ és $T_{NM}(i, m)$ pedig az egyenkénti maszkológörbék az előző lépésben számított módon. L és M pedig a tonális és zaj maszkolók száma rendre. Összességében, a maszkológörbe a hallóküszöb módosítását jelenti. Ez a módosítás az összes tonális és zaj maszkolótól függ, amely a jel spektrumában megtalálható.

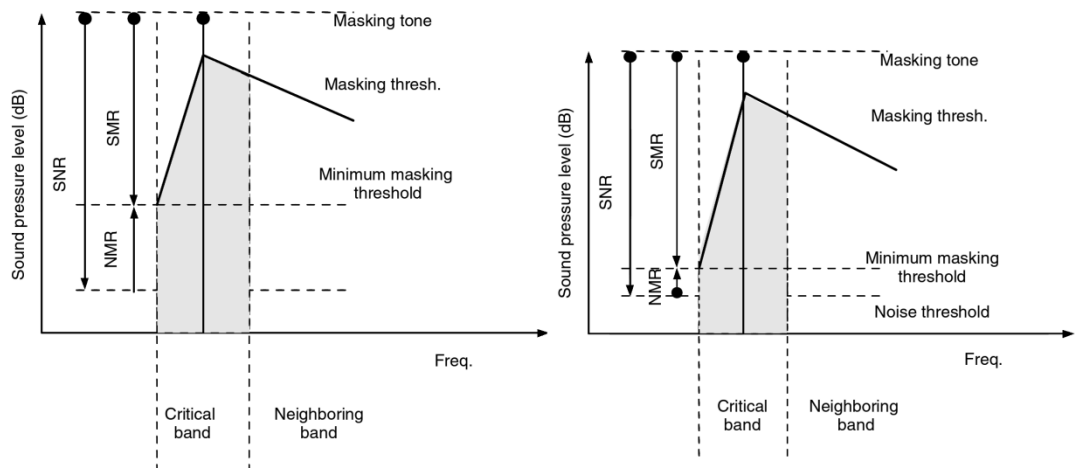
4.6 Érzékelésalapú bit-kiosztás

Az alábbi alfejezetben meghatározzuk az adott Bark-sávokhoz tartozó ideális bitszámot, ez előző lépések alapján számított globális maszkológörbe figyelembe vételével.

A jel-maszk arány (SMR, Signal to Mask Ratio) határozza meg az adott Bark-sávban a frekvenciatartománybeli minták kódolásához szükséges bitek számát. A zaj-maszk arány (NMR, Noise to Mask Ratio), úgy számítható ki, hogy kivonjuk az SNR jel-zaj viszonyból az SMR-t

$$NMR = SNR - SMR \quad [dB]$$

A fő célunk a kevesebb bit használatából adódó kvantálási zaj küszöbszint alatt tartása. Ez a küszöbszint a globális maszkológörbe adott szakasza az aktuális Bark sávban. Az alábbi 4.1 ábrán megfigyelhető az NMR számítása egy adott maszkoló szakaszhoz. A jobb oldali részábrán látható, hogy NMR értéke kisebb, így ebben a kritikus sávban kevesebb bit mellőzésére van lehetőség, mint a baloldali részábrán látható esetben, ahol NRM értéke nagyobb.



4.1. ábra NMR számítása különböző maszkológörbe esetén [8]

5 Audiojelek tömörítése a MODEL-1 felhasználásával

A 4. fejezetben leírt lépések alapján Matlab környezetben implementálható egy program mely képes a Model-I alapján kódolni hangminta szeleteket. Jelen fejezet a Model-I megvalósítását mutatja be, különböző tesztjeleken vizsgálva a modellt.

5.1 Szinusz-jel elemzése



5.1. ábra. Szinusz jel kódolásának első két lépése. Hallásküszöb és maszkolók

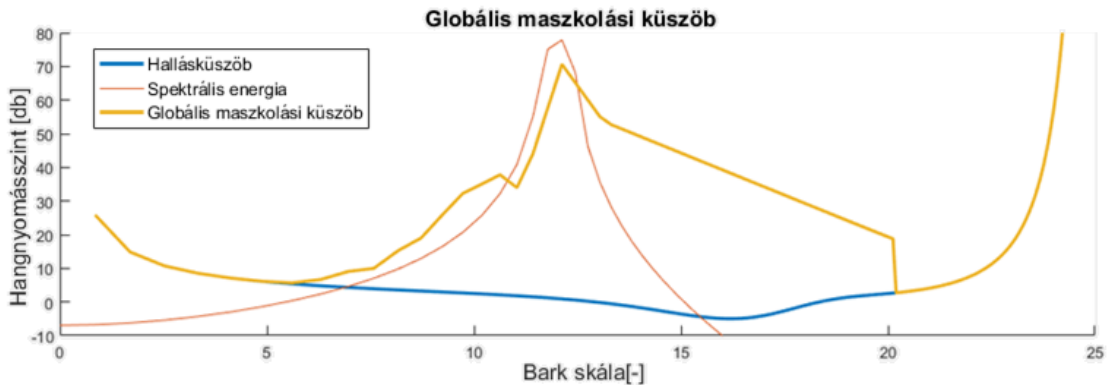
Az 5.1-es ábrán megfigyelhető a kódolás első két lépése. A vízszintes tengelyen Bark-skálán ábrázoljuk a frekvenciát, a függőlegesen pedig a hangnyomásszint olvasható le dB-ben. A kék vonal ábrázolja a hallásküszöböt, amelyet a model használ. Ezen az ábrán példaként egy szinusz jel látható. A vékonyabb piros vonal jelzi a jelenergiát. A piros kereszt és a fekete pontok a tonális és zaj maszkolók rendre.



5.2. ábra. Maszkolók kiválasztása és az egyéni maszkológörbék

Az 5.2-es ábrán szintén az előző jel látható. Megfigyelhetjük, hogy a kódoló a hallásküszöb alatti zaj maszkolókat a továbbiakban nem veszi figyelembe. Az összes

maszkolóhoz rendelt maszkológörbék is láthatóak továbbá. A piros görbe tartozik a tonális maszkolóhoz.



5.3. ábra. A szinusz jel globális maszkológörbéje

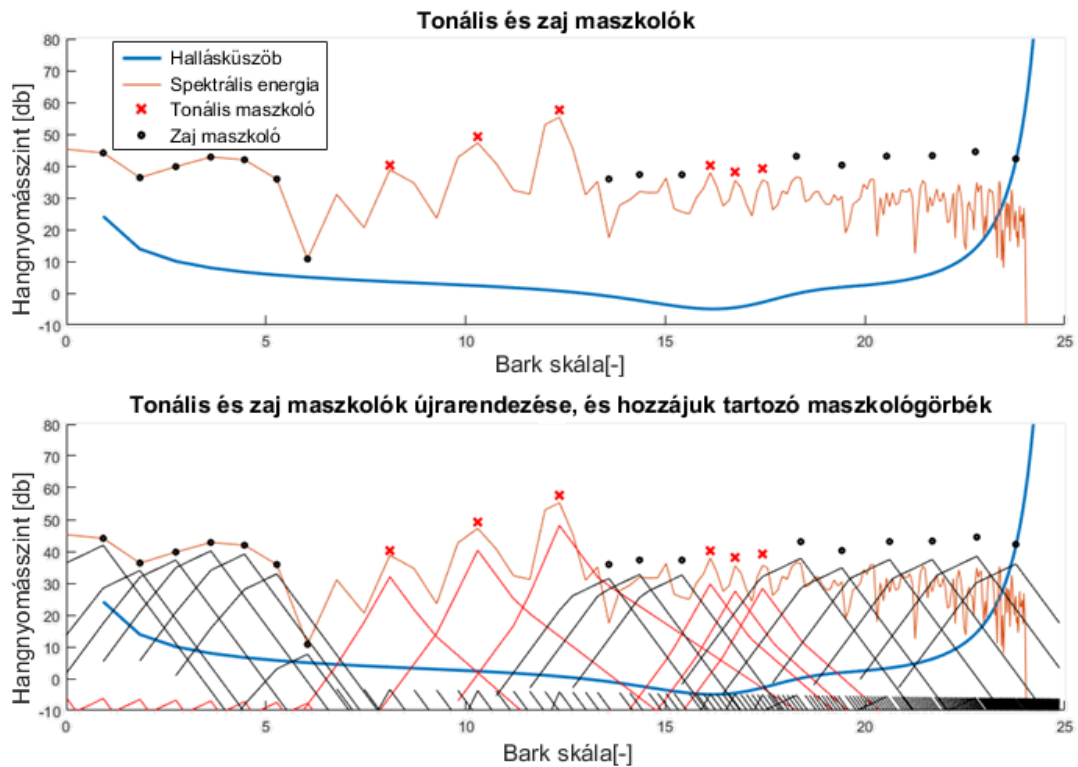
Az 5.3. ábrán az összegzett globális maszkológörbe látható sárga vonallal ábrázolva. Megfigyelhető, hogy azokon a részeken, ahol nincsenek egyedi maszkológörbék, mivel a korábbi fejezetben említett hallásküszöböt is hozzáadja, a két görbe együtt fut.



5.4. ábra. A globális maszkológörbéből számított bitnyereség szinuszjel esetén

Az 5.4. ábra a bitnyereséget ábrázolja kék vonallal. Megfigyelhető, hogy a görbe alakja követi az előző 5.3. ábrán ábrázolt globális maszkológörbe alakját. A szinusz jel frekvenciáján és környékén akár 9 bittel kevesebben is kódolhatunk, a Model-I szerint az emberi hallás nem veszi észre a különbséget.

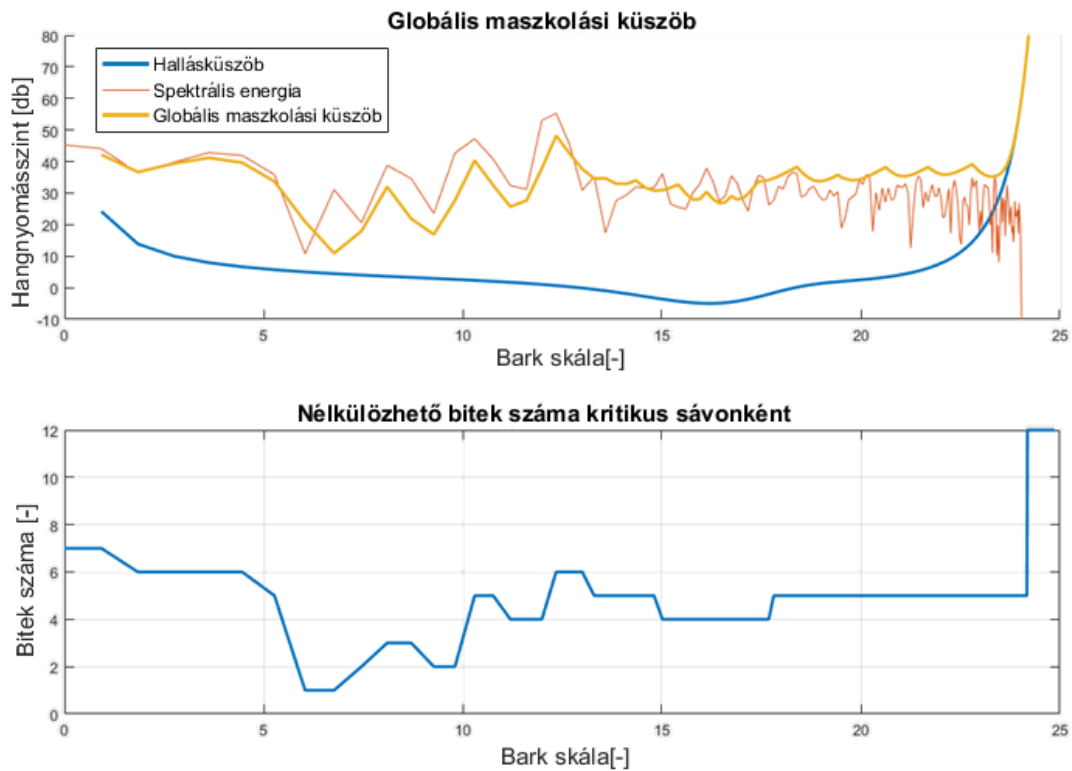
5.2 A kódoló bemutatása zenei hangminta egy szeletére



5.5. ábra. Zenei minta egy időszelétére elvégzett kódolás első három lépése

Az előző alponthoz hasonlóan, ebben az alponthoz is ugyanazokat a lépéseket végezzük el, ezúttal egy zene file egy időszelétére. Az 5.5. ábrán az 5.1. és 5.2. ábrákhoz hasonlóan a kiválasztott maszkolókat, azok szelektálását és a hozzájuk rendelt egyedi maszkológörbéket láthatjuk. A piros kereszt és a fekete pont továbbra is a tonális és zaj maszkolókat ábrázolja rendre. Megfigyelhető a tonális és zaj maszkológörbék közötti jellegbeli különbség is az alsó ábrán.

Az alábbi 5.6. ábrán a kódolás további lépései láthatóak. Az 5.3. ábrához hasonlóan a sárga vonal a globális maszkológörbét jelzi. Megfigyelhető, hogy zene esetén sokkal összetettebb a görbe. Az ábra alsó felében pedig a Bark-sávonként számított maximum megengedhető bitelhagyást ábrázolja a kék vonal. Ebben a példában maximum 6 bit hagyható el néhány frekvencián.



5.6. ábra. Zene időszeletének globális maszkológörbéje és számított Bark-sávonkénti bitnyereség

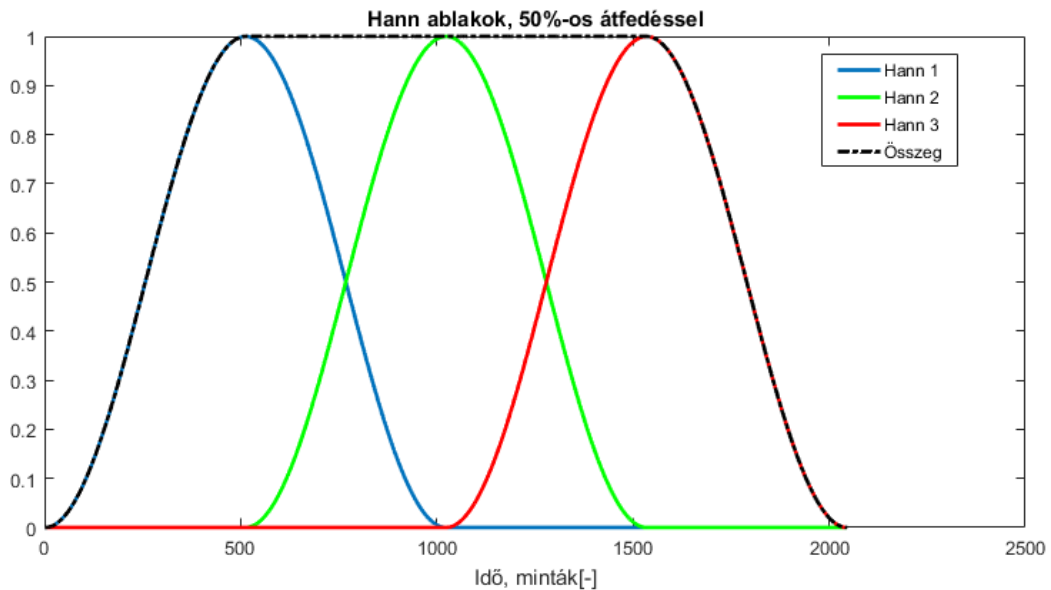
5.3 Több időszelét kódolása

5.3.1 Hann-ablak

Ahhoz, hogy egy teljes audio-file-t kódolni tudjunk, ahhoz tovább kell fejleszteni a kódolót. A kódolandó audioanyagot időben ablakozással bontjuk fel. Erre megfelelő például a Hann-ablak. Az ablak súlyozza az értékeket attól függően, hogy az ablak melyik részén található. Az ablak közepén 100%-ban megtartja a tartalmát, szélei felé pedig folyamatosan csökken nulláig. Ha ilyen Hann ablakokat alkalmazunk 50%-os átfedésben, akkor adatot nem veszítünk, hiszen a szomszédos ablakkal összeadjuk az értéket, és együtt már kiteszik az egészet. Ez látható az alábbi 5.7. ábrán.

Az ablakozásra azért van szükség, mert ha átfedés nélküli időszelleteket használnánk, akkor kódolás után összeillesztésükkor jelben ugrások alakulnának ki, ha nem illeszkedik az előző ablakban található utolsó időbeli minta az aktuális ablak első mintájához. Mivel a két ablakban különböző bitszámokkal kódolhatunk, lényegében minden keretnél ez az eset áll fenn. Az ablakok közti átlapolás a két ablak közti átmenetet simítja ki.

Hann-ablakokkal készített időszeletek sokaságát a kódoló képes kódolni, majd az átlapolást az összeillesztéskor a dekódoló veszi figyelembe.



5.7. ábra. 50%-os átfedésű Hann-ablakok és összegük

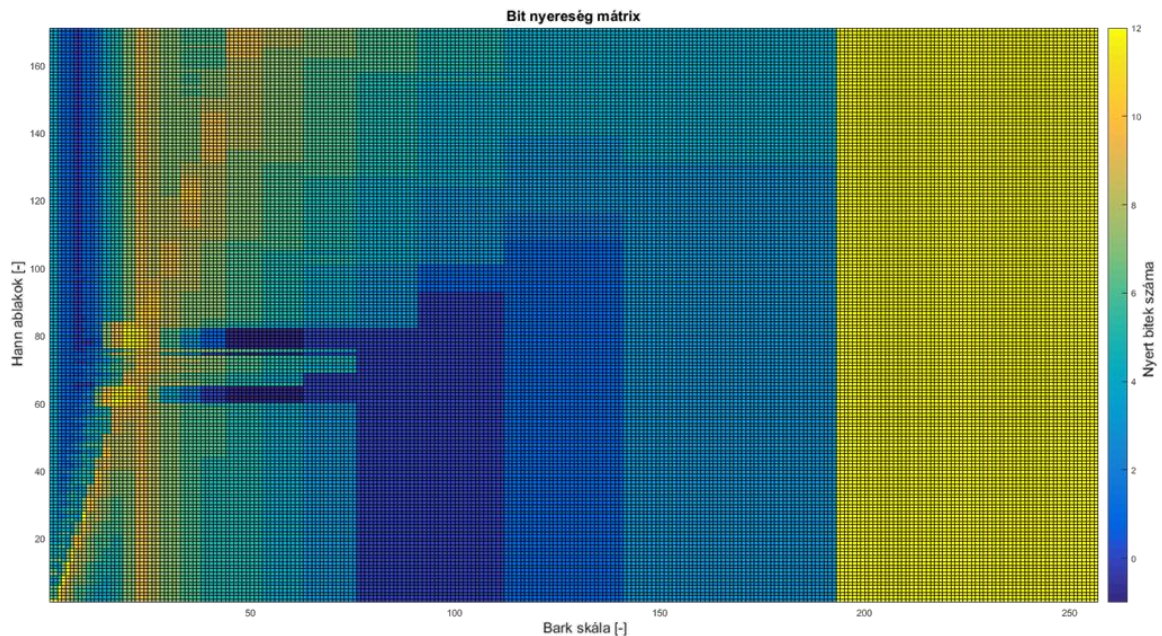
5.3.2 Hosszabb hangminták kódolása

Az alábbi 5.8. ábrán az előzőekben leírt hosszabb hangminta kódolásának megvalósítására láthatunk példát. A nélkülözhető bitek számát ábrázoló grafikon helyett most egy színezett mátrix látható. A vízszintes tengelyen továbbra is Bark skálán ábrázolt frekvencia található. A függőleges skála pedig az időt mutatja ablakokban mérve. A színezés szerint a világosabb színek jelentik a kevesebb biten lehetséges kódolást. A kódolást egy szinusz és egy sepert szinusz összegére végezzük el, ezek azonos amplitúdójúak.

Az ábrán megfigyelhető, hogy nagyfrekvencián, ahol a hallásküszöb jóval magasabb a sávközépi hallásküszöbnél, az összes bitet eldobja a kódoló, ezért ez a terület sárga. Az állandó frekvenciájú szinusz a függőleges narancssárga sáv. Látható hogy nagyobb frekvenciák irányában kicsit szélesebb a maszkolás hatása. A sepert szinusz esetén is szélesebb a maszkolás a nagyobb frekvenciák irányában, viszont az is megfigyelhető, hogy a maszkolás mértéke a frekvenciával együtt nő.

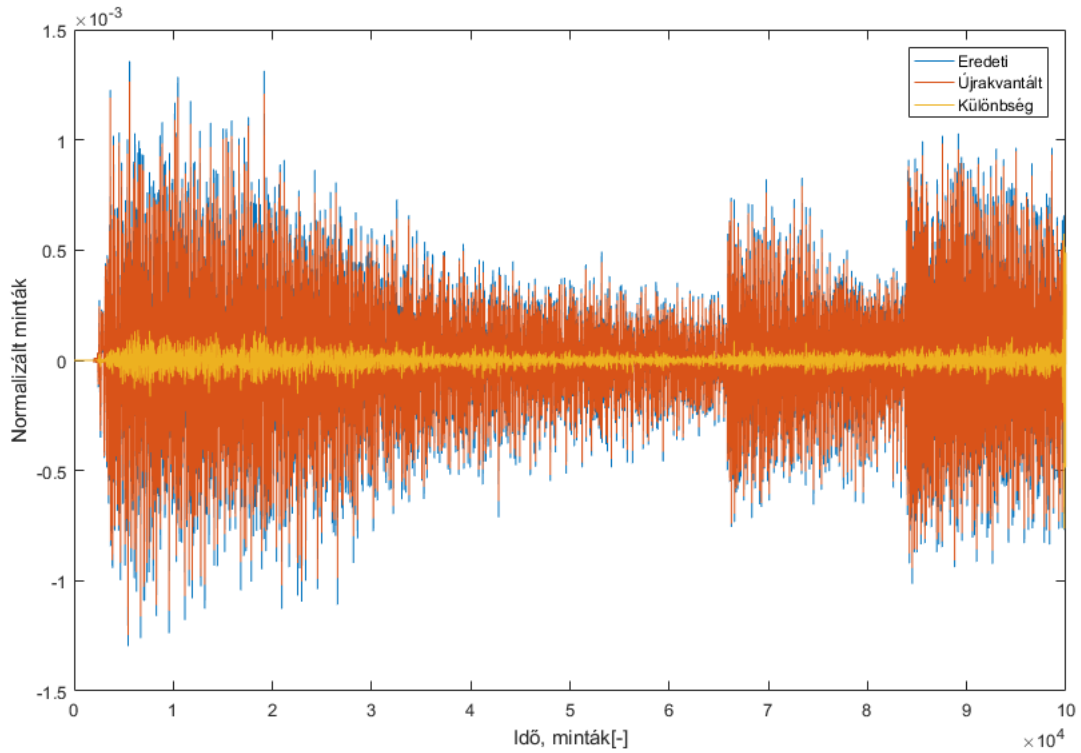
Továbbá, ahol a sepert szinusz eléri a másik szinusz frekvenciáját, ott az ábrán keresztezik egymást. Itt megfigyelhető, hogy jelentősen világosabb lesz a terület, vagyis a két szinusz maszkológörbéje összegződik, és itt több bit elhagyása lehetséges. Ettől a területtől kicsit jobbra, nagyobb frekvenciák irányába, megfigyelhető egy sötét terület,

ahol látszólag kevesebb, mint nulla bit elhagyása lehetséges. Ez a hiba abból fakad, hogy a frekvenciákhoz és hangnyomásszintekhez definiált maszkoló görbék nem teljesek nagy hangnyomásszint esetén. Ez korábban is megfigyelhető az 5.2. és 5.3. ábrákon, ezeken is egy ugrás látható, itt is ez történik.



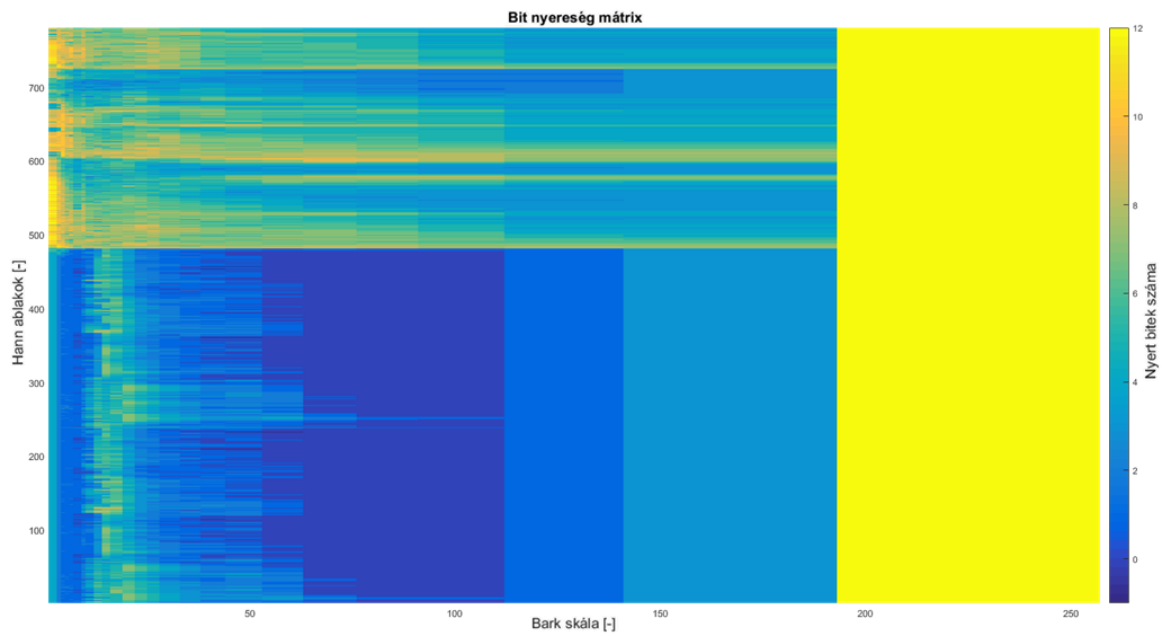
5.8. ábra. Bitnyereséget ábrázoló mátrix színekkel megjelentve. Szinusz és sepeert szinusz összegére elvégzett kódolás után

Az alábbi 5.9. ábrán egy gitáros zongorás hangminta kódolását ábrázoltuk. Vízszintes tengelyen időtartományban, függőleges tengelyen pedig a normalizált minták olvashatóak le. Kékkel ábrázolt az eredeti hangminta, amelyet szinte teljes mértékben eltakar a narancssárga kódolt hangminta, néhol látszanak kék részek, ebből is tudhatjuk, hogy a kódoló valóban tömörítést végez. Továbbá sárgával a két minta különbsége is látható, így jobban szemléltetve a kettő közötti különbséget.



5.9. ábra. Zene kódolása, eredeti, kódolt és a kettő közötti különbség az időtartományban

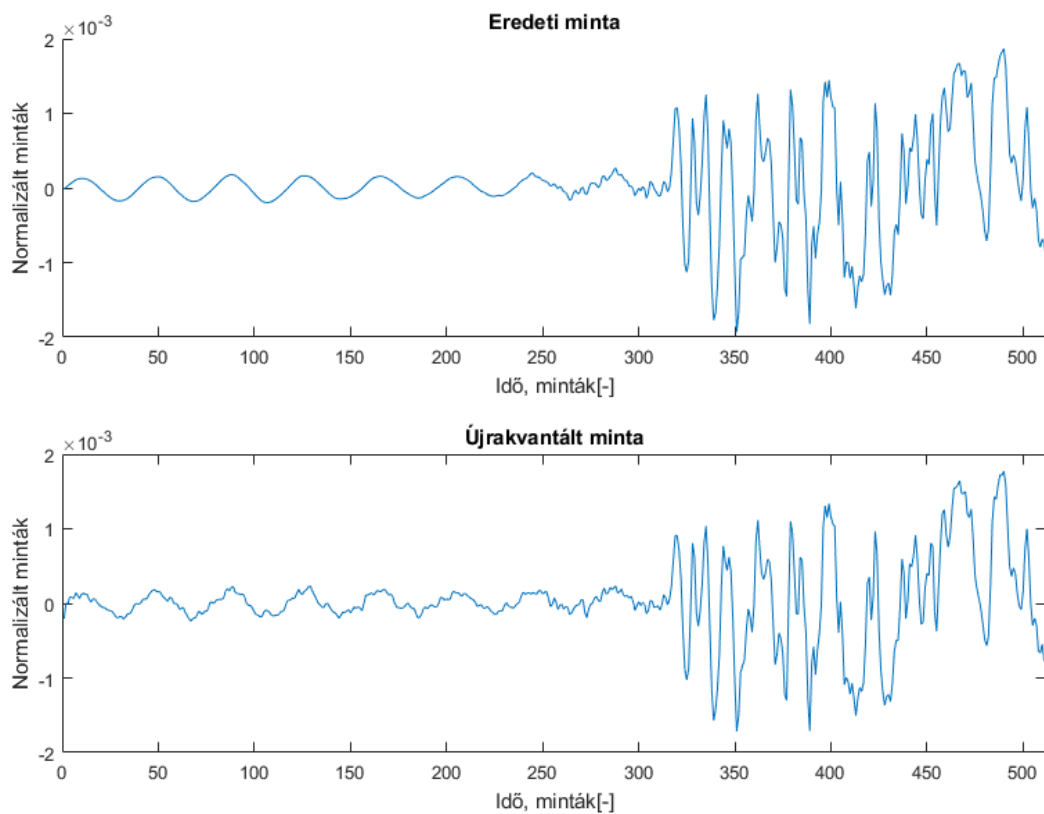
5.3.3 Hirtelen tranziensek hatása



5.10. ábra. Bitnyereség mátrix, zenére, amely hirtelen változást tartalmaz

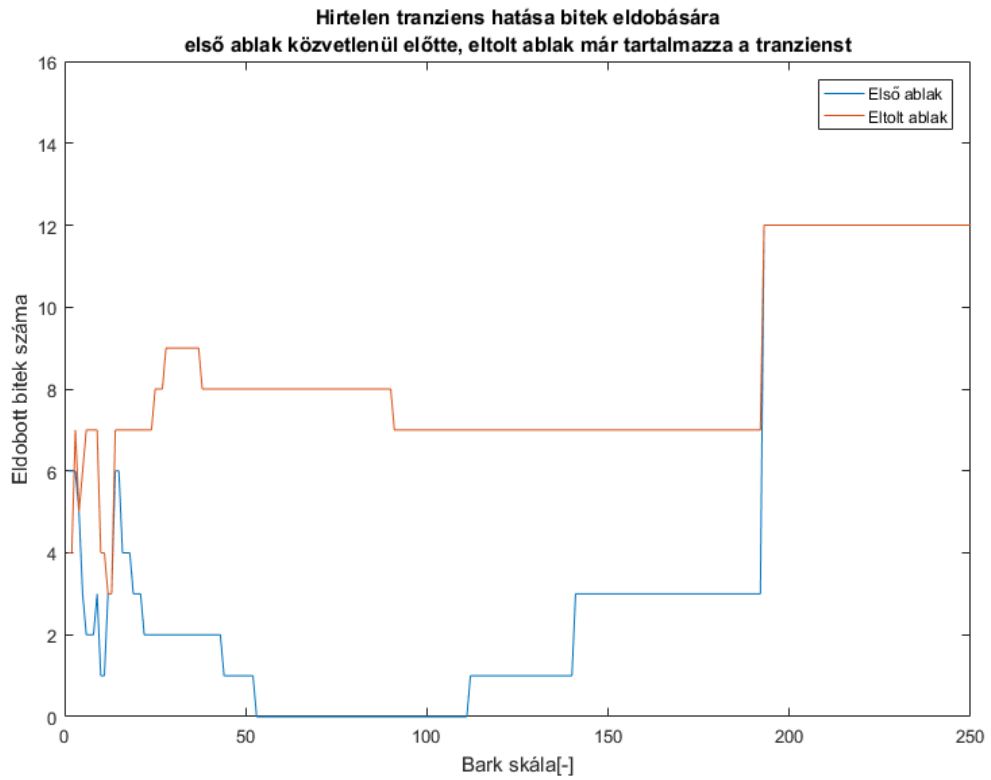
A fentü 5.10. ábrán egy olyan hangminta bitnyereség mátrixa látható, amely egy hirtelen változást tartalmaz, valamivel az 500. ablak előtt. Ebben a hangmintában az első pár másodpercben füttyülés szól, majd egy erős dobütéssel kezdődik a zene.

Ez az alábbi 5.11. ábrán látható módon okoz problémát. A kvantálás által okozott zaj az egész ablakra kiterjed, ez pedig jól megfigyelhető az alsó ábrán, amely a kódolt mintákat ábrázolja. Egy közel szinuszos jelre ráül a kvantálási zaj. Ha a következő ablak pont a hirtelen változás előtt kezdődik, akkor ez a jelenség nem figyelhető meg. Viszont nehéz előre tudni, hogy ez mekkora ablakmérettel lehetséges. Ezért tranziens detekcióval változó méretű ablakokkal küszöbölhető ki ez a hatás.



5.11. ábra. Eredeti és kódolt minta hirtelen változás esetén

Az alábbi 5.12. ábrán, ugyanazon hangminta egy-egy szeletének frekvenciánként elhagyható bitmennyiségeit ábrázolja. A késsel ábrázolt vonal egy olyan ablakhoz tartozik, mely véget ér a zenében található hirtelen változás előtt. Ezen elvégezve az analízist, azt kapjuk, hogy bizonyos frekvenciákon nagyon kevés vagy egy bit sem hagyható el. Míg ha ezt az ablakot úgy választjuk meg, hogy tartalmazza a változást – ezt jelöli az ábra pirossal – akkor ugyanezen a frekvencia-szakaszokon lényegesen több bitet számol elhagyhatónak a kódoló. Ebből következik az előző 5.11. ábrán is megfigyelt kvantálási zaj észrevehetősége. A probléma megoldásának megvalósítását egy későbbi fejezetben részletezem.



5.12. ábra. Hirtelen változást tartalmazó hangminta két ablakából számolt elhagyható bitek száma. Az első ablak véget ér a hirtelen változás előtt, még a második tartalmazza azt.

6 Eredeti és kódolt hangminták objektív összehasonlítása PEAQ modell segítségével

Eddigi eredményeket csak szubjektíven ellenőrizhetjük. Meghallgatjuk az eredeti és a kódolt hangmintát, észlelhető-e a hangminőség romlása, mennyire zavaró a kódolásból adódó zajhatás és torzítás. Ezt hasonlíthatjuk a tömörítés mértékéhez, de ez sem feltétlenül mond sokat. Ezért szükség van egy objektív összehasonlító, elemző módszerre. Ez lesz a PEAQ (Perceptual Evaluation of Audio Quality – hanganyag minőség érzeti értékelése). A PEAQ két hangminta összehasonlítására képes, melyre kiszámít egy ODG (Objective Difference Grade – osztályzat objektív különbségre) mérőszámot.

A veszteséges audiotömörítő eljárások lényege, hogy kisebb bitsebesség mellett ne halljuk meg a hanganyag hiányosságait. Tehát egyszerűen meghallgatva nem feltétlenül kapunk választ arra a kérdésre, hogy jobb-e a kódolás. Az emberi hallás szubjektív. Továbbá a szubjektív meghallgatásos tesztek elvégzése időigényes. Ezt figyelembe véve egy objektív elemzőre van szükség, ez lesz a PEAQ. Ezt az objektív elemzőt nem könnyű elkészíteni, ugyanis egy egyszerű különbség számítás nem elég a hiba meghatározására, hiszen ez vagy ehhez hasonló módszerek nem vennék figyelembe az emberi hallás tulajdonságait, a hangérzékelés folyamatának komplexitását. A megfelelő mérőszám kiszámításához szükséges figyelembe venni, hogy különböző hatások, változások különböző módon és eltérő mértékben befolyásolják az érzett minőséget.

6.1 PEAQ kialakulásának története

A PEAQ kialakulását három korábbi modell előzte meg. Ezt követte egy versengő szakasz, amelyben különböző cégek próbáltak kialakítani egy megfelelő modellt, amely több számításra képes, mint a korábbi három kezdetleges modell. Végül pedig egy együttműködő szakasz segítségével jött létre a megfelelő rendszer. A továbbiakban ezeket a modelleket és szakaszokat jellemzem.

Az első objektív minőség mérő módszer az NL (Noise Loudness – zaj hangosság) nevű rendszer. Ez a módszer az érzékelt zaj hangosságát mérte, mely zajt

egy beszédkódoló algoritmus ad hozzá a mintához. A két beszéd minta közötti különbség maga a zaj, ezt kb. 20 milliszekundumos időkeretekre határozták meg. A zaj érzékelt hangossága függ attól, hogy milyen mértékben maszkolja azt a beszéd minta. Ezeket a maszkoló határokat kísérletek alapján számították ki. A teljes beszéd minta minőségére ez a módszer nem számított mérőszámot, csak időkeretenként. []

A második módszer az előző ötleteire alapozva az ASD (Auditory Spectral Difference – hallás alapú spektrális különbség) nevű rendszer. Az időkeretek helyett egy átlapolt szűrőrendszer szolgáltatja az vizsgálendő egységeket. Hozzávettek egy modellt az időbeli maszkolás számításba vételéhez. Továbbá módosították azt a módszert mely kiszámítja az abszolút küszöböt. Mindkét vizsgálendő jelet ugyanúgy dolgozták fel, ezekből készült belső reprezentációkat hasonlították össze. Ezúttal sem készül el a teljes minta minőségére vonatkozó átfogó értékelés. Ez a módszer már jobban közelíti az emberi hallás által hallott különbségeket, de az algoritmus jelentősen bonyolultabb lett.

A harmadik módszer NMR (Noise to Mask Ratio – zaj maszkolásra vett aránya) a tömörítő audiokódolás segítésére készült el. NL-hez képest csökkent a komplexitás, és a maszkoló komponensek worst-case számítását alkalmazta. Az előzőekkel ellentétben NMR már próbát tett a minta teljes egészére vonatkozó érzékelt hangminőség meghatározására, számszerűsítésére.

Az objektív minőség meghatározásának megbízhatóságát, csak szubjektív meghallgatásos tesztekkel összehasonlítva kaphatjuk meg. Az ITU-R ajánlása egy olyan módszer mely leír egy szubjektív döntéseken alapuló számítást egy eszköz általános audiomínőségének meghatározására. A PEAQ kifejlesztésének során a cél az volt, hogy meghatározhassuk az általános audiomínőséget objektív mérési metódusok segítségével.

Az ITU-R ajánlása definiál módszereket audiojelek kis romlásának mérésére, és ezzel együtt az adat megfelelő szubjektív vizsgálatára is. A vizsgálat során a hallgató bármelyik audioforrást meghallgathatja, melyek rendre A, B vagy C. Ezek közül A a referencia-jel. Viszont B és C közül az egyik szintén a referencia-jel, míg a másik a teszt-jel, hogy melyik a referencia- és melyik a teszt-jel az véletlenszerű az adott vizsgálat során. Betanítás után a hallgatónak értékelnie kell a B és C jelet a referencia jelhez képest. Vagy B vagy C jel nem mutat különbséget az A jelhez képest, míg a másik minden kis különbsége romlásnak tudható be. Ebből kifolyólag a hallgató által adott értékelés egy globális tulajdonság, amely az összes érzékelhető különbséget mutatja. A hallgató által adott osztályzatok különbségéből számítják az SDG számot, ez a

szubjektív különbség osztályzat. Ha a hallgató helyesen azonosítja a teszt és referencia jeleket, akkor SDG értéke negatív. A nulla érték pedig a megkülönböztethetetlen jeleket jelenti. A mínusz négy érték esetén a különbség pedig nagyon zavaró mértékű.

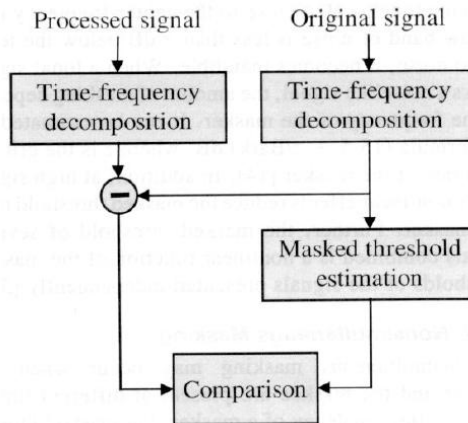
1994-ben az ITU-R kezdeményezte, hogy létrehozzanak egy olyan módszert, amely képes objektív mérőszám számítására. Hét modell közül kellett választani. A modellek vizsgálata során a közepes és jó minőségű audiojelek közötti különbség tétele volt a főszempont. Tehát az objektív minőségmérő pontosságát az ITU-R ajánlásához alkalmazkodó hallgatói teszteken keresztül állapíthatták meg. A vizsgálatok után a hét modell eredményei között nem találtak jelentős különbséget, illetve egyik sem érte el a felhasználók által meghatározott igényeket. Ebből kifolyólag együtt fejlesztettek ki egy jobb modellt, a cél pedig egy korábbi létező modell teljesítményének meghaladása. Ezzel a versenyző szakaszt az együttműködő szakasz váltotta fel.

Az együttműködő szakasz célja, a különböző módszerek legjobb részeinek felhasználásával egy új módszer létrehozása. Továbbá két verzió fejlesztése a cél, amelyek legjobban kiszolgálják a felhasználók igényeit. Az egyik valós idejű elemzésre alkalmas, míg a másik komplexebb, de megbízhatóbb számítást végez.

6.1.1 Érzeti mérési elvek

Az érzeti mérési módszerek között két fő elv létezik a hang változásának becslésére [10]. A maszkoló küszöbtől való távolság számítása és az ún. belső ábrázolás összehasonlítása. Ezen felül figyelembe vehetünk egy harmadik elvet, amely azon alapul, hogy néhány hatást egyszerűbb lineáris spektrumon modellezni a baziláris membrán feletti ábrázolás helyett, ez a hibák spektrális analízise elv.

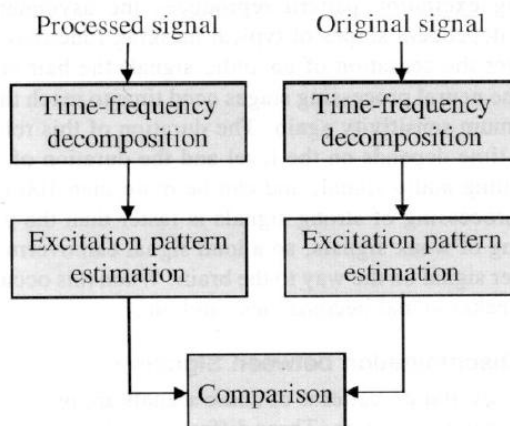
A maszkoló küszöbtől való távolság számításának menete az alábbi 6.1. ábrán látható. Egy hiba észlelhetősége akkor nulla, ha a nagysága nem haladja meg az előre definiált maszkoló küszöböt. Ennek a megközelítésnek legnagyobb előnye, hogy a modell paraméterei, maszkoló küszöb függvényei, közvetlenül meghatározhatóak kísérletek alapján.



6.1. ábra. Maszkoló küszöbtől való távolság számításának elve [10]

Az eredeti és feldolgozott jel idő-frekvencia felbontása után különbségüket képezzük. Ezt a különbséget hasonlítjuk össze az eredeti jelből becsült maszkoló küszöbvel. Az ábrán bal oldalt a feldolgozott jobb oldalt pedig az eredeti jel található.

A belső ábrázolás összehasonlítása a baziláris membrán gerjesztésre való reakciójának modellezésén alapul. Ebből úgy kapunk minőséget jellemző számokat, hogy összehasonlítjuk az eredeti és feldolgozott hangminta által létrehozott gerjesztési mintákat. Ez a megközelítés lényegesen közelebb áll az emberi hallórendszer érzékeléséhez, mint a maszkoló küszöb elmélete. Tehát ez egy jobb alap az összetettebb hallási jelenségek modellezéséhez. Ennek a folyamatnak a blokkvázlatát az alábbi 6.2. ábrán láthatjuk.



6.2. ábra. Belső ábrázolás összehasonlítása elv [10]

Az idő-frekvencia felbontás után egy gerjesztés-minta becslést számít a rendszer, mind az eredeti és a feldolgozott jelre, majd ezeket hasonlítja össze.

A hibák spektrális analízise elvére azért van szükség, mert néhány hatást, például az alapfrekvencia és a harmonikusok érzékelését, egyszerűbb modellezni lineáris spektrum használatával, mint a baziláris membrán gerjesztési mintái alapján. Jobb becsléseket kapunk az objektív érzeti minőségre, ha a lineáris spektrumon alapuló vizsgálatot kombináljuk ez előző két elvvel.

6.1.2 Fül-modell

A korábbi 2.1. alfejezetben tárgyalt tulajdonságait a hallásnak ebben a modellben is figyelembe vesszük. A hallás-küszöb frekvencia-függősége, a baziláris membrán tehetetlenségéből adódó elfedési jelenségek és a nemlineáris frekvencia érzékelés. Ezen felül a Bark-skála használata is fontos a modell szempontjából.

6.1.3 Észlelési modell

A fül által létrehozott érzeti képzethez társul egy észlelési képzet, amelyet nehezebb specifikálni. Ennek ellenére lehetnek feltevéseink arra, hogy az észlelési képzet hogyan befolyásolja a minőség megítélését. Például ha a hallgató nem ismeri a feldolgozott jelet, akkor a véleményét jelentősen befolyásolja a tudása. Lehet mindenkinek más véleménye arról, hogy mi is a jó zongorahang, de mindnyájuknak van egy általános tudása a zongorahangot illetően. Ha a hullámforma torzítást szenved, akkor mindenki egyetért abban, hogy rossz minőségű a felvétel.

Mivel a számítógépeknek nincsen ilyen világképük, ezért csak az eredeti hanghoz tudják hasonlítani a feldolgozott hangot. Ez akkor lehet probléma, ha az eredetinek zajjellegzetességei vannak, ekkor a gép lehet, hogy jobbnak osztályozza a feldolgozott hangot, míg egy valós hallgató rosszabbnak érezné. Ezt a szempontot kifejezetten nehéz modellezni, hiszen nem tudjuk, mi van a hallgató fejében az ideális hangképéről.

Ezen felül az is különböző hatással van a hallgatóra, hogy bizonyos frekvenciákon információt veszünk, vagy hozzáadunk. Az adatok hiánya kevésbé zavaró, mint azok hozzáadása. Tanulás alapú észlelési jelenségnek nevezzük azt az esetet, amikor egy ismeretlen torzítást kevésbé tud a hallgató kivenni, mint egy ismerős torzítást. Ez a hatás az információ alapú maszkolás.

Végül azt is érdemes megjegyezni, hogy az audiojel némely része több információt hordozhat, ezért ez lehet fontosabb, amikor felmérjük a torzításokat.

6.2 PEAQ

Az együttműködő fejlesztési szakasz eredménye a PEAQ. A PEAQ fül modellje FFT-n és egy szűrősoron alapul. A modell kimenete 6.1.1. érzeti mérési elvek nevű alponthban tárgyalt mindhárom elven alapul. A megfelelő kimeneti értékekből számolódik egyetlen mutató egy neurális hálózaton keresztül.

6.2.1 Analízisre való előkészítés

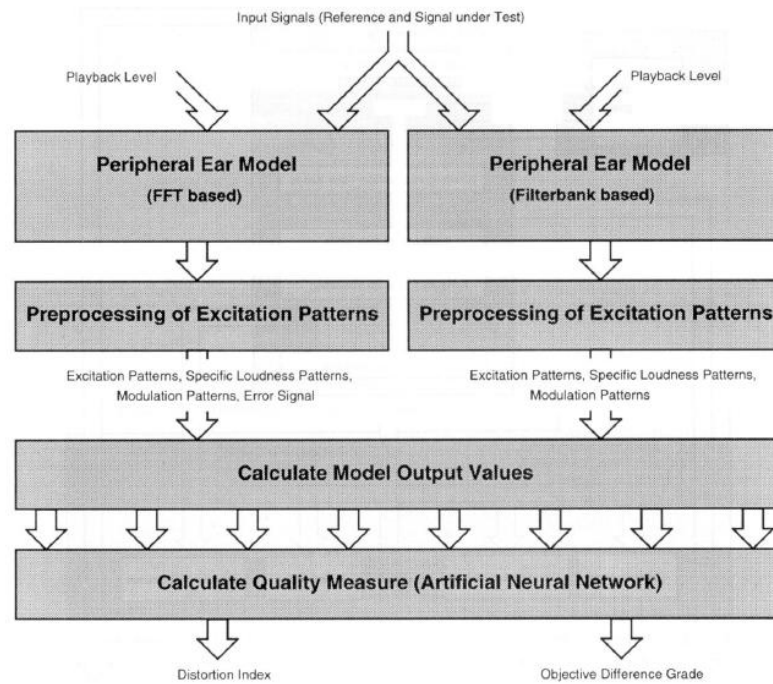
A modellt hozzá kell igazítani a teszt jel visszajátszási szintjéhez. Az adott L_{MAX} hangnyomásszinthez egy skálázási faktort számol a modell a bemenet számára, úgy hogy az 1 értékű amplitúdónak 0 dB feleljen meg, az adott L_{MAX} érték pedig 85–100 dB tartományban található. Ha nem ismert a visszajátszás szintje, akkor ezt 92 dB-re kell állítani.

Az időbeli összehangolás nem része a modellnek, tehát az eredeti és feldolgozott jeleket időben össze kell hangolni, mielőtt a mérőmódszerrel megvizsgáljuk a minőséget.

6.2.2 PEAQ verziók

A modellnek két verziója van. Az egyik olyan felhasználásra alkalmas, amely gyors feldolgozást igényel, így kisebb a komplexitása, ez a „basic version”. A másik pedig a legnagyobb elérhető pontossággal való számítást igénylő alkalmazások számára készített „advanced version”.

A basic verzió csak az FFT alapú fül modellt használja, és alkalmazza mind a két érzet mérési elvet. Az FFT alapú fül modellből következő rossz ideiglenes felbontás miatti korlátokat több kimeneti változóval és megnövelt spektrális felbontással kompenzálja ez a verzió. Ez a spektrális felbontás 0,25 Bark szélességű sávokat jelent. Ezzel szemben az advanced verzió az FFT-alapú fül modellen kívül a szűrősor alapú fül modellt is használja. A maszkoló küszöb elve az FFT-alapú modellen keresztül kerül alkalmazásra, a belső ábrázolás összehasonlítása pedig a szűrősor alapú modellen keresztül lesz figyelembe véve. Ennek a verziónak a spektrális felbontása pedig 0,5 Bark szélességű sávokból áll.

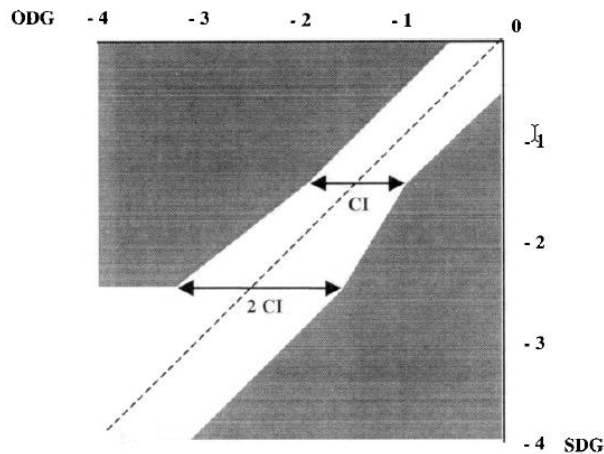


6.3. ábra. A mérés rendszerének blokk diagramja [10]

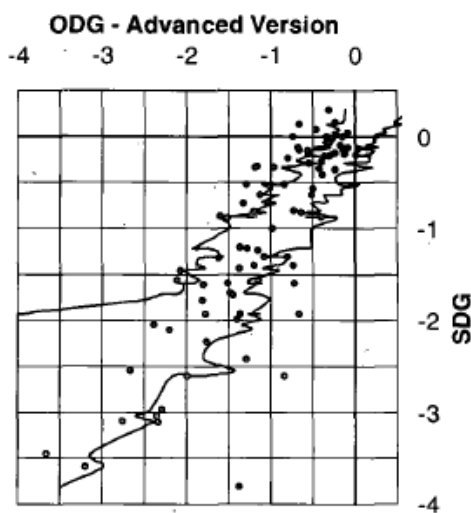
Az fenti 6.3. ábrán található a mérés lépéseinek sorrendje. A bemeneti jelek mellett a visszajátszás szintjét is megadjuk. FFT és szűrősor alapú fül modell számol, majd a gerjesztés minták elő-feldolgozása következik. A kiszámolt gerjesztés mintákból, jellegzetes hangosság mintákból és moduláció mintákból, az FFT alapú modell esetén hiba jelből is, kiszámoljuk a modell kimeneti értékeit. Ezekből az értékekből tudunk számolni minőség-mérőszámokat egy neurális hálóval. A végső kimenetek a torzítási index és az ODG mérőszám.

Az alábbi jellemzőket használjuk fel a hang minőségének becslésére: burkoló-moduláció, modulációs különbség, részleges zaj hangosság, hallható lineáris torzítás, zaj-maszk arány, jel sáv szélesség, detektálás valószínűsége és harmonikus struktúra hiba. Ezeket az értékeket a hallgatók által meghatározott minőség-értékeléssel összekötő függvényt a hallgatói tesztekben kapott SDG adatokkal hitelesítették. Így megkapjuk az ODG (objektív) mérőszámot, amely így kapcsolatban áll az SDG (szubjektív) mérőszám értékekkel.

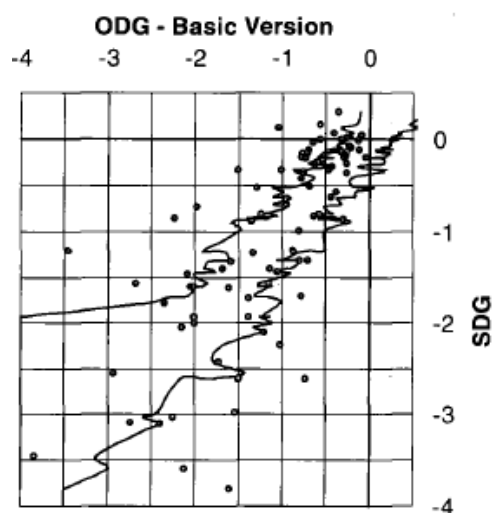
Az alábbi 6.4. ábra azt ábrázolja, hogy mekkora eltérés engedett meg az adott SDG értéktől. Megfigyelhető, hogy kisebb, tehát rosszabb minőséget jelző, SDG értéktől nagyobb eltérés is megengedett.



6.4. ábra. Az ODG megengedett eltérése SDG-hez képest [10]

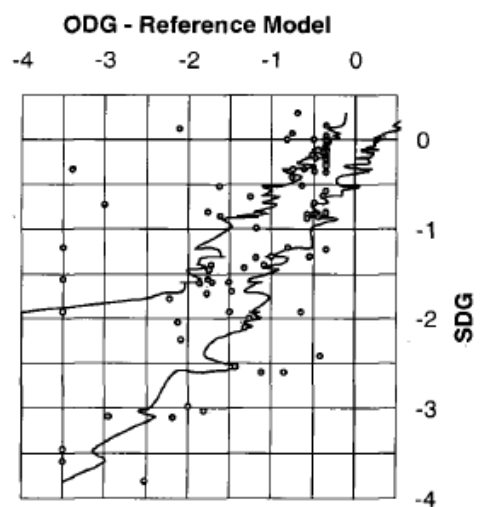


6.5. ábra. Advanced Version eredményei adott hangmintára [10]



6.6. ábra. Basic Version eredményei adott hangmintára [10]

A 6.5., 6.6. és 6.7. ábrákon azonos hangminta és kódolt párjára végeztek összehasonlítást az új modell két verziójával és a korábbi modellel. A vonalak jelzik a tolerancia határát az adott hangmintára. Megfigyelhetjük, hogy az advanced verzió esetén elenyésző a vonalakon kívülre eső pontok száma. Míg a basic verzió esetén több pont esik a határon kívülre, viszont az átlagos távolságuk lényegesen kisebb így is, mint a korábbi verzió esetén.



6.7. ábra. Korábbi modell eredményei adott hangmintára [10]

6.3 PEAQ Basic verzió megvalósítása Matlab környezetben

A korábbi alfejezetekben tárgyalt PEAQ modell megfelelő az általam készített audiokódoló vizsgálatára. Eddig a kódoló képes elkészíteni a kódolt hangmintát, amit meghallgatva próbálhatjuk összehasonlítani az eredetivel. Ez nem feltétlenül ad összehasonlítható értékeket. Meg tudjuk állapítani, hogy rosszabb-e a hangminőség, de egy másik hangminta ugyanolyan módon kódolt változatáról nem biztos, hogy meg tudjuk ítélni, hogy ez pont ugyanolyan, vagy mennyivel jobb vagy rosszabb a minősége, mint az előzőnek. Ezért van szükség egy objektív összehasonlítóra, ami a fent bemutatott PEAQ modell lesz. Létezik egy Matlab-ban implementált verziója a PEAQ Basic modellnek. Ezt még nem tesztelték alaposan, viszont szabadon felhasználható és letölthető. Ez Stephen Welch és Matthew Cohen munkájának köszönhető.

6.3.1 PEAQ működése Matlab-ban

A `scratch.m` nevű Matlab file-ból kiindulva megismerhetjük a kód működését. Ez a file végzi el az audioanyagok összehasonlítását.

1. Adatok beszerzése, tehát a két minta beolvasása és bufferbe betöltése
2. Határok kiszámítása, vagyis a minta kezdetének és végének meghatározása
3. További számításokhoz szükséges, egyelőre üres mátrixok létrehozása
4. PQCB kiszámítja a kritikus sávokat, ezt a függvényt több további függvény is alkalmazza
5. Szűrőmemóriák előkészítése
6. Belső pufferből adatok beolvasása
7. PQeval függvény végzi a kiértékelést
 - a. Diszkrét Fourier transzformációval keretekre bontás
 - b. Szint és mintázat adaptáció
 - c. Modulációs mintázat feldolgozása
 - d. Hangosság mértékének kiszámítása
 - e. Modulációs különbségek számítása
 - f. Zaj hangosságának kiszámítása
 - g. Sáv szélesség tesztek

- h. Zaj-maszkoló arány számítás
 - i. Detekció valószínűség kiszámítása
 - j. A hiba harmonikus struktúrájának kiértékelése, erős kvantálás esetén a hiba és a referenciajel harmonikus struktúrája erősen korrelált, ebből lehet következtetni az észlelt minőség romlására
8. Új struktúrába helyezés
9. Neurális hálóval ODG mérőszámok kiszámítása a kimenetekből

6.4 Saját audiokódoló vizsgálata a PEAQ basic modellel

A továbbiakban az általam készített audiokódoló vizsgálatom. Több audiorészletet azonos feltételek mellett kódolva hasonlítok össze. A PEAQ által számolt ODG értékek mellett, még hasznos információval szolgál, ha azt is vizsgáljuk, hogy milyen tömörítési arányt érünk el bizonyos változtatásokkal. Ezeket a tömörítési arányokat össze is vethetjük az ODG számokkal.

Három teljesen különböző hangmintát vettem a teszt elvégzéséhez.

1. blackbird: zongorás hangminta egy kis kísérettel
2. castanets: gitár kasztanyetta kísérettel, sok hirtelen változást tartalmaz
3. pipes: skótduda

A kódolást három különböző ablakmérettel végeztem el a teszt során. A választott ablakméretek rendre 256, 512 és 1024, a kisebb ablakméret jobb minőséget eredményez. A keretek 64 minta átfedésben álltak egymáshoz képest ebben a kísérletben.

A tesztet egyszerűen egy Matlab szkript végzi el. Első lépésként mono hangmintákat készít a két hangsáv mintákból. Majd mind három ablakmérettel elkészíti a kódoló az újrakódolt hanganyagokat, ezeket ki is menti különböző neveken, így meg is hallgathatjuk a kódoló eredményeit. A tömörítés arányát elmenti egy belső mátrix változóban. Ezeket a kódolt mintákat egyenként az eredetihez hasonlítja a PEAQ modell segítségével, az ODG értékeket pedig eltárolja egy másik mátrix változóban. Ezt a két lépést elvégzi az összes hangmintára. Tehát végeredménynek kapunk összesen kilenc kódolt hanganyagot, és két háromszor hármas mátrixot. Minden kódolt mintához tartozik egy tömörítési arány és egy ODG mérőszám a két mátrixból. A különböző hangminták a mátrixok sorainak felelnek meg, a különböző ablakméretek pedig az

oszlopoknak. Ebből a két mátrixból készült az alábbi 6.1. táblázat. Az adott ablakmérethez párba rendezve a tömörítési arányt és az ODG számot.

ablakméret	256		512		1024	
	<i>Tömörítés</i>	<i>ODG</i>	<i>Tömörítés</i>	<i>ODG</i>	<i>Tömörítés</i>	<i>ODG</i>
Blackbird	46,1%	-2,0256	28,2%	-2,3897	21,2%	-3,2891
Castanets	45,1%	-1,2618	25,7%	-2,7577	19,3%	-3,5409
Pipes	46,2%	-1,6528	32,2%	-2,0323	23,2%	-2,3950

6.1. táblázat. Tömörítési arányok és ODG mérőszámok

Megfigyelhető, hogy 256-os ablakmérettel nagyjából egyforma mértékben tömöríthető a három hanganyag. A növekvő ablakmérettel viszont a kasztanyettás minta esetén lényegesen kisebb méretet érhetünk el. A skótdudás hangminta pedig kevésbé tömöríthető, mint a zongorás (zenekaros) hangminta. Ezeket figyelembe véve vizsgáljuk meg az ODG mérőszámokat.

A zongorás hangminta esetén lényegesen rosszabb minőséget érzékel a PEAQ, mint a másik két minta esetén a 256-os ablakméretet nézve, viszont kevésbé romlik az 512-es ablakméret használatával. Az 1024-es ablakméretet választva hangmintától függetlenül rossz értékeket kapunk, de érdemes lehet megjegyezni, hogy a skótdudás hangminta (harmadik sor) esetén észrevehetően kevesebbet romlik a minőség az ablakméret növelésével. Nem meglepő a kasztanyettás hangminta rohamos romlása: míg a kis ablakméret esetén az eredményezte a legjobb minőséget, már az 512-es ablakmérettel lényegesen rontjuk a minőségét, nem is beszélve a 1024-es ablakméretekről, amivel már zavaróan rossz hangminta készül el, ez a mintában található hirtelen változásoknak tudható be. Ezt a romlást tükrözte a tömörítési arányok értéke is. A skótdudás hangminta kevésbé tömöríthető, viszont sokkal jobb eredményeket mutat hangminőség szempontjából.

A különböző ablakméretek mellett még egy tesztet elvégeztem. Mi történik, ha mindenhol eggyel kevesebb biten kódolunk, mint a modell által kiszámított. Előre sejthetjük, hogy ezzel lényegesen rontjuk a minőséget. Meghallgatás után az egy bit hiánya észrevehető, de ha már két bittel kevesebben kódolunk, akkor szinte hallgathatatlan lesz az újrakvantált hangminta.

Korábban az 5.3.3. alfejezetben volt szó a hirtelen tranziensekről. A kasztanyettás hangminta szintén ilyen nagy különbségű változásokat tartalmaz. Az ilyen hanganyagok indokolják a változó ablakméretű kódolást.

7 Változó ablakméretek

Többször is ugyanabba a nehézségbe ütköztünk, ez a hangmintában található hirtelen változások jelensége. Ha egy ablakban egy hirtelen változás az ablak első felében található, akkor kódoláskor kiszámolt bitek számából adódó kvantálási zajt, amely az egész ablakra kiterjed, az időtartománybeli utóelfedés elfedi, így nem zavarja a hallgatót. Viszont ha ugyanez a változás egy ablaknak a második felében található, akkor a frekvenciatartománybeli elfedéséből kiszámított bitek számának hatása az egész ablakra kihat ugyanúgy, mint az előző esetben. Azonban az időtartománybeli elő elfedési jelenség lényegesen rövidebb időtartamú, mint az utóelfedési jelenség. Így az ebből adódó kvantálási zaj zavarhatja a hallgatót, romlik a minőség.

Az előző fejezetben PEAQ által vizsgált kasztanyettás hangminta esetén pont ezt figyelhettük meg. Nagy ablakméret esetén sokkal rosszabb minőséget érzékel az emberi fül, és ugyanígy a PEAQ is ennek megfelelően rosszabb osztályzatot számol ki. De kis ablakméret esetén észrevehetően jobb eredményeket kaptunk ezzel a hangmintával, mint a többi esetén. Tehát a kis ablakméret valóban segít az ilyen hirtelen változásokkal teli hangminták kódolásában. Kis ablakméret esetén viszont csak kisebb tömörítést érhetünk el. A kompromisszumot a változó ablakméretű kódolás jelenti. Ez az elem az eredeti anyagban megtalálható Model-1-ben nem található meg.

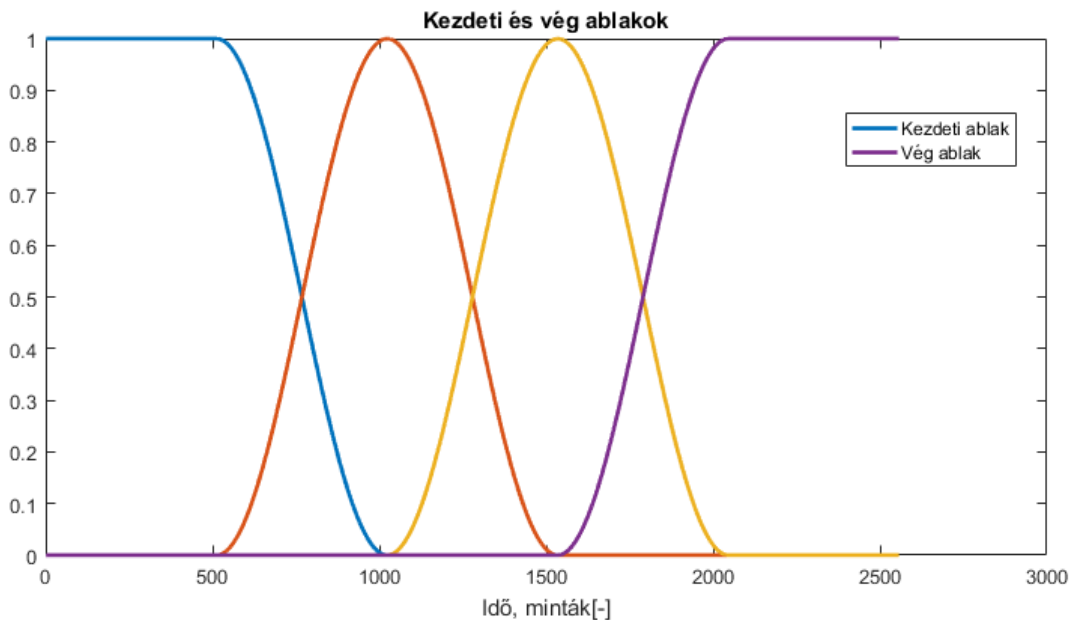
7.1 Változó ablakméret megvalósítása

Két fő változtatásból áll a változó ablakméretek megvalósítása. Először is ki kell számolni mely ablakok esetén szükséges kisebb ablakokra váltani, ezt a kis ablakot esetleg kell-e ismételni. Illetve a különböző Hann-ablakok és átmeneti ablakok elkészítése, amelyek szükségesek a kódolt hangminta összeállításához.

Egy egyszerű négyzetes összeg összehasonlítás fogja eldönteni, hogy a következő ablak esetleg kisebb méretű legyen-e vagy sem. Az ablak első és második felére kiszámítja a program a négyzetösszeget és ezek arányát számítja ki. Későbbi tapasztalatok alapján a négyzetösszeg-arányhoz a megválasztott határ az 5 lesz. Tehát ha az ablak két fele közötti energiaarány 5-nél nagyobb, akkor kis ablakra van szükség, ha ennél kisebb, akkor maradhat az eddigi nagyobb ablak.

A döntés után módosítani kell az ablak elhelyezkedését is nem csak a méretét. A kisebb méretű ablak, később kezdődik, mint ha egy nagyobb ablak lenne a helyén. Az indexelést ennek megfelelően kell alakítani. Nagy ablakok közötti ugrás, nagyról kicsire ugrás és kicsiről nagyra ugrás illetve a kisebbek közötti váltások is mind különböző esetet definiálnak. A kód, ha kisebb ablakra váltott, akkor még két további kis ablak beiktatása után vált vissza a hosszabb ablakokra.

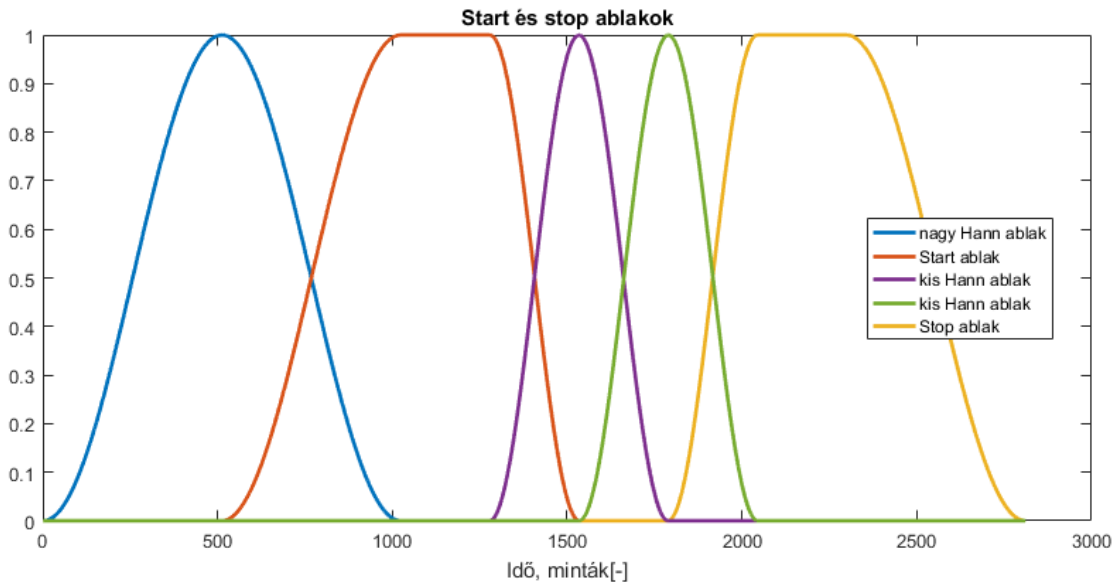
A kódolt hangminta visszaállításakor különböző Hann-ablakokra van szükség, ezeket egy külön függvény állítja elő. Szükség van egy kezdeti és vég ablakra, amelyek nem szimmetrikus Hann-ablakok, hanem egyik felük csak egyesekből áll nem emelt-koszinuszos függvényből. Ezt a két típusú ablakot az alábbi 7.1. ábrán figyelhetjük meg, kékkel a kezdeti ablak, lilával pedig a vég ablak jelölt.



7.1. ábra. Hangminta visszaállításához szükséges kezdeti és vég ablakok, közbeiktatott normál Hann-ablakokkal

Továbbá szükség van a nagyobb és a feleakkora, kisebb méretű ablakokra is, illetve az átmeneti ablakokra. A start nevű ablakra akkor van szükség, ha nagy ablak után kis ablak következik, első fele a nagy ablak első felével megegyezik, a második felében viszont egy negyed ablakméretig egyesek követik és utána a lefutás a kis ablak második felével egyezik meg. Így egy nagy–start–kis ablak sor összege ugyanúgy végig egy értékű, mint a korábban vizsgált egyforma méretű Hann-ablakoké. A stop nevű ablakra akkor van szükség, ha kis ablakméretről a nagyobbra váltunk vissza, ez a start ablaknak a tükörképe. Ezeket az ablakokat a következő 7.2. ábrán vizsgálhatjuk meg.

Az ábrán az is megfigyelhető, hogy a különböző méretű ablakok kezdőpontja eltérő, ez a korábban említett indexelést mutatja. Az általam készített kódoló, három kis ablakot használ egymás után, de az ábra egyszerűsége kedvéért csak kettő látható rajta.



7.2. ábra. Normál és kis Hann-ablakok közötti váltáshoz szükséges start és stop ablakok

Az átmeneteket figyelve láthatjuk, hogy a különböző ablakok összege továbbra is végig egy. Ezeken az ábrákon 50%-os átfedésű Hann-ablakokat használtam a különböző ablakok bemutatásához. A kódoló képes kisebb átfedésű ablakok kezelésére és készítésére is. Ezek annyiban különböznek, hogy középre konstans szakaszokat iktat be a kódoló, azonos ablakméret mellett, tehát a felfutás és a lefutás minden ablak esetén rövidebb lesz, így az átfedés is kisebb, ebből kifolyólag kevesebb ablakra van szükség a teljes hangminta kódolásához.

Még érdemes megjegyezni, hogy a pontos összeg számításához valójában páratlan mintaszámú Hann-ablakokra van szükség. Páros mintaszám esetén nincs egy értékű minta a Hann-ablak értékei között. Míg páratlan esetén a középső minta egy értékű. Ez az összegük számításakor fontos, páros mintaszám esetén nem tökéletesen egy az összeg, hanem kicsit kisebb változó értékeket vesz fel. Egy 1024 méretű ablak valójában egy $2N+1$ méretű Hann-ablak minden második mintája, és 1025 hosszú.

Tehát a teljes kódolt hangminta visszaállításához hat féle ablakra van szükség, normál és kis méretű Hann-ablakra, kezdeti és vég ablakra, továbbá az átmenethez szükséges start és stop ablakokra. Az ezeket előállító függvénynek két bemenete van, a nagy ablak mérete és az ablakok közötti átfedés mértéke. A kis ablak alapértelmezetten fele akkora méretű, mint a nagy ablak. A függvénynek két kimenete van. Az első

kimenete egy mátrix, amely tartalmazza az összes nagy méretű ablakot, a mátrix soraiban találhatóak az öt különböző ablak. A második kimenet pedig a kis méretű ablak.

Az állandó ablakméretet használó kódoló belső változója egy tömb, amely tartalmazza az egyes ablakokat kódolt formában. Ezt használja fel és állítja össze a kódoló a teljes kódolt hangmintát. A változó ablakméretű kódolás esetén ez a belső változó cella típusú lesz, erre a különböző ablakméretek miatt van szükség. A cella elemeinek mérete attól függ, hogy éppen mekkora méretű az ablak. Ezt a cellát vizsgálva dönti el a visszaállító ciklus, hogy milyen típusú ablakot kell használnia. Az ablakkal összeszorozza a kódolt ablakot és eltolva hozzáadja az előzőhöz, így alakul ki a teljes kódolt hangminta.

7.2 Változó ablakméreteket alkalmazó kódoló eredményeinek vizsgálata PEAQ-kel

A 6.4. alfejezetben vizsgált három hangmintát használtam továbbra is a kódoló vizsgálatához. Mint korábban láthattuk az ablakméret növelése, legjobban a kasztanyettás hangminta minőségét rontotta. Ezért a továbbiakban ezt a hangmintát kódolom az eredeti és az új változó ablakméretet alkalmazó kódolóval, ezeket az eredményeket hasonlítom az eredeti mintához a PEAQ segítségével. Ezek után pedig a PEAQ által számolt ODG értékeket hasonlítom össze. A választott ablakméret 512 és ezt különböző átfedésekkel is megvizsgáltam. Az alábbi táblázatban az eredmények olvashatóak.

ablakok közötti átfedés	változó ablakméretet alkalmazó kódoló	eredeti, azonos ablakméretet alkalmazó kódoló
256 – 50%	-2,4754	-2,4968
128 – 25%	-2,5784	-2,6757
64 – 12,5%	-2,6139	-2,7577

7.1. táblázat. Castanets hangminta ODG mérőszámai, különböző ablak átfedéssel készült minták, eredeti és módosított kódolóval történő kódolása után számítva.

Megfigyelhető, hogy az átfedés csökkentésével a minőség romlik. Ha egyenként vizsgáljuk az ODG érték párokat, akkor megfigyelhetjük, hogy nagyobb átfedés esetén kevésbé javít az érzékelt minőségen a kisebb és nagyobb ablakok kombinált használata. Ez annak tudható be, hogy az a kvantálási zaj, amely egy későbbi erősebb jel miatt a korábbi csendesebb részén az ablaknak megjelenik, lecsökken, amikor az előző ablakkal

összegezzük, mert az előző ablak nagyobb részét tartalmazza az aktuális ablaknak nagyobb átfedés esetén. Tehát kisebb átfedés esetén ezek a hatások zavaróbbak lehetnek, ebből kifolyólag a táblázat alsó sorában a változó és nem változó ablakméretű kódolás esetén lényegesen eltér a két ODG érték.

Ha 512-es ablak helyett, 256-os ablakot használunk, akkor nem lesz lényeges különbség, hiszen már az előző fejezetben is vizsgált kódoláskor a 256-os ablakkal kódolt kasztanyettás minta jó minősítést kapott PEAQ-tól.

A gitáros és skótdudás mintát hasonlóan 512-es ablakkal kódolva, és 64-es átfedést alkalmazva a 7.2. táblázatban feltüntetett értékeket kapjuk.

hangminta	változó ablakméretet alkalmazó kódoló	eredeti, azonos ablakméretet alkalmazó kódoló
Blackbird	-2,3904	-2,3897
Pipes	-2,0386	-2,0323

7.2. táblázat. Blackbird és Pipes hangminták ODG mérőszámai azonos paraméterekkel való kódolás esetén, változó és állandó ablakméretekkel

Tehát nincs lényeges különbség a változó és nem változó ablakméretű kódolás esetén ezekre a hangmintákra nézve. Ez a hangminták stacioner jellegének tudható be.

Összességében a változó ablakméretű kódolás megoldást jelent, ha a hangminta sok hirtelen változást tartalmaz, és kevesebb ablakot szeretnénk használni, ezért kisebb átfedést is alkalmaznánk.

8 Összegzés

Diplomadolgozatom témája manapság nagyon aktuális. A veszteséges audiotömörítő eljárásoknak köszönhetően mindenki számára könnyen elérhető lett a zene. Otthon ülve, az utcán sétálva vagy munka közben, mindenkit körülvesznek a hasonló módszerekkel kódolt audiojelek. Ezért érdemes megismerni a különböző tömörítő eljárások menetét, előnyeiket és hátrányaikat.

Munkám eredménye egy működő audiotömörítő Matlab környezetben, amely a MODEL-1 alapján készült. Különböző beállításokkal lehet kísérletezni vele. Kisebb vagy nagyobb ablakozással, vagy akár direkt kevesebb biten való kódolás eredményét is lehet vizsgálni, ezzel is jobban megértve a pszichoakusztikai modellek működését. Mindezt egy objektív összehasonlítóval is vizsgálhatjuk, mely a PEAQ. Nem csak meghallgatásos tesztek, de konkrét számszerűsített minőség jellemző értékeket is kapunk, melyekkel még inkább megismerhetjük a veszteséges audiotömörítő eljárások módszereit, és azt, hogy ezek miért is hatékonyak. Az eredeti Matlab-ban elkészített PEAQ modell nem volt alkalmas minden típusú mai hangminta beolvasására, a kód kis változtatásával ezt a problémát is megoldottam.

A MODEL-1 nem ad javaslatot változó méretű ablakok használatára. A kódoló fejlesztésének következő lépése, így a változó ablakméretek alkalmazására képes kódoló megvalósítása. Az így változott kódolt mintákat szintén megvizsgálhatjuk PEAQ segítségével. Így bármilyen a kódolóval készült hangmintára, meghallgatásos szubjektív tesztek mellett, a kódolással elért minőség objektív módszerrel is becsülhető.

A kódoló használata segíti a veszteséges audiotömörítő eljárások módszerének megismerését, így oktatási célokra is használható, mivel hasonló program eddig nem létezett.

Irodalomjegyzék

- [1] James D. Johnston: *Transform Coding of Audio Signals Using Perceptual Noise Criteria*, IEEE Journal on selected areas in communication, Vol. 6, No. 2, 1988
- [2] P. Kabal: *An Examination and Interpretation of ITU-T BS.1387: Perceptual Evaluation of Audio Quality*, 2003
- [3] Brian C. J. Moore, Brian R. Glasberg: *Suggested formulae for calculating auditory-filter bandwidths and excitation patterns*, 1983
- [4] Rassol Raissi: *The Theory Behind Mp3*, 2002
- [5] Arizona State University: *Analysis of the MPEG-1 Layer III (Mp3) Algorithm Using Matlab*, 2011
- [6] Seymour Shlien: *Guide to MPEG-1 Audio Standard*, IEEE Transactions on broadcasting, Vol. 40, No. 4, 1994
- [7] Ted Painter, Andreas Spanias: *Perceptual Coding of Digital Audio*, Proceedings of the IEEE, Vol. 88, No. 4, 2000
- [8] Andreas Spanias, Ted Painter, Venkatraman Atti: *Audio Signal Processing and Coding*, 2007
- [9] Xiph.Org Foundation: *Vorbis I specification*, 2015
- [10] Thilo Thiede, William C. Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G. Beerends, Catherine Colomes, Michael Keyhl, Gerhard Stoll, Karlheinz Brandenburg, Bernhard Feiten: *PEAQ – The ITU Standard for Objective Measurement of Perceived Audio Quality*, J. Audio Eng. Soc., Vol. 48, No. ½, 2000