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

Herczeg Ferenc

FORRÁS AZONOSÍTÁS ÉS SZÉTVÁLASZTÁS KEVERT AUDIOANYAGBAN

KONZULENS

Firtha Gergely

BUDAPEST, 2016

HALLGATÓI NYILATKOZAT

Alulírott **Herczeg Ferenc**, szigorló hallgató kijelentem, hogy ezt a szakdolgozatot 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, 2016. 12. 08.

...…….

Herczeg Ferenc

Tartalomjegyzék

Ábrajegyzék

Kivonat

A szakdolgozatom témája a forrás azonosítás és szétválasztás problémája kevert audioanyag esetén, amely a digitális jelfeldolgozásnak jelenleg is kutatás alatt álló tématerülete. A bevezetés után bemutatom a manapság használatos audio keverési módszereket és az ebből adódó nehézségeket a források szeparálására nézve. Továbbá ismertetek néhány eljárást, amely alkalmas a feladat végrehajtására.

Részletesen kifejtem az Azimuth Discrimination and Resynthesis algoritmust, annak előnyeit és hátrányait, valamint megvalósíthatóságát és alkalmazhatóságának feltételeit különféle források használata esetén. Az időtartományi jelek szintézise a frekvenciatartományban zajlik, az átjárást a legtöbb esetben a közismert Fourier-transzformáció biztosítja. Az analízis alternatívájaként szokták a Wavelet-transzformációt megemlíteni és használni. Ennek bemutatása után összevetem a Fourier-eljárással, valamint kifejtem a vele kapcsolatos várakozásaimat.

Ezt követően a már implementált algoritmus kimeneteleit hasonlítom össze a kétféle transzformáció alkalmazása esetén és értékelem az eredményeket az előzetes remények figyelembevételével. A szakdolgozatot összegzés zárja, amelyben összefoglalom az eredményeket és kitekintést nyújtok a továbbfejlesztési lehetőségek, jövőbeli célok felé.

Abstract

The topic of my thesis is the sound source separation in mixed audio signals, which is still an active research field of digital signal processing. After the introduction I present the audio mixing technics and the difficulties of the source separation. Additionally I present some methods, which are able to solute the problem.

Afterwards I expand particularly the Azimuth Discrimination and Resynthesis algorithm, its advantages and disadvantages, as well as practicability and conditions of applicability using different sources. The synthesis of the time domain signals happens in the frequency domain, the connection is provided by the well-known Fourier transform. The Wavelet transform is mentioned and used as an alternative variation of the Fourier analysis. After the presentation of its theory I resemble these methods, and explicate my expectations with it.

Then I compare the issues of the implemented algorithms using the different transforms and evaluate the results attending to the previous hopes. The paper is closed by summary, where I summarize the results and describe further development opportunities.

1 Bevezetés

Napjainkban folyamatos fejlesztéseket érünk el a különféle hangtechnikai eszközök és eljárások területén, amelyek célja a még teljesebb és még több élvezetet nyújtó hangélmény létrehozása. A monó hanghoz képest hatalmas előrelépést jelentett a sztereó bevezetése, amely még mindig a legjobban elterjedt technika. Mára már otthonainkban is elérhetővé váltak a sokcsatornás hangrendszerek, amelyek sokáig csak a filmszínházak különlegességei voltak. A digitális technikának köszönhetően a stúdiókban egyre komplexebb keverések és hanghatások érhetők el, és már házilag is elfogadható minőségű mixeket állíthatunk elő. Ezek célja a minél élet hűbb hangélmény elérése, a fizikai valóság pontosabb leírása, illetve olyan hanghatás keltése, amely lenyűgözi a hallgatót, függetlenül attól, hogy ez a világunkban megtalálható vagy sem.

Ezzel szemben a művelet fordítottja, azaz a kész audiomix forrásainak kinyerése a már kevert anyag alapján, még mindig nagy kihívást jelentő feladat. A működőképes módszer átjárást biztosíthat az úgynevezett csatorna alapú sokcsatornás hangrendszerek (Dolby) és a jelenleg is fejlesztés alatt álló modell alapú hangrendszerek (Wave Field Synthesis, Dolby Atmos) között, illetve lehetőséget ad zenei anyagok újrakeverésére, információk kinyerésére.

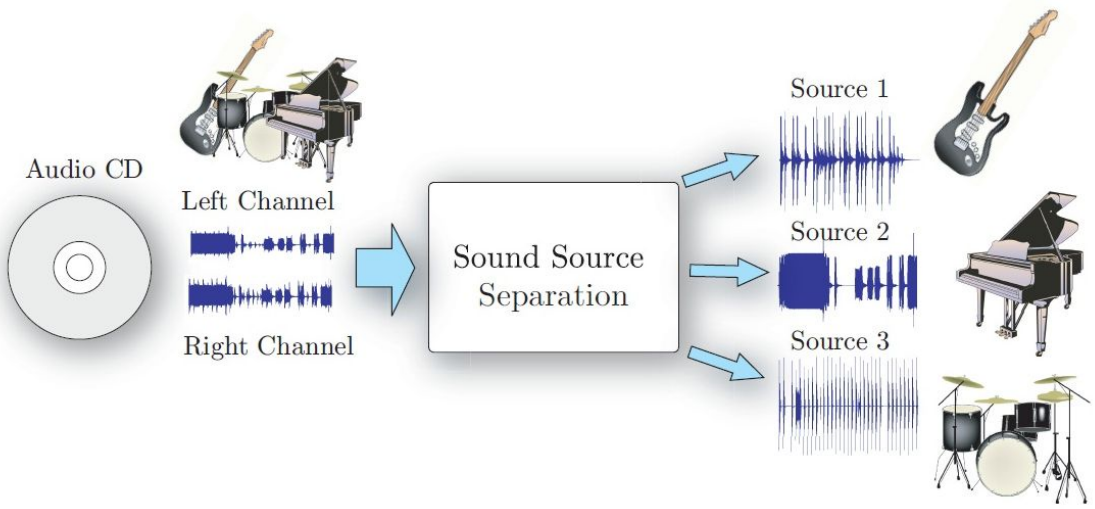
A forrás szeparálás nehézségeket rejt magában, hiszen egy kész audiomix számos forrást tartalmazhat és számos előállítási módon készülhetett, így a nem egységes hanganyagokat különbözőképpen kell feldolgozni és a szétválasztást elvégezni. A konkrét eljárások pedig csak bizonyos feltételek esetén működnek elfogadható minőségben, ezért korlátozottak a lehetőségek.

Ahogyan a tartalmi összefoglalóban is említésre került, a szakdolgozat célja betekintést nyújtani a digitális jelfeldolgozás ezen területére és bemutatni azon technikákat, amelyek segítségével elfogadható minőségű eredmény érhető el.

2 Sound Source Separation

2.1 Bevezetés

A források szétválasztásának számtalan alkalmazási területe van, egészen a kép és video feldolgozásán keresztül az orvoslásig. Gyakorlatilag egy olyan jelfeldolgozási feladattal állunk szemben, ahol több objektum egyesítésével létrejött információ halmazból akarjuk elkülöníteni az eredeti alkotókat. A hangfeldolgozás területén ezeket a módszereket *Sound Source Separation* (SSS) eljárásoknak nevezzük. Az SSS megoldást kínál többek között a témában gyakran emlegetett koktélparti-problémára [1] is. A koktélparti-jelenség az a képessége a hallásunknak, hogy ki tudjuk választani a számunkra fontos emberi beszédet egy olyan környezetben, ahol egyszerre több beszélgetés hallatszik. Ezt átvihetjük technikai értelembe is, azaz elkülöníthetjük a különféle hangjeleket egymástól egy adott audiomixben.



2.1. ábra: Forrás szétválasztás sémája [2]

Az SSS technikák az alábbi három csoportba oszthatóak: *Blind Source Separation* (BSS), *Semi-Blind Source Separation* (SBSS) algoritmus és *Computational Auditory Scene Analysis* (CASA) technikák. [2]

Annak függvényében, hogy mennyi információval rendelkezünk a forrásokat és a keverési eljárást illetően beszélünk BSS vagy SBSS módszerekről. A hanganyag tulajdonságai szorosan összefüggnek azzal a környezettel, ahol a rögzítés készült. Ennek ismeretének hiányában pusztán statisztikai jellegű információink vannak az adott jelsorozatot illetően, ezért szokták a BSS algoritmusokat általában statisztikai

algoritmusoknak nevezni. Persze ezekben az esetekben is élhetünk a hanganyaggal kapcsolatban előfeltevésekkel, sőt ilyenkor erőteljesebb feltételeket kell szabnunk az algoritmusok használhatóságát illetően. A CASA technikák esetén a matematikai módszerekkel ellentétben olyan eljárásokról beszélünk, ahol a számítógépeket programozzuk fel az emberi hallásmodellekkel. Az algoritmus az adott hanganyagból igyekszik felépíteni azt az eredeti környezetet, amelyben a jelenség lejátszódhatott az eredeti forrásokkal, és ennek segítségével próbálja szeparálni azokat. Természetesen ez a hallás jellemzőinek pontos ismeretét igényli. A CASA lehetővé teszi a monaurális (egy füllel érzékelt) hangok, azaz az egy csatornás hangok vizsgálatát. Viszont ezek az eljárások csak nagyon speciális feltételek esetén működnek kielégítően, így az alkalmazhatósága nem olyan széleskörű, mint a BSS esetében. [2] [3]

A forrás szétválasztás nehézségét általában a csatornák és források egymáshoz viszonyított száma határozza meg. Ha az M csatornájú anyagban N forrás van jelen, akkor a probléma az alábbiak szerint osztályozható:

1. M > N esetén túlhatározott
2. M = N esetén pontosan-meghatározott
3. M < N esetén alulhatározott

Zenei anyagok esetén általában a legutóbbi esettel találkozhatunk, hiszen a leggyakoribb zenei felvétel a sztereó, amelyben rendszerint kettőnél több hangszer szokott szerepelni. [2]

A keverési eljárás során használt modell matematikai leírására a *mixing matrix* szolgál, amely a források vektorából szorzással, illetve elemenkénti konvolúcióval állítja elő a csatornákat. A forrás szétválasztás leegyszerűsíthető gyakorlatilag arra a matematikai problémára, hogy a rendelkezésünkre áll a kimenet vektora és keressük az ismeretlen mátrixot és a bemenet vektorát.

2.2 Sztereó hangrögzítési és keverési technikák

A 1960-as években került bevezetésre és azóta széles körben elterjedt a sztereó rendszer, amely jobban illeszkedik a hallásunk kétfülű (binaurális) jellegéhez. A kétcsatornás hangzás annyival jobb zenei élményt nyújtott, és olyan könnyen meg lehetett valósítani, hogy rövid időn belül szabványos lett. Előnye továbbá, hogy a sztereó hangtér visszaállításához használt hangszóró elrendezés a rögzítésekor használt mikrofontechnikától független. Napjainkban is ez a legelterjedtebb hangzás, gondoljunk csak a televízióra, rádióra, CD és az újra divatba jövő LP lemezekre. A multimédiás tartalmak esetén is eléggé elterjedt még a sztereofónia, illetve az otthoni zenehallgatási lehetőséget vizsgálva – árban, megvalósításban és élményben is összevetve - sokáig még fenn fog maradni. Ezért érdemes az általában használt kétcsatornás felvételi és keverési technikákat megvizsgálni.

A sztereó rögzítési technikák alapvetően két csoportba foglalhatóak. A legkézenfekvőbb lehetőség, hogyha a felvétel csupán két mikrofonnal történik (klasszikus felvétel). A két mikrofon egymáshoz, illetve a forrásokhoz viszonyított helyzete alapján más-más hangjelet kap és különösebb utólagos manipulációk nélkül lesznek a bal illetve a jobb csatorna forrásai. A másik lehetőség, hogy minden forrás külön-külön kerül rögzítésre (közvetlen felvétel), a mikrofonok a lehető legközelebb kerülnek a hangforrásokhoz, hogy minél kevesebb zaj illetve áthallás keletkezzen a bementen. A sztereofon anyag pedig az utólagos munkálatok során készül el. [4]

A rögzítési technikától függetlenül három keverési módszert szoktak megkülönböztetni: *instantaneus*, *anechoic* és *convolutive*. [2]

2.2.1 Instantaneous Model

A legáltalánosabb keverés az instantaneous (pillanatnyi) vagy lineáris modell (2.2. ábra) alapján készülhet. Ebben az esetben a kimenetek a források lineáris kombinációjaként állnak elő:

$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t),\ m = 1, \dots, M, \tag{2.1}$$

ahol s_n a forrás, x_m a csatorna, a_{mn} az úgynevezett skaláris tényező az n. forrás és m. csatorna között.

2.2. ábra: Instantaneous Model [2]

Ez történhet az előző pontban bemutatott felvételi technikákkal. Egyrészt a források külön, szeparálva és közvetlenül kerülnek rögzítésre majd a jelek a bal és jobb csatornában történő különböző erősítésével kapják meg a helyüket a sztereó képben. Az eljárást, mikor a forrás pozíciója csupán az erősítések függvénye *amplitude panning*-nek nevezzük. A két leggyakoribb amplitude panning módszer:

Konstans amplitúdó szabály: a forrás amplitúdója állandó, ez oszlik meg a két csatorna között.

$$\alpha_n^L + \alpha_n^R = 1$$
$$\alpha_n^L = (1 - \phi_n)$$
$$\alpha_n^R = \phi_n$$
$$\phi_n = \frac{\alpha_n^R / \alpha_n^L}{1 + \alpha_n^R / \alpha_n^L} \tag{2.2}$$

Konstans teljesítmény szabály: a forrás energiája állandó, ez oszlik meg a két csatorna között.

$$(\alpha_n^L)^2 + (\alpha_n^R)^2 = 1$$
$$\alpha_n^L = \cos\left(\frac{\phi_n \pi}{2}\right)$$
$$\alpha_n^R = \sin\left(\frac{\phi_n \pi}{2}\right)$$
$$\phi_n = \tan^{-1}\left(\frac{\alpha_n^R}{\alpha_n^L}\right) \cdot \frac{2}{\pi} \tag{2.3}$$

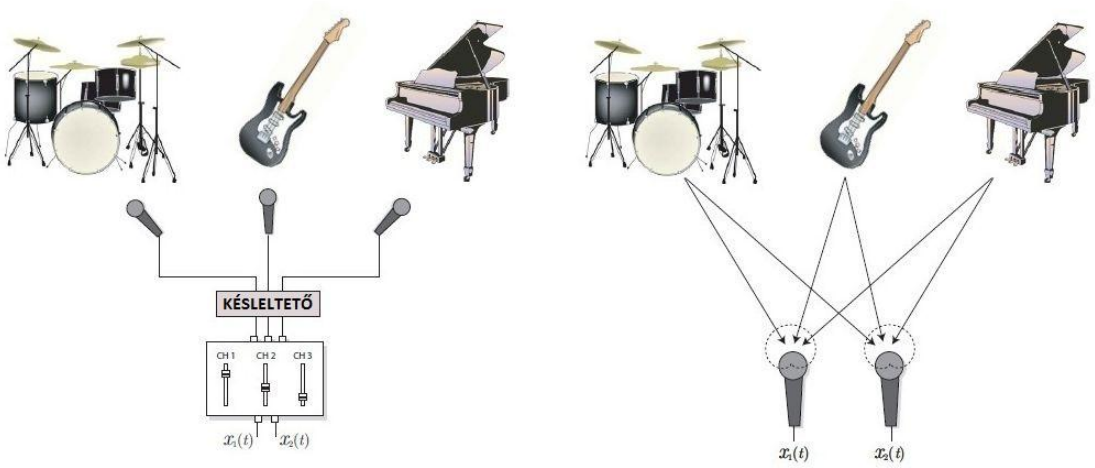
Másrészt a felvétel történhet közvetetten, azaz állítható iránykarakterisztikájú mikrofon párt helyeznek el a hangtér közepén. Mivel a mikrofonok egy helyen vannak, egy adott pontból érkező hang két vett jele között semmilyen késleltetés sincs. Az eltérő iránykarakterisztikák miatt azonban a két jel között jelentős intenzitáskülönbség lesz, ebből pedig bizonyos korlátokkal az irány meghatározható. [2] [5]

2.2.2 Anechoic Model

Az anechoic (visszhangmentes) vagy késleltetett modell (2.3. ábra) az előzőnek kiterjesztett változata, ahol a különböző erősítés értékek mellett különböző késleltetések is beállításra kerülhetnek:

$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t - \delta_{mn}),\ m = 1, \dots, M, \tag{2.4}$$

ahol δ_{mn} a hang késleltetése az m. mikrofon és n. forrás között.



2.3. ábra: Anechoic Model [2]

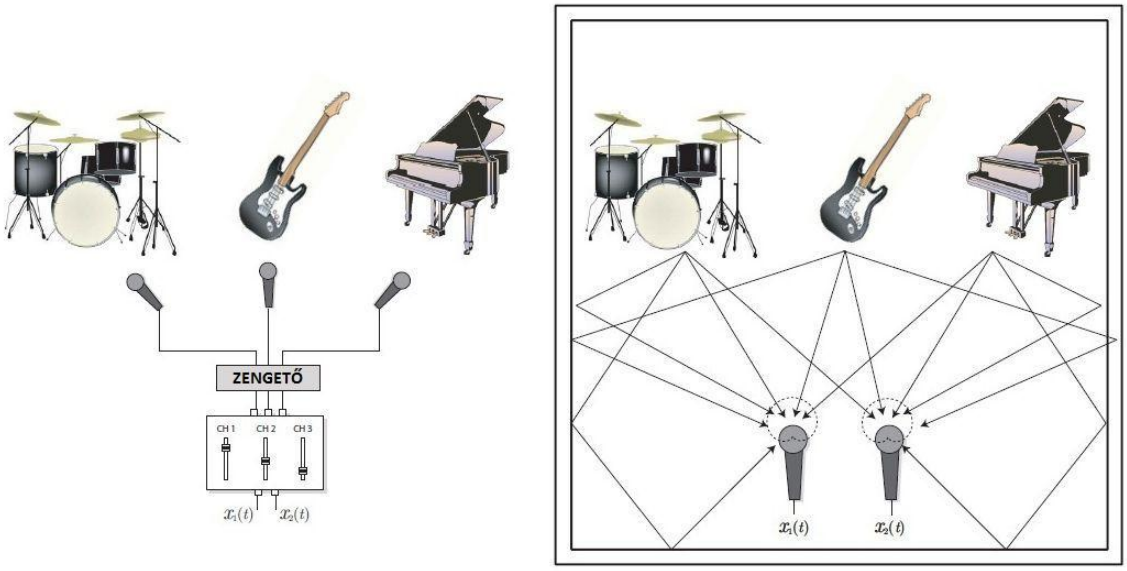
Ez megvalósítható egy késleltető és egy erősítő segítségével, vagy használhatunk két külön mikrofont, amelyeket egymástól megadott távolságban helyezünk el. A távolság akár több méter is lehet, a helyiség méretétől függően. A két mikrofon iránykarakterisztikái azonosak. A két jel között jelentős időkülönbség lesz, valamint részben intenzitás különbség is, ezekből pedig az irány meghatározható. [2] [5]

2.2.3 Convolutive Model

A convolutive (konvolúciós) vagy visszhangos modell (2.4. ábra) az eddigiekhez képest figyelembe veszi azt a környezetet, ahol a felvétel készül, azaz a reflexiókat, a visszhangokat:

$$x_m(t) = \sum_{n=1}^{N} \sum_{\tau=1}^{L_{imp}} a_{mn\tau} s_n(t - \delta_{mn\tau}),\ m = 1, \dots, M, \tag{2.5}$$

ahol L_{imp} azon utak száma, amelyet a hang meg tud tenni a forrástól a szenzorokig.



2.4. ábra: Convolutive Model [2]

A rögzítés ebben az esetben is történhet a források egyesével történő felvételével, amelyet utólag lehet manipulálni. Illetve lehet egy zárt, visszhangos környezetben mikrofon párral rögzíteni. [2] [5]

2.3 Forrás szétválasztásra alkalmas eljárások

A következőkben röviden ismertetésre kerül néhány SSS algoritmus, amelyek a legeredményesebbek és a leggyakrabban fellelhetőek a különféle irodalmakban és tudományos értekezésekben. Az itt bemutatandó módszerek a BSS körébe tartoznak.

2.3.1 Independent Component Analysis (ICA)

Az ICA jól működő eljárás a pontosan-meghatározott (M = N) esetekben, vagyis mikor annyi csatorna áll rendelkezésünkre, ahány forrás szerepel az audioanyagban. (Természetesen csak M > 1 esetekről van értelme beszélni.) Az algoritmus a sajátvektor keresésen alapszik. Az M csatornás keverékről feltételezzük, hogy minden csatorna N

darab független vektor lineáris kombinációja. A csatornák jeleiből alkotott mátrix SVD felbontása adja meg a források időtartományi jeleit. [6]

2.3.2 Degenerate Underdetermined Estimation Technique (DUET)

A sztereó aluldeterminált anyagok esetében az egyik elterjedt alkalmazás a DUET, amely elsőként volt alkalmazható anechoic keverés esetén. Az eljárás pont az időkülönbséget használja ki a két csatorna mintái közötti arányt vizsgálva, és a relatív csökkenés és késleltetés értékéből következtet a források pozíciójára. Több csatorna esetén a Multiple sENsor dUET (MENUET) szokták használni. [2] [7]

2.3.3 Panning Index Window (PIW)

A PIW esetében a két csatorna között korrelációs függvények kerülnek kiszámításra. Ezek a függvények mutatják meg a lineáris összefüggést a két csatorna mintái között. A hasonlósági függvényt vizsgálva különböző korrelációs szintek és értékek, amelyeket panning indexeknek nevezünk, határozhatók meg. Az indexek maximális értékeinél források találhatók, majd ezek csoportosításával határozhatók meg az eredeti jelalakok. [8]

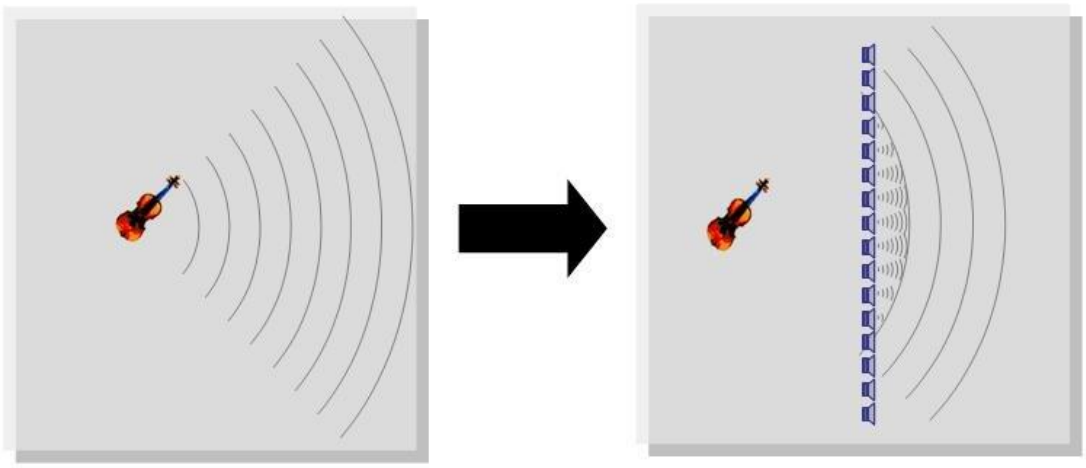
2.3.4 Multi-Level Thresholding Separation (MuLeTS)

MuLeTS módszer elsősorban a képek feldolgozásához lett kifejlesztve, amelynek során a kép egyes részei elkülöníthetőek egymástól. A képek hisztogramjai alapján különböző küszöbértékeket lehet meghatározni és két küszöbszint közötti terület a hisztogramon a kép egy adott részéhez tartozik. Ezek különválasztásával a kép szegmentálhatóvá válik. Hasonló az eljárás az audioanyagok esetében is. A két csatorna amplitúdó különbségét kell vizsgálni az idő-frekvencia tartományban, majd annak hisztogramjából egy speciális küszöbözési algoritmust használva bizonyos frekvencia tartományok meghatározhatók, amelyek egy adott forráshoz tartoznak. [2] [8]

2.4 Wave Field Synthesis

A bevezető fejezetben említésre került, hogy a forrás szétválasztás átjárást tud biztosítani a csatorna alapú sokcsatornás és az úgynevezett modell alapú hangrendszerek között. A DVD és Blu-ray lemezek megjelenésével otthonainkban is könnyen elérhetővé vált az 5.1-es, 7.1-es térhangzás, amely célja többek között a minél valóság hűbb hangzás megteremtése. Azonban ezekben a technológiákban közös, hogy a térhatást csak egyetlen pontban, az úgy nevezett *sweet spot*-ban képesek elérni.

Ezzel szemben más technológiák, például a *Wave Field Sythesis* (WFS) ezt a pontot kiterjesztik egy nagyobb térrészre, így képes a hangtér teljes, fizikai reprodukálására, persze bizonyos korlátozásokkal. A WFS technológia a Huygens-Fresnel elven alapul, amely kimondja, hogy valamennyi hullámfront előállítható gömbhullámok szuperpozíciójaként. [9]



2.5. ábra: Valós hangtér leképződése WFS-re [10]

Ez a gyakorlatban egy vonalmentén elhelyezett hangszóró sokaságot jelent, ahol a hangszórók szolgálnak a gömbhullámok forrásaiként. (2.5. ábra) A két rendszertechnológia között az is eltérés, hogy még egy sztereofon rendszer esetén a csatornák jeleit tároljuk, és térhatást pusztán a csatornák közötti intenzitás- és fáziskülönbség alkalmazásával érjük el, addig a WFS esetén a virtuális források jelei kerülnek tárolásra. Így WFS alkalmazása során az adott hangszóró elrendezésre valós időben tudunk renderelni. Ekkor megjelenik a két hangrendszer közötti kompatibilitás vizsgálata iránti igény. Egy már meglévő kevert hanganyagot, csak akkor tudunk egy WFS rendszeren kielégítően megszólaltatni, ha a kevert audioanyagból ki tudjuk nyerni a virtuális forrásokat, a két rendszer között nem elegendő egy lineáris leképzés.

A SSS még nagyobb jelentőséget kapna a WFS rendszer térhódítása esetén, azaz ha minél több filmszínházat, koncerttermet felszerelnének ilyen technikával, mert akkor a már meglévő hanganyagokat is le lehetne vetíteni különösebb újrarögzítés és újrakeverés nélkül. Ez idáig csak kevés, például a Fraunhofer cég által létrehozott IOSONO rendszerrel felszerelt vetítőtermek alkalmasak nagyobb közönség befogadására.

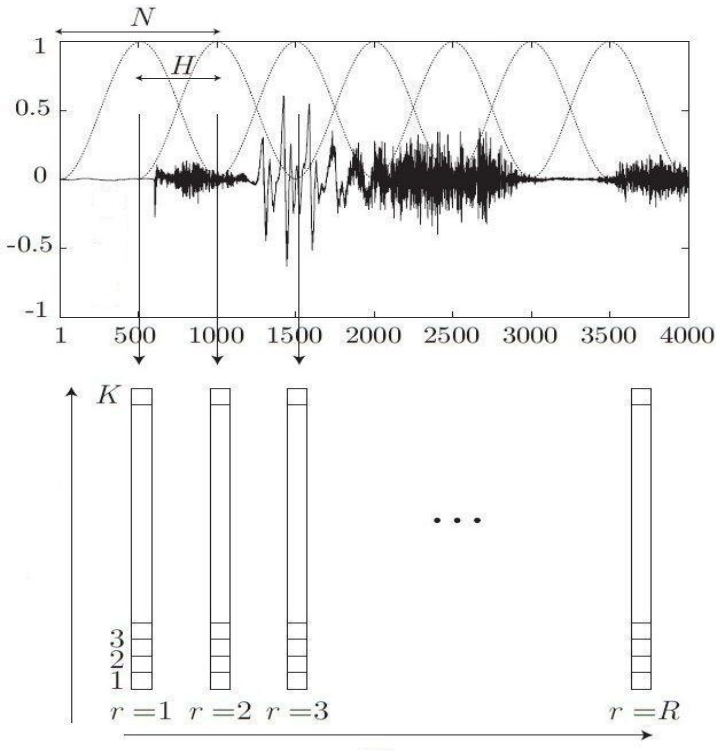
2.6. ábra: IOSONO rendszer Hollywoodban, Mann's Chinese Six Theatre [11]

3 Azimuth Discrimination and Resynthesis

3.1 Bevezetés

A következőkben részletesen bemutatásra kerül, hogy az *Azimuth Discrimination and Resynthesis* (ADRess), amely a tématerülten szintén gyakran alkalmazott megoldás, miként választja ki a virtuális forrásokat kevert sztereofon audioanyagból.

A hanganyag feldolgozása, mint a legtöbb SSS módszer esetén, idő-frekvencia tartományban történik. Időtartományból frekvenciatartományba az áttérést a Fourier-transzformáció biztosítja. Idő-frekvencia tartománybeli képet úgy kapunk, ha transzformációt ablakozva végezzük el, ez a *Short Time Fourier Transformation* (STFT). Ennek eredményeként minden időszegmensben megkapjuk a jel spektrum komponenseit. Egy hosszabb jel STFT-je a jelből H ugrásméretenként vett N minta N hosszú diszkrét Fourier-transzformáltját jelenti. (3.1. ábra)



3.1. ábra: STFT szemléltetése [2]

Jelen esetben is különböző megkötésekkel kell élni. Az ADRess olyan anyagokat tud kezelni, amelyek instantaneous keveréssel készültek, így a források pozíciója a két

hangszóró között annak az eredménye, hogy egy forrás mekkora intenzitással foglal helyet a bal illetve a jobb csatornában. Ezt az intenzitás különbséget használja ki ugyanis az algoritmus. Továbbá feltételezzük, hogy rendelkezünk egy előzetes becsléssel a virtuális források számáról. Illetve megkötjük, hogy teljesüljön a W-diszjunkt ortogonalitási feltétel, azaz időszegmensekben ne legyen frekvenciatartománybeli átfedés a források között. Ez a gyakorlatban azokra az anyagokra igaz, ahol párbeszéd hallható, a zenei anyagokra csak nagyon ritka esetben teljesül. [12] [13]

3.2 Működés [12]

Az amplitude panning keverés alapján így fejezhetőek ki a bal és jobb csatorna időtartománybeli jelei:

$$x_L(t) = \sum_{n=1}^{N} a_{Ln} s_n(t) \tag{3.1}$$

$$x_R(t) = \sum_{n=1}^{N} a_{Rn} s_n(t), \tag{3.2}$$

ahol s_n az n. független forrás, a_{Ln} és a_{Rn} a bal illetve jobb csatorna erősítési tényező. Látható, hogy a leírás konzisztens a (2.2)-vel. Az n. forrásra nézve meghatározhatjuk az *intenzitás tényező*t:

$$g(n) = \frac{a_{Ln}}{a_{Rn}}. \tag{3.3}$$

Ezt rendezve kapjuk:

$$a_{Ln} = g(n) \cdot a_{Rn}. \tag{3.4}$$

Innen látszik, hogyha a jobb csatornát felerősítjük a $g(n)$ intenzitás tényezővel, akkor a bal és jobb csatorna egyenlő erősségű lesz, tehát $x_L(t) - g(n) \cdot x_R(t)$ különbségképzéssel kiejthető az n. forrás. Ha az n. forrás a jobb csatornában dominánsabb, akkor $x_L(t) - g(n) \cdot x_R(t)$ különbséget vizsgáljuk, ha a balban, akkor a $x_R(t) - g(n) \cdot x_L(t)$. Az ADRess lényege tehát a források elnyomásán alapszik. Mivel a források intenzitás arányát előre nem tudjuk, ezért meg kell keresni. Ezt az intenzitás tényezővel tesszük, amely gyakorlatilag egy 0-tól 1-ig futó változó, tetszőleges léptékben. Természetesen ez az idő-frekvencia tartományban történik. Tehát:

$$X_L(k) = \sum_{s=0}^{S-1} x_L(s) \cdot e^{-j\frac{2\pi ks}{S}} \tag{3.5}$$

$$X_R(k) = \sum_{s=0}^{S-1} x_R(s) \cdot e^{-j\frac{2\pi ks}{S}}, \tag{3.6}$$

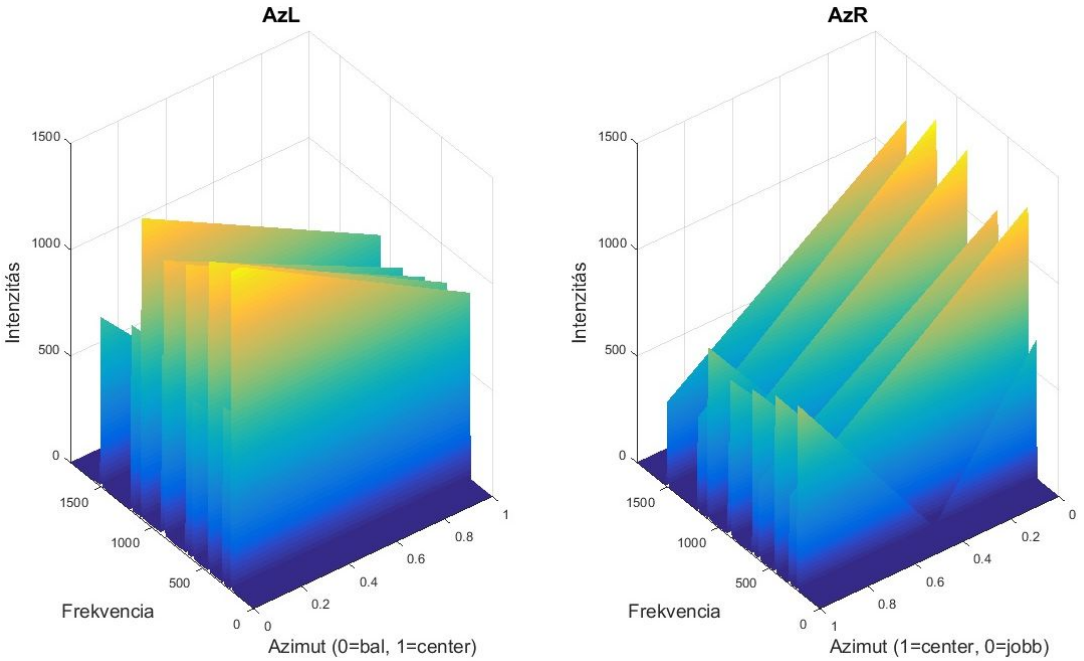
ahol $X_R(k)$ a jobb csatorna, még $X_L(k)$ a bal csatorna frekvencia tartománybeli képe egy időablakra nézve, S (jelen esetben N már a források számát jelenti) pedig a minták száma az ablakban. Képezzük a következő különbségeket:

$$AzL(k,g) = |X_R(k) - g \cdot X_L(k)| \tag{3.7}$$

$$AzR(k,g) = |X_L(k) - g \cdot X_R(k)|. \tag{3.8}$$

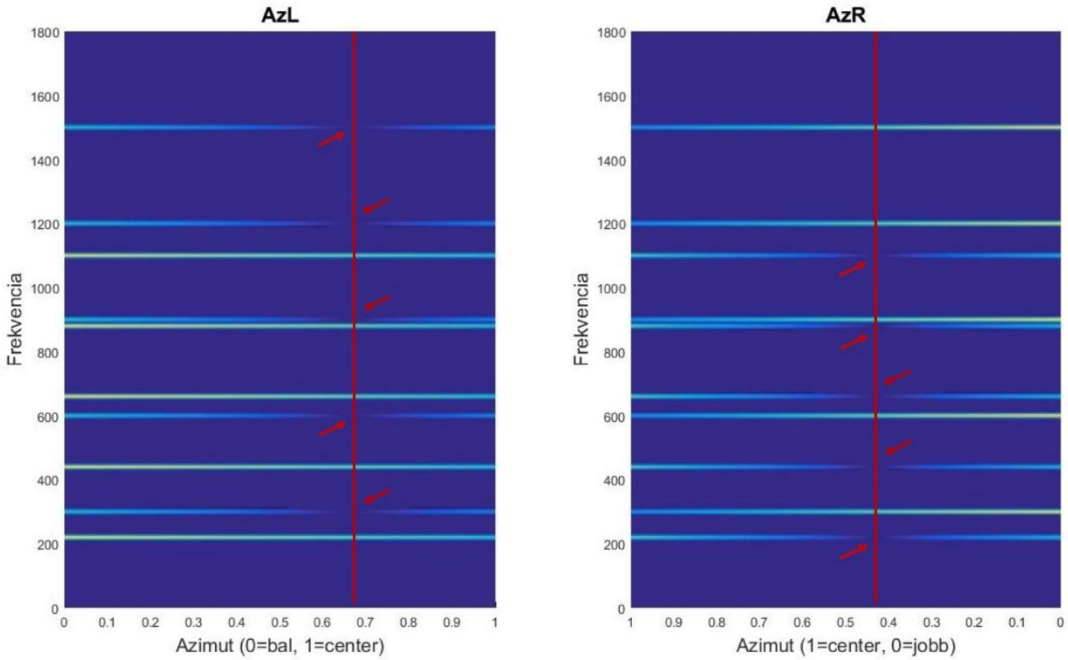
A különbségképzést minden időszegmens esetén elvégezzük, minden frekvencia binre (k) és minden g index értékre.

Ha az intenzitás tényezővel egybeesik a bal és jobb csatorna közötti intenzitás érték különbség, akkor a különbségképzés során a g helyeken kioltás keletkezik azokon a frekvencia bineken, amelyek az adott forráshoz tartoznak. A minimum helyek keresését az egész sztereó tartományra el kell végezni, azaz ha a forrás a bal csatornában dominál, akkor az (3.7)-t, még ha a jobb csatornában dominál, akkor a (3.8)-t vesszük figyelembe. (3.2. ábra) A g indexet szokás azimutnak nevezni, ami vízszintes irányszöget, a vízszintes síkban fekvő referenciaegyenessel bezárt szöget jelent a terminológia szerint, főként a csillagászatban, földrajztudományban használják. Az azimut érték a forrás irányát határozza meg a sztereó képben.



3.2. ábra: A különbségképzés eredménye a bal és a jobb csatorna esetén

A 3.2. ábrát felülről tekintve, az azimut-frekvencia síkon jól láthatóak a minimum értékek.



3.3. ábra: Azimut-frekvencia sík

Ezek után a minimum helyekből maximumokat, a maximum helyekből minimumokat képzünk az alábbiak szerint:

$$AzL(k,g) = \begin{cases} AzL(k)_{max} - AzL(k)_{min}, \ ha\ AzL(k,g) = AzL(k)_{min} \\ 0, \ k\ddot{u}l\ddot{o}nben \end{cases} \quad (3.9)$$

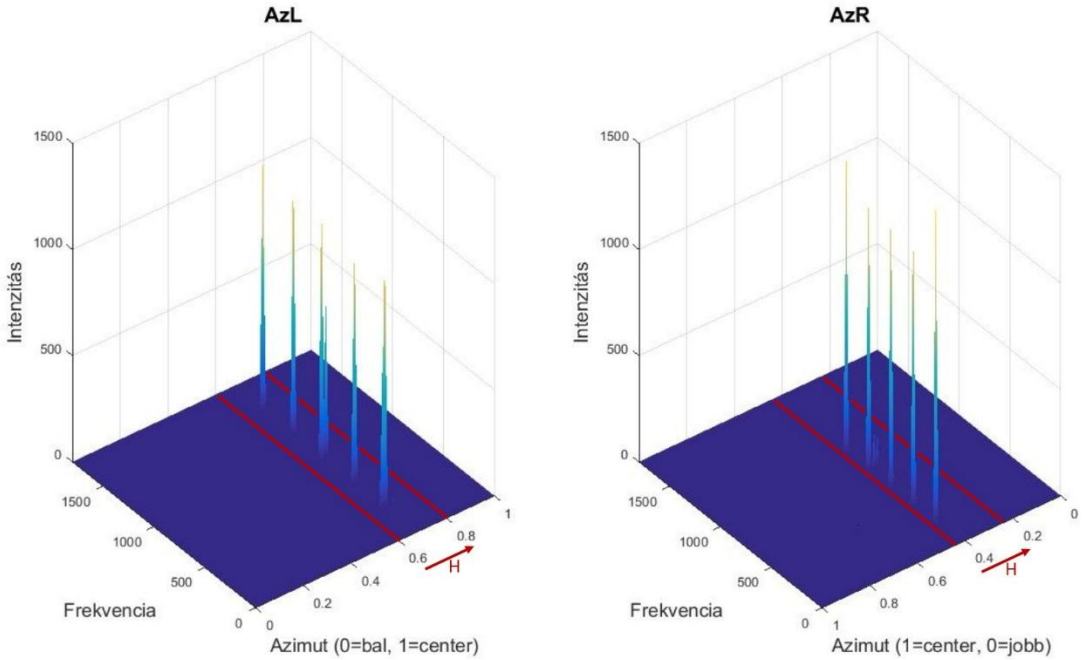
$$AzR(k,g) = \begin{cases} AzR(k)_{max} - AzR(k)_{min}, \ ha\ AzR(k,g) = AzR(k)_{min} \\ 0, \ k\ddot{u}l\ddot{o}nben. \end{cases}. \quad (3.10)$$

Majd g mentén frekvenciánként csoportosítjuk a forrásokat. (3.4. ábra) A csoportosítás során megválaszthatjuk az ablak méretét, illetve, hogy alkalmazunk-e valamilyen ablakozó függvényt például Gauss-ablakot. Az összegzés az alábbiak szerint zajlik:

$$Y_L(k) = \sum_{g=d-H/2}^{g=d+H/2} AzL(k,g) \quad (3.11)$$

$$Y_R(k) = \sum_{g=d-H/2}^{g=d+H/2} AzR(k,g), \quad (3.12)$$

ahol $Y_L(k)$ és $Y_R(k)$ a kinyert források frekvenciatartománybeli képe, d az úgynevezett *diszkriminációs index*, amely a csúcsokhoz tartozó azimut érték, H pedig azon ablak szélessége, amelyen belül összegzünk. Az ábrákból látható, hogy ebben az esetben két forrás van jelen a sztereó képben. Az egyik a bal, a másik a jobb csatornában hozzávetőleg $d = 0.7$, illetve $d = 0.4$ azimut értékeknél.

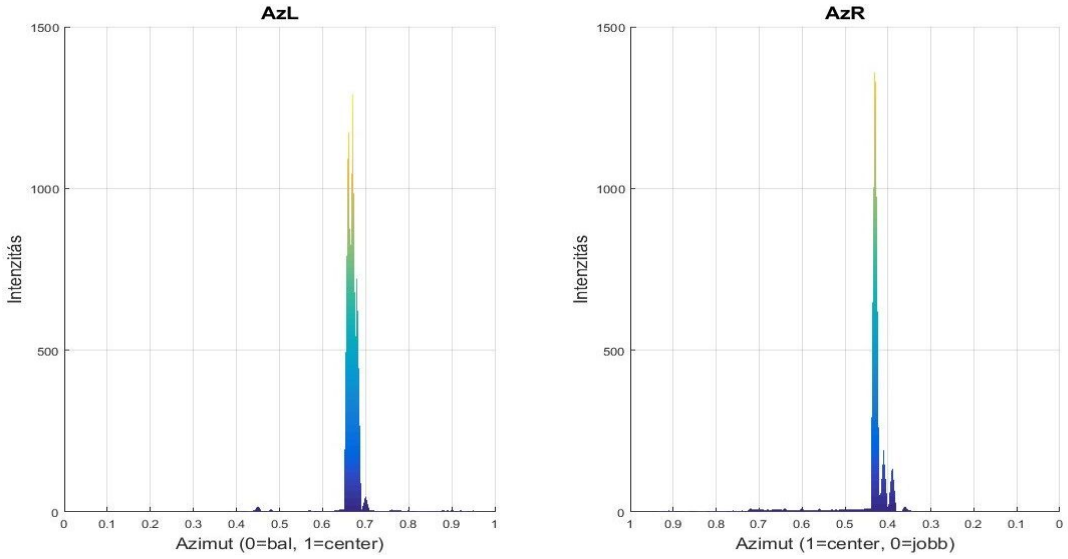


3.4. ábra: A maximumhelyek összegzése

Ezek után következik a csoportosított részek inverz transzformálása. Az így kapott eredmény csak a forrás spektrális összetevőjének intenzitását adja. Fázisinformációként az eredeti $X_L(k)$ és $X_R(k)$ spektrumokat használjuk, a megfelelő frekvencia binek fázisainak kinyerésével.

3.3 Klaszterizáció

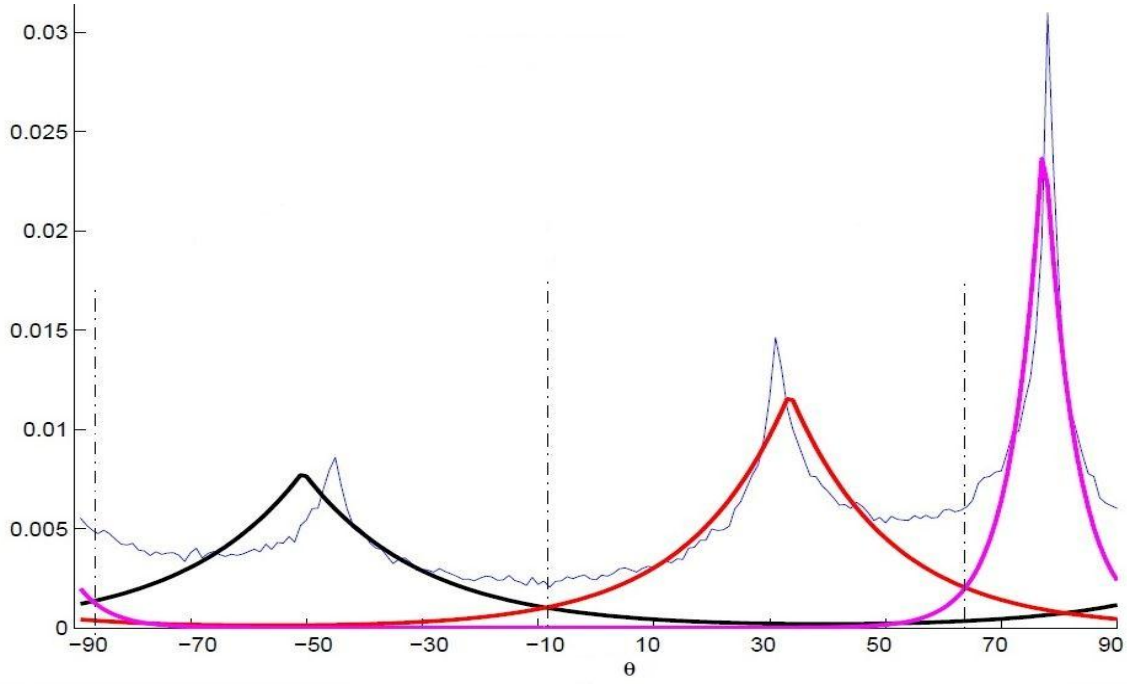
Vizsgáljuk meg a maximum értékek eloszlását az 3.5. ábrán.



3.5. ábra: Azimut-intenzitás sík

Az előző pontban látható volt, hogy a csúcsértékek csoportosítása egy egyszerű összegzés eredménye egy H szélességű ablakon belül a diszkriminációs index környezetében. Ennél kifinomultabb csoportosítás, azaz klaszterizálás is lehetséges.

Ennek egyik megoldása a *Laplacian Mixing Model* (LMM). Az LMM olyan úgynevezett Laplace-függvényeket használ, amelyek a hisztogram maximum értékeire ülnek rá, úgy hogy a legjobban illeszkedjenek az adott eloszláshoz. (3.6. ábra) [15]



3.6. ábra: LMM alkalmazása 3 forrásos minta esetén [15]

A Laplace-függvény az alábbi formulával adható meg:

$$\mathcal{L}(\theta, c, \theta_0) = ce^{-2c|\theta-\theta_0|}, \tag{3.13}$$

ahol c és θ_0 rendre a szélessége és a középértéke a függvénynek. Ebből az LMM definíciója:

$$p(\theta) = \sum_{i=1}^{N} a_i \mathcal{L}(\theta, c_i, \theta_i) = \sum_{i=1}^{N} a_i c_i \, e^{-2c_i|\theta-\theta_i|}, \tag{3.14}$$

ahol a_i a súlyozása a Laplace-függvényeknek. Az alábbi paraméterek az úgynevezett *Expectation-Maximination* algoritmussal becsülhetőek meg. [14] [15]

Ezután következik a források összegzése. Kétféle küszöbszámítási stratégiát lehet használni. Az egyik az úgynevezett *hard threshold*. Ebben az esetben minden megjelenő minta csak egy forráshoz tartozik. A források közötti határokat a szomszédos Laplace-függvények metszéspontja adja meg. Ez látható a 3.6. ábrán. A két küszöbérték közötti

pontok fognak egy forráshoz kerülni. A másik lehetőség a *soft threshold*. Ebben az esetben bizonyos értékek több forráshoz is tartozhatnak. Így ez abban különbözik az előbbitől, hogy a határok egy bizonyos környezetében lévő értékek az összegzés során mindkét forrásban megjelennek. [15]

3.4 Kiterjesztés sokcsatornás hangrendszerre

Az ADRess eljárás kiterjeszthető a sztereóról sokcsatornás, például 5.1-es hanganyagokra is. A megoldás lényege a szomszédos csatornák párba állítása, és azok jeleinek sztereó képként való kezelése. Először is a mély (subwoofer) csatorna jelét figyelmen kívül hagyjuk, az összegzett alacsony frekvenciás hangok nem szolgálnak információval. Ha az audiomix szintén amplitude panninggel készült, feltételezhetjük, hogy egy forrás pozícióját csak két szomszédos csatorna jele határozza meg, a többit pedig figyelmen kívül hagyhatjuk és nullának tekintjük. Így összesen 5 pár hozható létre a szomszédos csatornák párba állításával, illetve külön figyelembe kell venni az elülső bal és jobb csatornát, kizárva ilyenkor a középső (center) sávot. Ugyanis többcsatornás hanganyagok esetén gyakori egy forrás pozícionálása pusztán a bal és jobb csatorna felhasználásával. Majd az immáron 6 sztereó párból a fentieknek megfelelően történik a szétválasztás. [14]

4 Wavelet-transzformáció

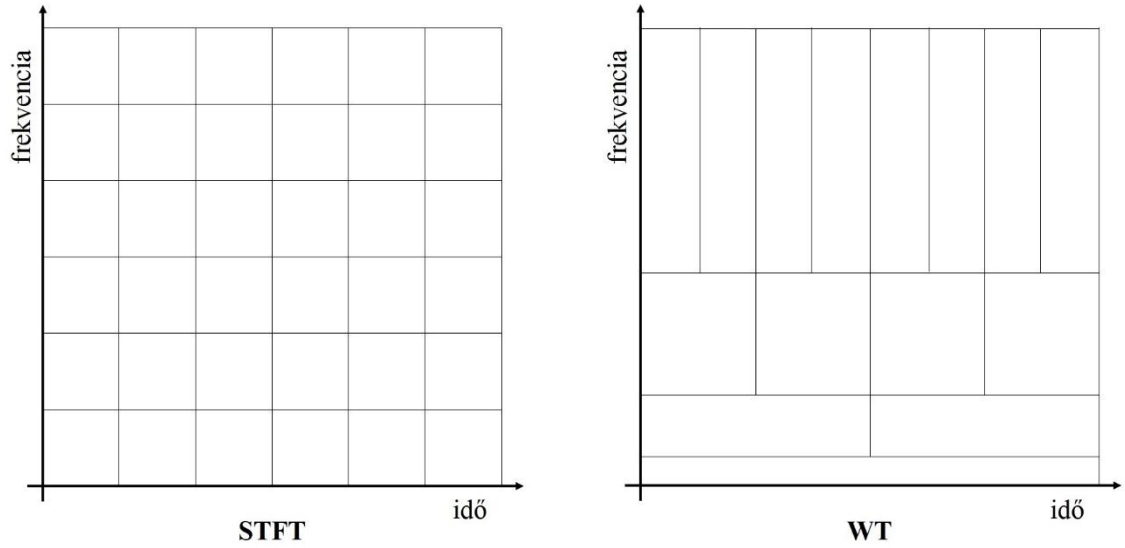
4.1 Bevezetés

Mint az előző fejezetben látható volt, az időtartományi jelek vizsgálata és feldolgozása legtöbbször a frekvenciatartományban történik. Az átjárást a Fourier-transzformáció biztosítja, amelynek jól ismert formája a következő:

$$\mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}\,dt. \tag{4.1}$$

A digitális jelek esetében a transzformáció diszkrét változatát használjuk a minták feldolgozása során, ez a *Discrete Fourier Transform* (DFT). A digitális számítógépes analízis esetén pedig, a számítási idő lerövidítése érdekében a *Fast Fourier Transform* (FFT) kerül alkalmazásra. Az így kapott eredmények azt mutatják meg, hogy milyen frekvencia komponensek vannak jelen a jelben, de az időbeli elhelyezkedésükről nem adnak információt. A sok esetben szükséges idő-frekvencia ábrázolást úgy kaphatjuk meg, ha az FFT-t ablakozva, időszegmensenként végezzük el, így jutunk az előző fejezetben már használt STFT-hez. Az ablak mérete befolyásolja az ábrázolás felbontását. Nagy ablak választása esetén a frekvenciatartományban jó, de időtartományban gyenge felbontást kapunk, még rövid ablak esetén ezek fordítottja áll fenn, de mindig lineáris felbontásban kapunk információt az adott jelről az idő- és frekvenciatartományban is. Azonban sok jel ennél rugalmasabb megközelítést kíván és adott jel esetén különböző felbontást kívánunk elérni különböző frekvenciák esetén, de a transzformáció során az ablak mérete nem változtatható. Ennek következtében minden frekvencia komponensre ugyanazt az ablakot alkalmazzuk, tehát az analízis felbontása ugyanaz lesz minden helyen az idő-frekvencia síkon. [16]

Erre kínál megoldást a Wavelet-transzformáció, amely nem rögzített méretű ablakokkal dolgozik, így nagy frekvenciák esetén nagy idő és alacsony frekvencia, alacsony frekvenciákon pedig alacsony idő és nagy frekvenciás felbontást ad. (4.1. ábra) Ebben a tekintetben hasonlít az emberi halláshoz, amely hasonló idő-frekvencia karakterisztikát mutat. [17]



4.1. ábra: Az STFT és WT együtthatóinak eloszlása az idő-frekvencia síkon

A wavelet szó a transzformációhoz használt függvényekre utal, amelyekre nézve követelmény, hogy legyenek jól lokalizálhatók az időskálán, azaz véges tartójú bázisfüggvények legyenek. A Fourier-transzformációhoz tartozó szinusz és koszinusz függvényekre ez nem igaz, hiszen mindkét függvény az egész (-∞,+∞) intervallumon értelmezve van. A speciális tulajdonsága azonban a waveleteknek az, hogy előállíthatók egy úgynevezett *mother wavelet*ből (anyawavelet). Ez az alapfüggvény egy kisebb hullám (egy pulzus). A transzformáció során ezen függvény átskálázott és eltolt változatait használjuk, így ezek jelentik ebben az esetben az ablakozást. Ebben a felfogásban a waveleteket az idő-frekvencia sík helyett sokkal inkább az idő-skála síkon szokták emlegetni. [16]

A két gyakran használt formája az analízisnek a folytonos (*Continuous Wavelet Transform* - CWT) és diszkrét (*Discrete Wavelet Transform* - DWT) transzformáció. A CWT a jel analizálásakor használatosabb, a DWT pedig a jel feldolgozásakor.

4.2 Folytonos Wavelet-transzformáció

A folytonos Wavelet-transzformáció az alábbi formulával adható meg:

$$W(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-\tau}{s}\right) dt, \tag{4.2}$$

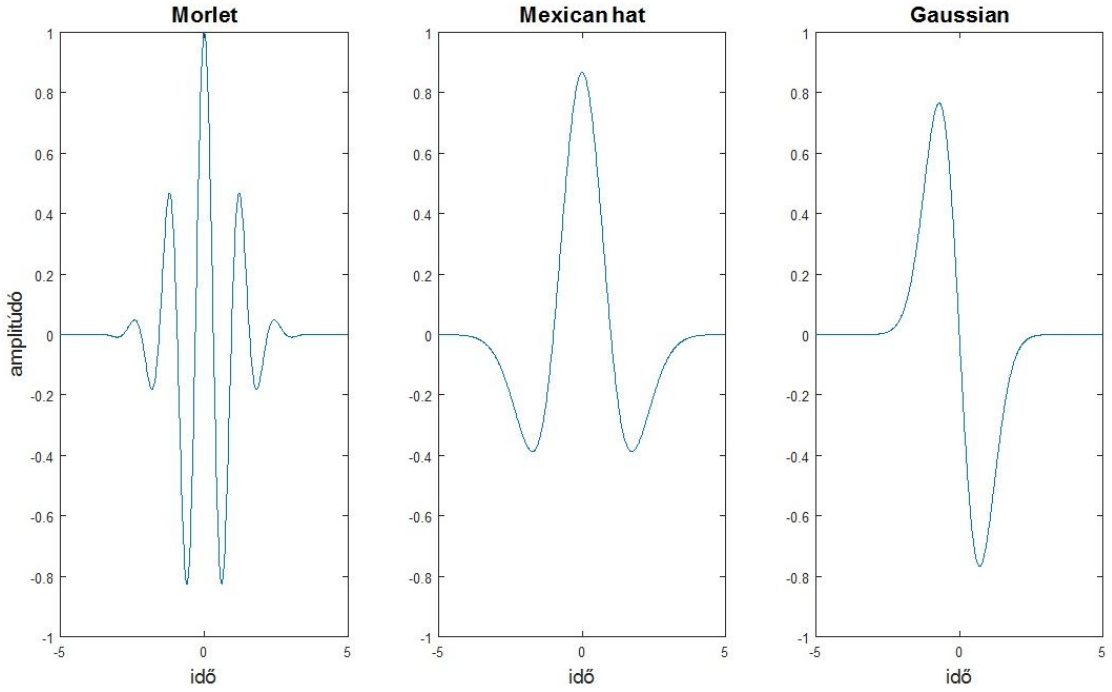
ahol $\psi(t)$ jelenti az mother waveletet (bázis vagy elemző függvényt), s a skálatényezőt, τ az eltolást és $*$ pedig a komplex konjugálást. [19]

Minden wavelet pontosan leírható egy egyenlettel, azok térben és időben lokalizáltak. Mindegyik rendelkezik egy középfrekvenciával, amit az egyenlete határoz meg. A skálázás során, a skála tényező növelésével az eredeti jelalak az idő tengely mentén széthúzódik, az amplitúdó tengely mentén pedig zsugorodik. Így a wavelet frekvenciája csökken. A waveletekre az alábbi kikötéseket kell tenni:

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0 \tag{4.3}$$

$$\psi(\pm\infty) = 0, \tag{4.4}$$

azaz néhány oszcilláció után a végtelenben eltűnő jelet kapjunk. (4.2. ábra) [16]

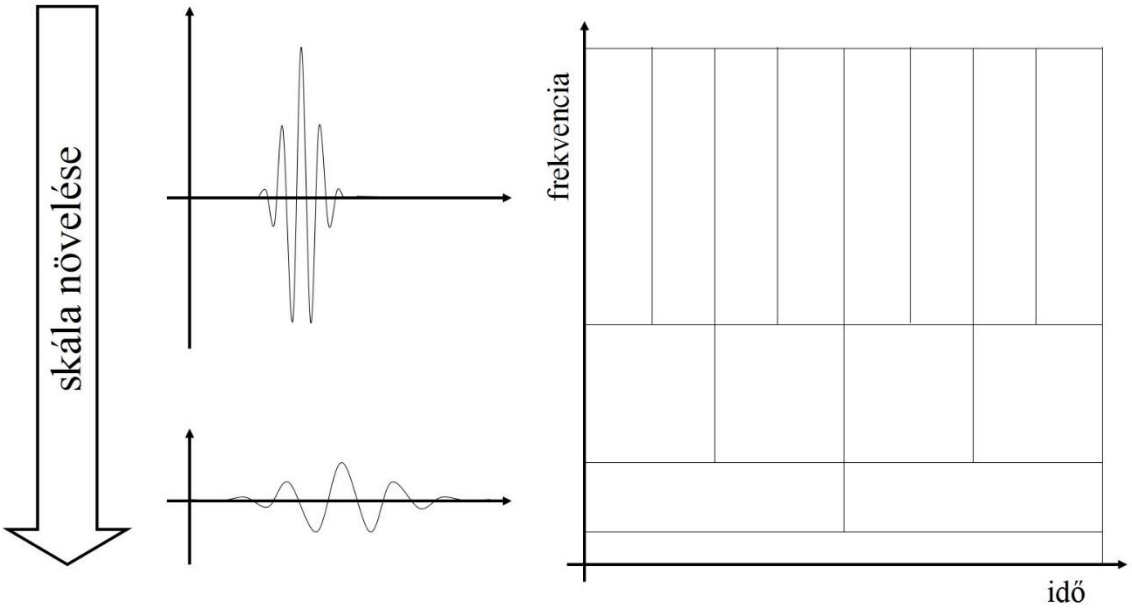


4.2. ábra: Különböző mother waveletek

A waveletek lehetnek komplexek és valósak. A komplexek általában a jelek amplitúdójáról és fázisáról adnak információt, még a valósak a csúcsok, szakadások

azonosítására, finomabb felbontására adnak lehetőséget. Az elemző wavelet típusának megválasztása némiképp önkényes, a cél, hogy az alakja minél jobban illeszkedjen, igazodjon a vizsgálandó jelfolyamhoz.

A (4.2) egyenlet alapján látható a transzformáció menete, amelynek során a waveletet összehasonlítjuk a jel egy részletével és korrelációt számolunk, majd eltoljuk a waveletet és újra számolunk, az eltolást egészen a jel végéig tesszük. Majd növeljük a skála tényezőt és elölről végezzük a műveletet. A transzformálás során minden skála értékre, minden eltolás esetén elvégezzük a korrelációs számítást. A skálát minél tovább növeljük az elemző függvény frekvenciája annál kisebb lesz, így annál pontosabb felbontást kapunk az alacsony frekvenciák esetében, de az idő felbontás rovására. (4.3. ábra) [19]



4.3. ábra: Wavelet analízis ablakozása

A folytonos transzformációval a jelek nemlinearitásai és szakadásai jobban megjeleníthetőek, a gyorsan változó tranziensek elemezhetőek vele.

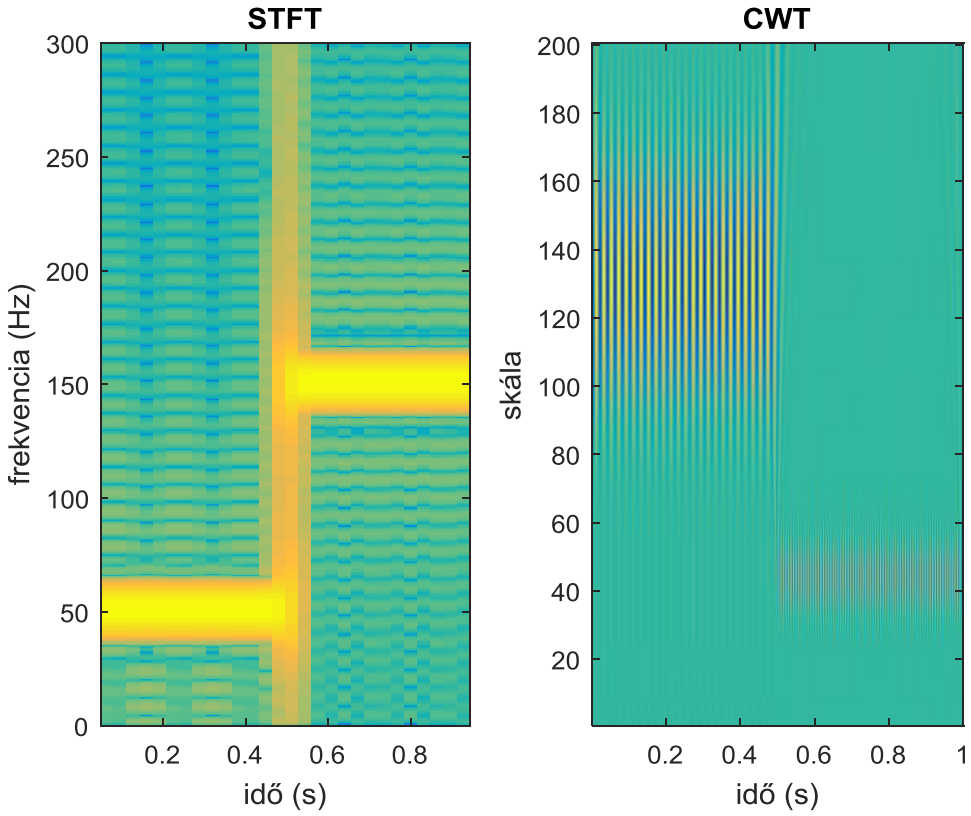
Az eljárás eredménye látható a 4.4. ábrán. A tesztjel egy 50 Hz-es, majd azt követő 150 Hz-es szinuszos jelből áll 0.5-0.5 másodperc időtartamig. Az STFT esetén ez egyértelműen megjelenik a diagramon. CWT esetén a frekvencia helyett skála tengely szerepel. Mivel a skála fordított arányban áll a frekvenciával, ezért az alacsony

frekvenciás komponensek nagyobb értéknél szerepelnek és ott nagyobb felbontással. (4.4. ábra) [18]

A frekvencia és a skála között az alábbi kapcsolat írható fel:

$$s = \frac{F_c}{f \cdot \Delta}, \tag{4.5}$$

ahol s a skála, f a frekvencia, F_c az alkalmazott wavelet középfrekvenciája, Δ pedig a mintavételi periódus. [20]



4.4. ábra: STFT és CWT összevetése

4.3 Diszkrét Wavelet-transzformáció

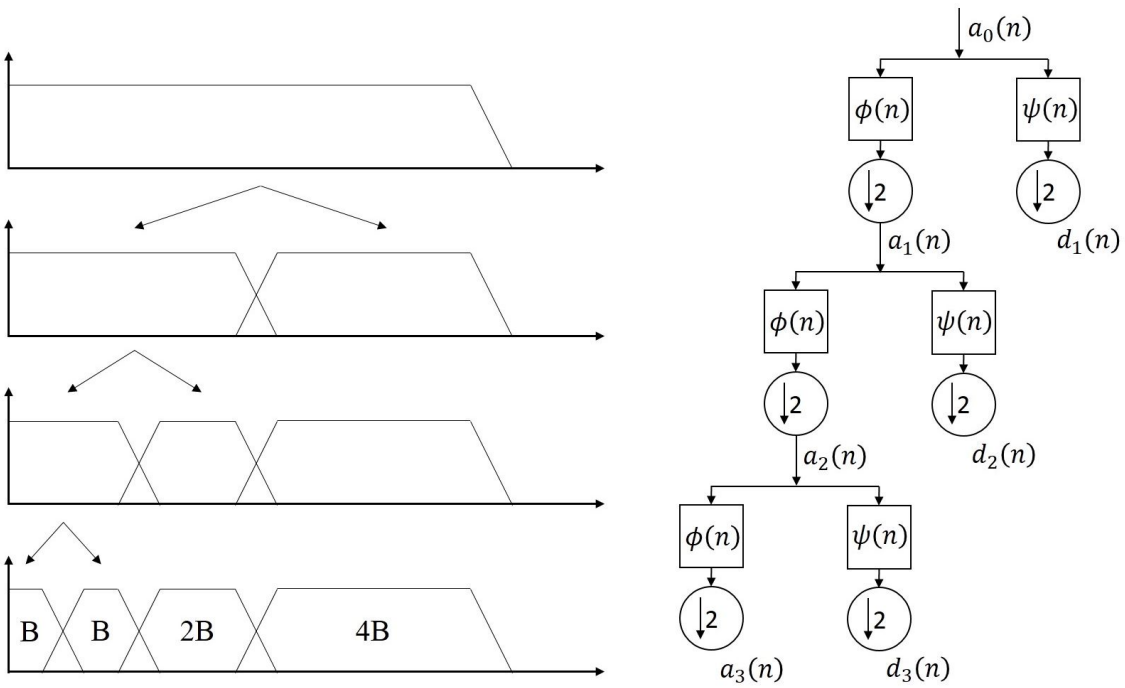
A mintavételezett jelek digitális számítógépes feldolgozását nem teszi lehetővé a folytonos transzformáció, és mint a Fourier-analízis esetében is, kidolgozásra került az eljárás diszkrét megfelelője. A CWT esetében redundáns összetevőket is kapunk, valamint megvalósíthatósága lassú, ezért gyorsabb, hatékonyabb számítás kell. A diszkretizálás során cél, hogy a végtelen sok paraméterből véges számút kapjunk.

A DWT az alábbi alakban írható le:

$$W(m,n) = 2^{-\frac{m}{2}} \sum_k x[k]\, \psi(2^{-m}k - n), \tag{4.6}$$

ahol m és n a diszkrét skála és eltolás paraméterek, $x[k]$ pedig a mintavételezett vizsgálandó jel. [19]

Az eltolás és a skála paraméterek diszkretizálásával még mindig végtelen számú waveletünk van. Az eltolás természetesen véges, hiszen n diszkrét érték és limitált az analizálandó jel időtartama (hossza). A kérdés, hogy mennyire kell skálázni a waveleteket, hogy analizálni tudjuk a jelet. Ezt úgy tudjuk meghatározni, hogy a sávkorlátolt jelünket lefedjük a waveletek spektrumával a frekvencia tartományban, ahol a waveletek sáváteresztő szűrőként funkcionálnak. A szűrő spektruma minden skálázásnál feleződik. A jel teljes frekvenciatartományát csak végtelen számú sávszűrővel fedhetnénk le. A fennmaradó rész egy aluláteresztő szűrővel tudjuk lefedni. A vizsgálandó jel tehát a transzformáció során egy szűrőrendszeren halad át. (4.5. ábra)



4.5. ábra: DWT megvalósítása szűrőkkel

Mivel sávkorlátolt jeleket vizsgálunk a sáváteresztő szűrő gyakorlatilag felüláteresztő szűrővel helyettesíthető, amely a $\psi(n)$ wavelet függvénynek, az aluláteresztő pedig az úgynevezett $\phi(n)$ skála függvénynek felel meg. Minden szűrő kimenetén a DWT egy újabb együtthatója jelenik meg. Az aluláteresztő szűrő kimenetén az úgynevezett *approximation* együtthatók, a felüláteresztőén pedig az úgynevezett *detail*

együtthatók jelennek meg. Minél több szinten engedjük át a jelet, annál részletesebb skálázást kapunk. A transzformációhoz választott waveletek karakterisztikája az alkalmazott szűrőkben jelenik meg. Mivel egy adott szinten a jelet mindkét szűrőn átengedjük a kimeneten kétszer annyi jel fog megjelenni. Ezen hatás kiküszöbölésére minden kimenet után alulmintavételezzük a jelet. [18] [19]

5 Eredmények

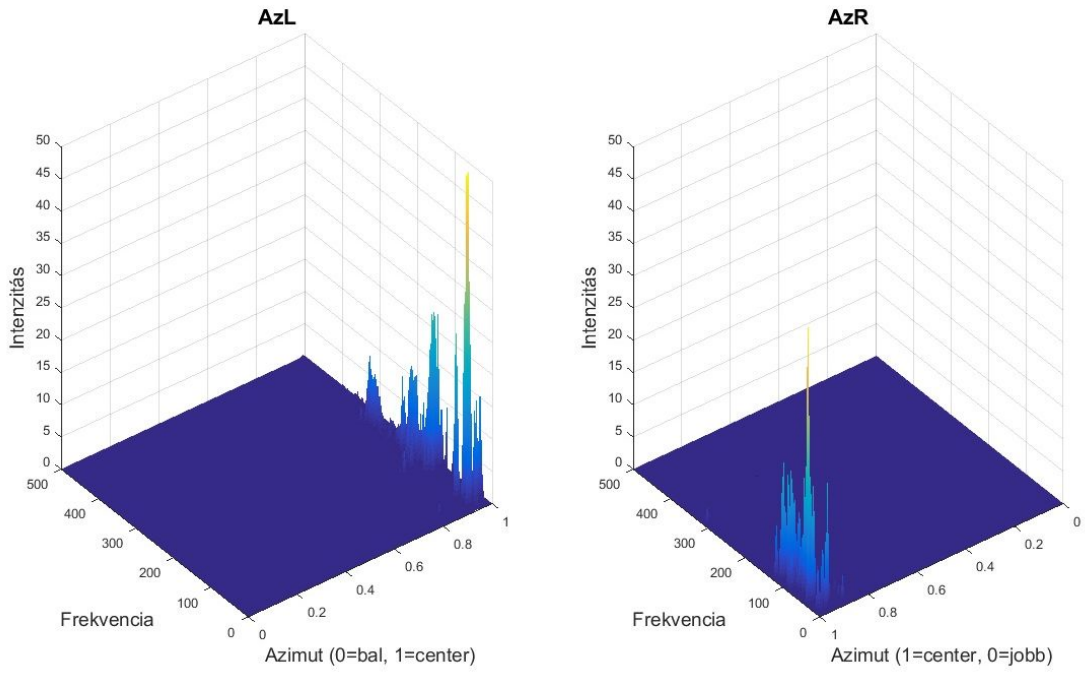
5.1 Bevezetés

A fent bemutatott ADRess algoritmus implementálása, illetve a tesztjelek feldolgozása a MATLAB környezetben történt, ahol a transzformációkat már előre beépített függvények segítették. Az implementáláshoz más szoftvert, illetve hardver egységet nem használtam.

Az időtartományi jelek beolvasását követően a minta sorozat áttranszformálása a frekvenciatartományba, illetve az algoritmus végrehajtása beleértve a különbségképzést, maximumhelyek keresését, csoportosítást, majd a jel időtartományba való inverz transzformálása ablakozva történt.

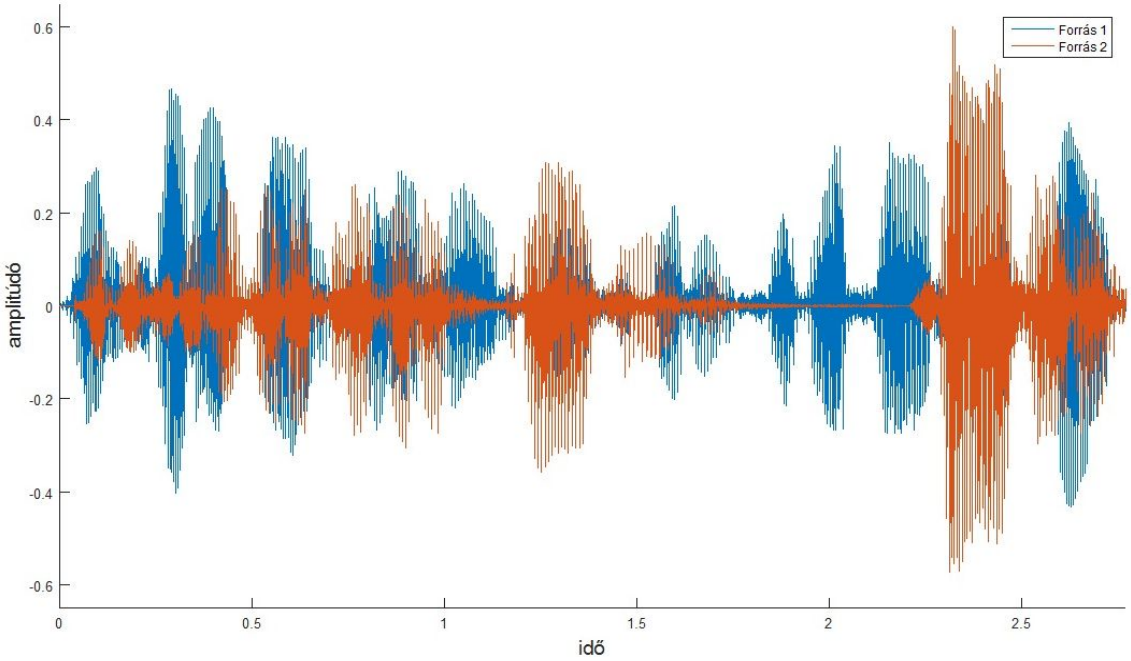
5.2 Fourier-transzformáció használata

Az első esetben a frekvenciatartományba való áttérés a Fourier-transzformációval történt, amely a beépített FFT függvénnyel valósítható meg. Tesztjelnek két egyszerre beszélő férfi hangjából álló audioanyagot hoztam létre. A források elhelyezkedését az 5.1. ábra mutatja. A források frekvencia komponensei a $d = 1$ azimut érték környezetében



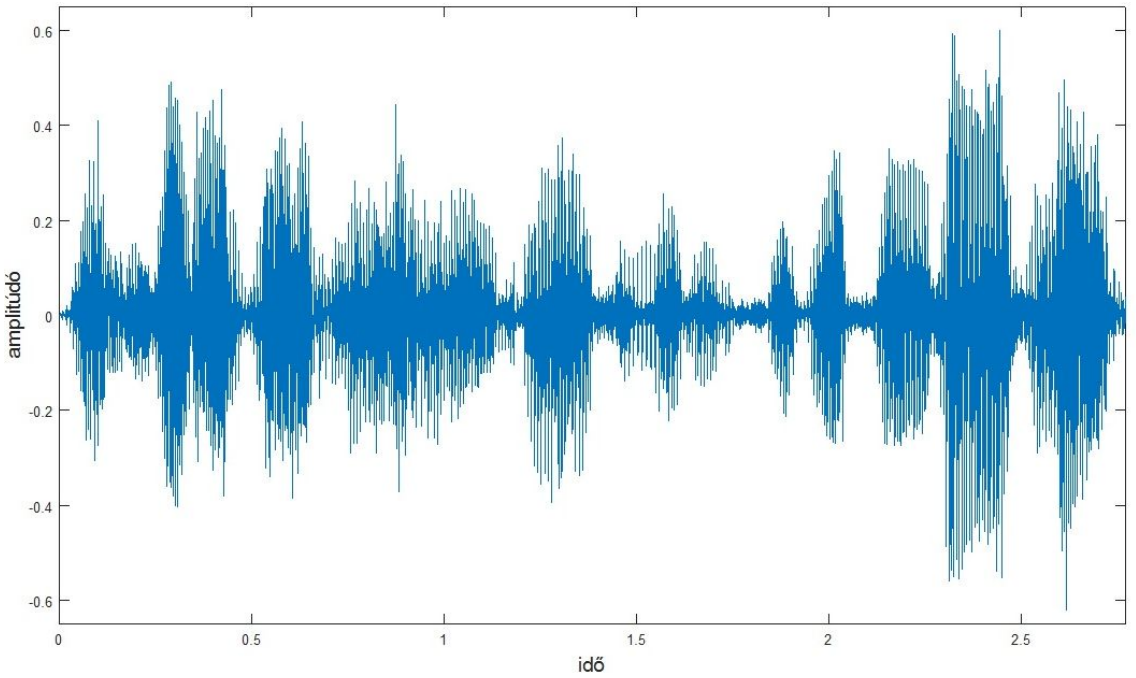
5.1. ábra: A komponensek intenzitása egy adott időszegmensben

találhatóak, amelyből látható, hogy a két forrást a sztereó képben majdnem középre panorámáztam, ez a gyakorlatban egysávos hanghatásnak érzékelhető. A két forrás jelét az 5.2. ábra mutatja.



5.2. ábra: A források időtartományi jele

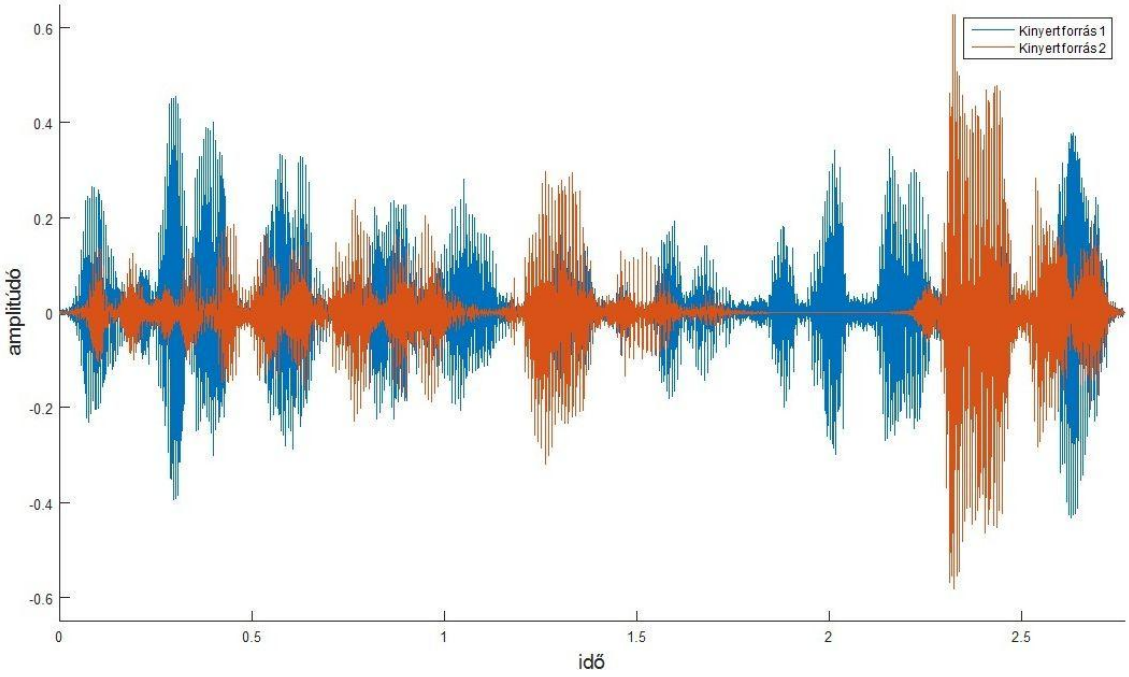
Az 5.3. ábrán pedig látható a sztereó jel bal csatornája, amely ténylegesen az előbbi két jel összegzésének az eredménye, tehát a két csatorna között minimális intenzitás különbség áll fenn.



5.3. ábra: A sztereó jel bal csatornája

Az STFT megvalósításához 4096 mintánként ablakozva, 1024 minta ugrásokkal végeztem el az FFT-t. Az algoritmus során a 0-tól 1-ig futó azimut skála beosztását százados nagyságúra állítottam, a klaszterizáláshoz pedig az egyszerű téglalap alapú összegzést használtam, amihez $H = 10$ szélességű ablakot választottam. A különböző értékek, a szélsőséges eseteket leszámítva ($H = 80$ vagy $H = 1$), nem befolyásolták jelentősen a kimeneteli értékeket. A csoportosított jelek fázisát az ablakozott jel FFT-jének fázisából állítottam be.

Az algoritmus lefuttatása a tesztanyagon az 5.4. ábrán szereplő jeleket eredményezte.



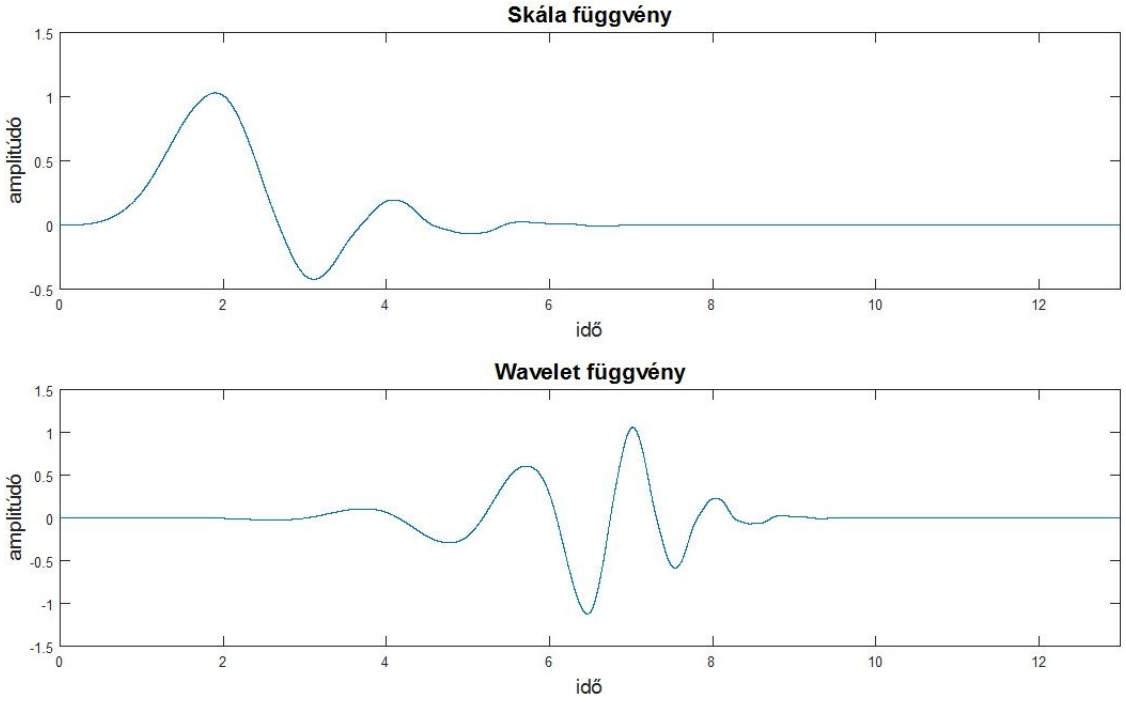
5.4. ábra: A kinyert források időtartományi jele

A kinyert forrásokat meghallgatva az eredmény szubjektíven értékelve kielégítő lett. Tehát elmondható, hogy a bal és jobb csatorna közötti minimális intenzitás különbség hatására is működőképes a módszer, viszont további források azonos időben való jelenléte akkora átfedést eredményez a frekvenciatartományban, hogy a W-diszjunkt ortogonalitási feltétel nem tud teljesülni kielégítően, így a visszaállított jelekben több forrás is erőteljesen megjelenik.

5.3 Wavelet-transzformáció használata

Az eredmények meghallgatása után érzékelhető egy kevés minőségbeli romlás az eredeti jelekhez viszonyítva. Ennek javítására, másodszorra a frekvenciatartományba való áttérés a wavelet transzformációval történt. A MATLAB rendelkezik beépített CWT és DWT függvénnyel, ahol többféle mother wavelet közül lehet választani. A CWT viszont olyan hosszú futási időt eredményez, hogy nem lehet vele feldolgozni a jelet. DWT-vel az FFT-hez hasonlóan megvalósítható az ADRess. A különbség abban áll, hogy a fenti FFT komplex eredményt ad, amelyből kinyerhető a fázis és felhasználható a források csoportosítása során. Addig a DWT esetén csak valós wavelet függvényeket lehet használni, amely szintén valós eredményt ad. Ezért abból fázis információ nem nyerhető ki, pusztán az eredeti jelek előjelét lehet felhasználni. Így ugyanúgy végrehajtható a forrás szétválasztás, mint a Fourier-analízis esetében, de a minőség nem annyira kielégítő. A fő problémát tehát az abszolútérték képzés okozza, amit FFT esetén a fázissal való szorzás állít vissza. DWT esetén pedig csak az eredeti transzformált előjelét tudjuk figyelembe venni, tekintve hogy valós adatok állnak a rendelkezésünkre.

A futtatáshoz az előbbi esetben megegyező paramétereket állítottam be, a DWT-hez pedig az úgynevezett *Daubechies* waveletet használtam.



5.5. ábra: Daubechies wavelet

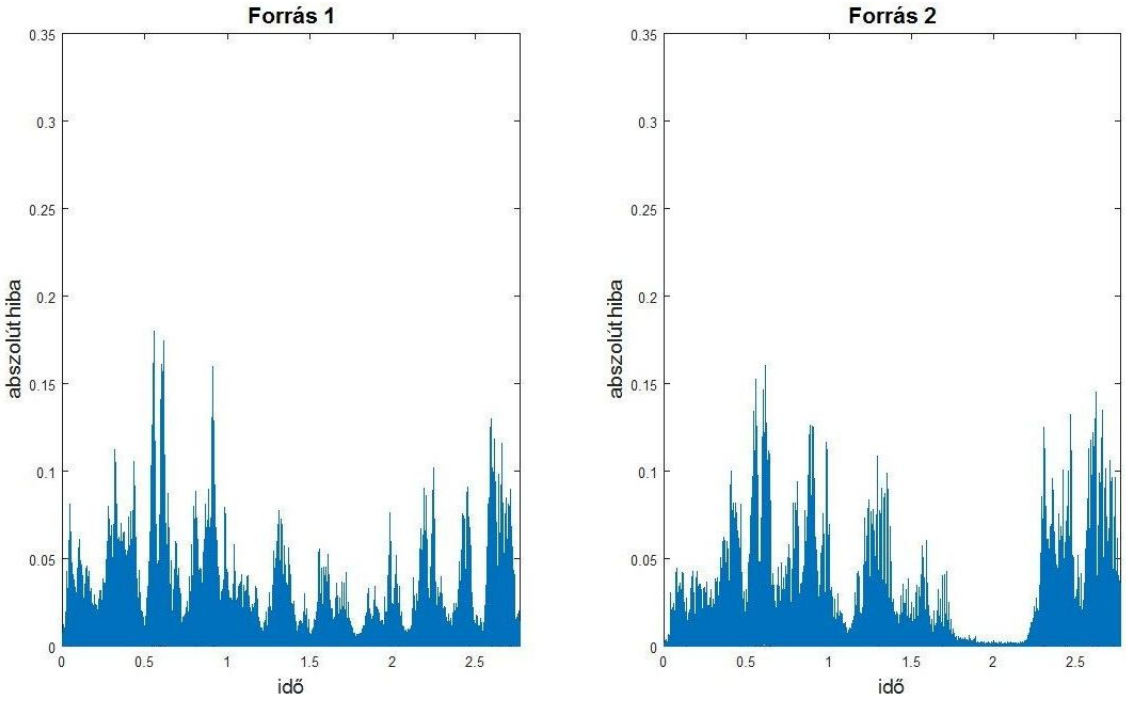
Az 5.5. ábrán látható az alkalmazott wavelet két függvényének alakja. A DWT során az alul- és felüláteresztő szűrő karakterisztikáját ez a két függvény határozza meg.

Az mintasorozatok objektív értékelésére képeztem az alábbi különbséget:

$$x_{hiba} = abs\left(x_{kinyert} - x_{eredeti}\right), \tag{5.1}$$

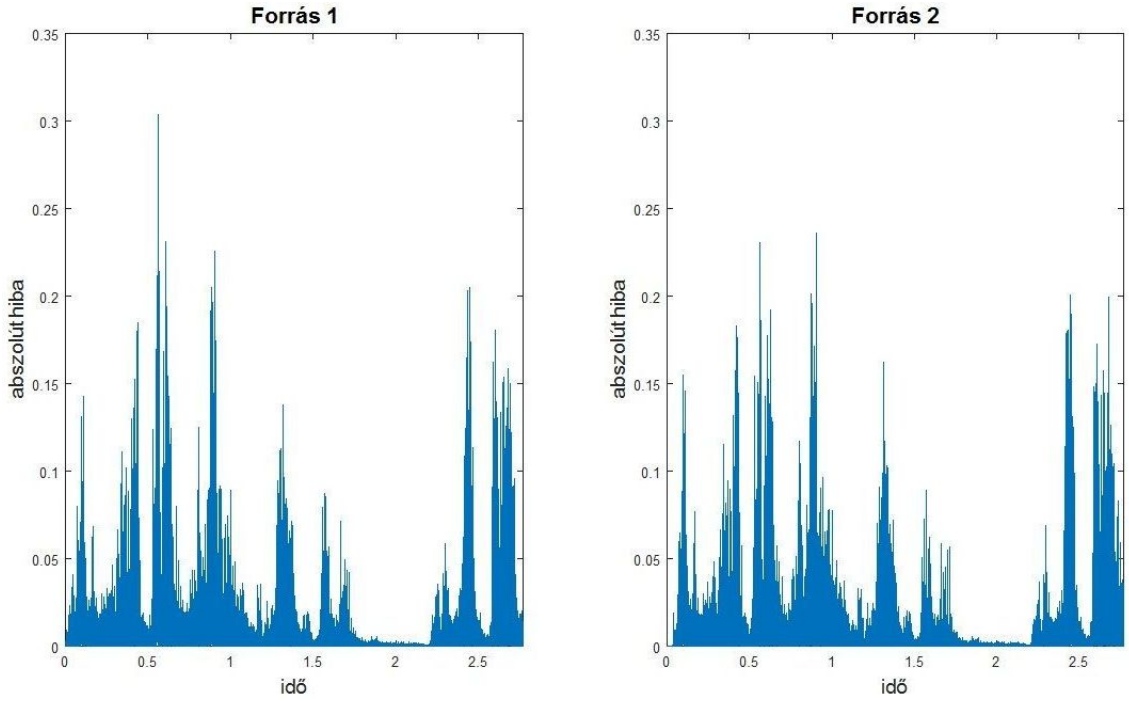
ahol x_{hiba} az algoritmus abszolút hibája egy forrásra mintáira vetítve, $x_{kinyert}$ a kinyert forrás, $x_{eredeti}$ pedig az eredeti forrás mintasorozata. Az abszolút hiba tehát az amplitúdó értékek közötti különbséget mutatja meg a források esetében.

Az alábbi eredmények születtek az FFT-vel végrehajtott ADRess esetén a két forrásra nézve:



5.6. ábra: Abszolút hiba FFT esetén

Illetve a DWT használata esetén:



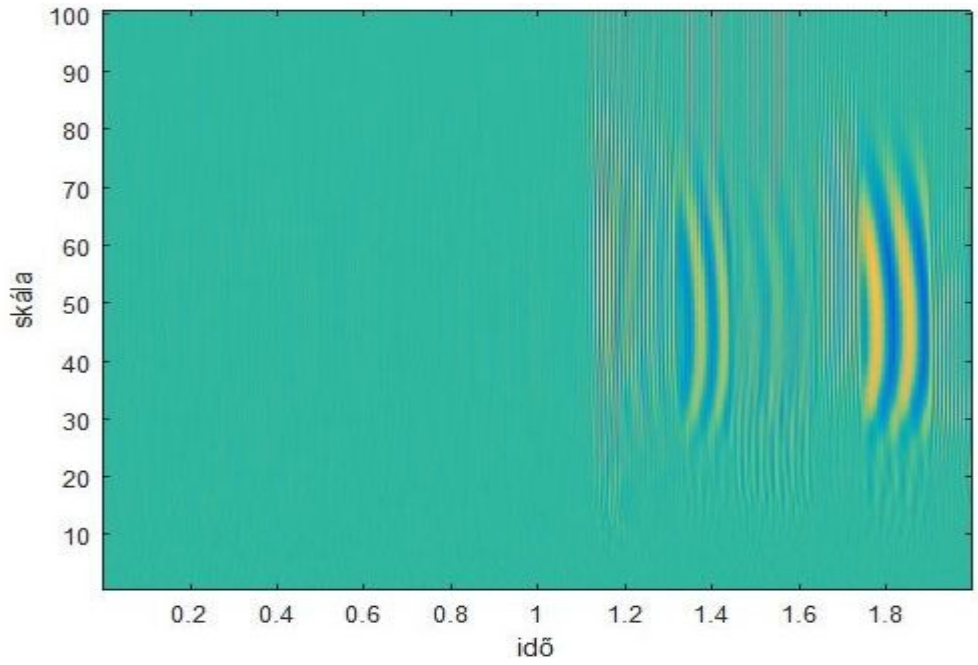
5.7. ábra: Abszolút hiba DWT esetén

A hibaértékek eloszlását vizsgálva látható, hogy a waveletek alkalmazása esetén nagyobb eltérést mutat az eredeti érték a szétválasztottól. Sajnos a transzformáció kicserélésének következtében nem növekedett, sőt kissé romlott a módszer hatékonysága.

A waveleteknek különösen a tranziens jelek esetében kellene jobban teljesíteni. Gyorsan változó jelmintára jó példa a különböző dob ütések. Az algoritmust lefutattam ilyen hanganyagokra is, de a fázisprobléma miatt nem jelentkezett a minőség javulása.

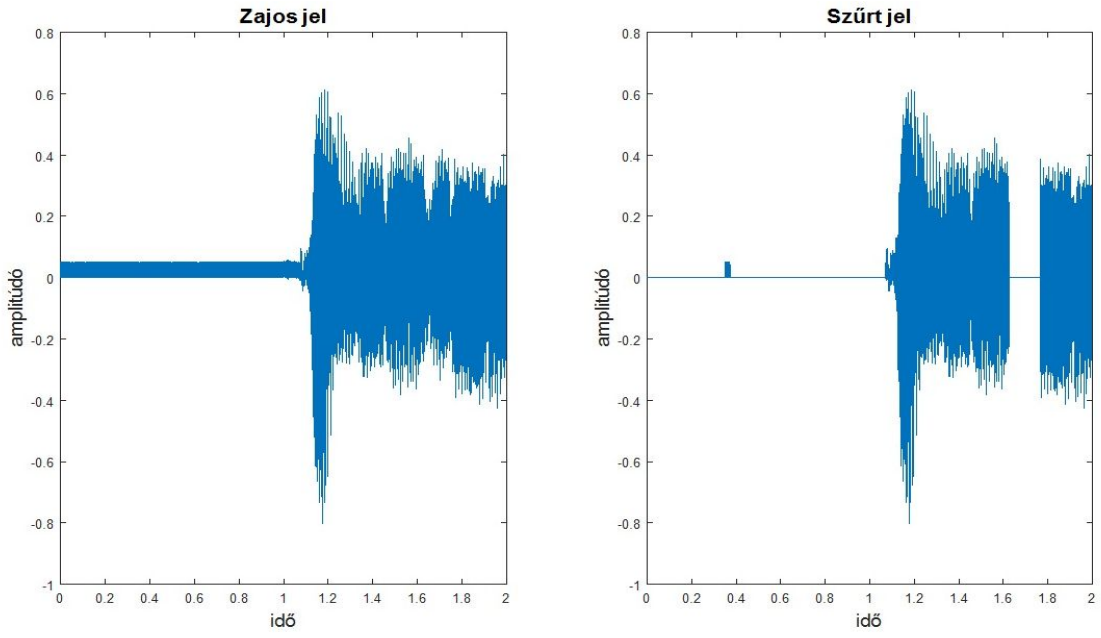
5.4 Waveletek alkalmazása zaj szűrésére

Amennyiben minden forrás független volt a frekvencia tartományban az ADRess kielégítően működik. De amint a különböző források megosztoznak egy frekvencia binen a szétválasztás nem tud eredményes lenni. Ilyenkor a csoportosítás során szórványos frekvenciák is megjelenhetnek és ez konstans zajt jelent a visszaállított időtartományi képben. A megjelenő zaj a forrás szétválasztás eredményességét gyengíti, és főképpen akkor zavaró, mikor az adott sávon épp nem hallható semmi és a forrás még nem lépett be, vagy már ki lépett. Ezen időintervallumok szűrésére és simítására képezzük az adott sáv wavelet térképét.



5.8. ábra: A zajjal terhelt jel CWT diagramja

A cél az, hogy észleljük az energia értékének megnövekedését illetve csökkenését, amelyeket az adott forrás belépésének, kilépésének tulajdonítunk. A térképet vizsgálva válasszuk azt a skála értéket, amelyik a legalkalmasabb az energia szintjének vizsgálatára, majd ezen skála érték mentén végighaladva számolunk energiát, amelyet mindig összehasonlítva az előző értékekkel észlelhetjük, hogy mikor lesz az a forrás hallható az adott sávban és mikor némul el. Ha meghatároztuk ezeket a szakaszokat, akkor nullázhatjuk azokat a helyeket, amelyekben nincs jelen a forrás. [13]



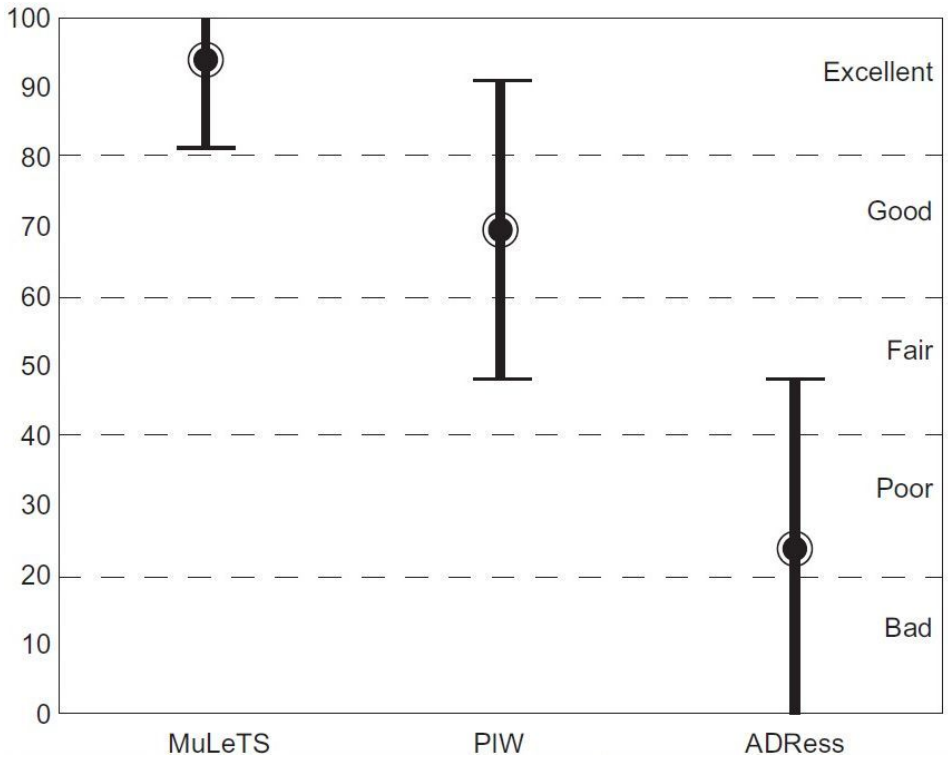
5.9. ábra: A zajos és a szűrt hangminta

Látható, hogy a szűrés után a hasznos jelből is eltűnhetnek részletek, ezeken a paraméterek finomításával lehet javítani. A zaj tehát nem a teljes jeltartományból tűnik el, csupán a csatornák úgymond néma részleteiből, amikor a forrás nem szólal meg.

Tehát ebben az esetben a Wavelet-transzformálást nem az ADRess végrehajtásához használjuk fel, hanem a már kinyert forrás jelének vizsgálatára. Ilyenkor célszerű a folytonos transzformációt alkalmazni, hiszen az pontosabb felbontást ad, de jelfeldolgozáskor, amikor a mintasorozat részleteit folyamatosan transzformálni kell, akkor a DWT képes a számítási idő lecsökkentésére és a műveletek gyors végrehajtására.

5.5 Értékelés

Az ADRess forrás szétválasztó eljárás előnye, hogy egyszerű működésű, könnyen implementálható. Illetve látható volt, hogy a források szoros elhelyezkedése, a bal és jobb csatorna minimális intenzitásbeli különbsége esetén is működőképes. Hátránya viszont, hogy általában gyenge minőséget eredményez és nagy az áthallás az egyes kinyert források között. A meghallgatásos szubjektív tesztek is ezeket támasztják alá. Az 5.5. ábrán egy ilyen próba eredménye látható, ahol a különböző módszereket hasonlították össze.



5.10. ábra: A különböző algoritmusok értékelése a szubjektív tesztek alapján [8]

Látható, hogy az ADRess teljesít a leggyengébben, ettől messzemenően jobb a PIW, illetve a MuLeTS eljárás.

Az algoritmus hatékonyságának feljavítása érdekében megpróbáltam a Fourier-transzformáció helyett waveleteket alkalmazni, ugyanis a különféle mother waveletek az átskálázások során rugalmasabban illeszkednek a vizsgálandó jelalakhoz, így arról pontosabb felbontást tudnak adni. Azonban a várt eredmény elmaradt. Az ADRess esetleges további finomításokat igényel, hogy minél jobban a DWT tulajdonságaihoz tudjon igazodni, illetve a jelek fázisainak hatékonyabb beállítását igényli.

6 Összefoglalás

A dolgozatom célja az audio forrás szeparálás problematikájának a prezentálása volt. Bemutatásra kerültek a különféle hang keverési eljárások a leggyakrabban használt audio technika – sztereofónia – esetében, illetve a lehetséges módszerek az eredeti komponensek visszaállítására. Látható, hogy sokszor nem áll rendelkezésre elegendő információ a hang keverését illetően, ilyenkor csak a meglévő csatornák jeleit lehet felhasználni a jelfeldolgozás során.

Bemutattam az ADRess algoritmust, amely egy viszonylag könnyen megvalósítható, ám kevésbé eredményes eljárás. A hatékonyság növelésének érdekében megismerkedtem a Wavelet-analízissel, bemutattam röviden annak működését és alkalmazásának esetleges pozitív hatásait. A használatával szubjektív minőségjavulás, kevesebb áthallás keletkezés volt várható.

MATLAB környezetben implementáltam az ADRess eljárást a kétféle transzformáció felhasználásával, azonban a várt javulást nem lehetett elérni a waveletekkel, sőt eredményekben alulmaradt a kiváltásra kerülő Fourier-analízissel szemben. A Wavelet-transzformáció jobb felhasználásának érdekében a későbbiekben tovább vizsgálható és fejleszthető az ADRess algoritmus. Illetve jövőbeni cél lehet újabb módszerek megismerése és ezek tesztelése. Az értékelésnél látható volt, hogy a MuLeTS jól teljesített a felmérések során, célszerű lenne a működésének részletes elsajátítása, majd implementálása és tesztelése különböző audioanyagok esetén.

Irodalomjegyzék

[1] E. Collin Cherry: *Some Experiments on the Recognition of Speech, with One and with Two Ears*, The Journal of the Acoustical Society of America, Vol. 25, No. 5, pages 975-979, September 1953

[2] Máximo Cobos Serrano: *Application of Sound Source Separation Methods to Advanced Spatial Audio Systems*, Doctoral thesis, June 2009

[3] Guy J. Brown and Martin Cooke: *Computational auditory scene analysis*, Computer Speech and Language, No. 8, pages 297-336, 1994

[4] Edwin Verheijen: *Sound Reproduction by Wave Field Sythesis*, Thesis, 1997

[5] Dr. Huszák Árpád: *Médiakommunikáció*, előadás diasor, BME-HIT, 2016

[6] Aapo Hyvarinen, Juha Karhunen, and Erkki Oja: *Independent Component Analysis*, JOHN WILEY & SONS, INC., March 2001

[7] Shoko Araki, Hiroshi Sawada, Ryo Mukai and Shoji Makino: *Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors*, Signal Processing, Vol. 87, pages 1833–1847, March 2007

[8] Maximo Cobos, Jose J. Lopez, Alberto Gonzalez and Jose Escolano: *Stereo to Wave-Field Synthesis Music Up-mixing: An Objective and Subjective Evaluation*, ISCCSP, Malta, 12-14 March 2008

[9] Firtha Gergely: *Wavefield szintézis*, Diplomaterv, 2011

[10] Trevor Cox: *Wavefield synthesis: 'souped-up' surround sound*, The Sound Blog https://acousticengineering.wordpress.com/2013/12/08/wavefield-synthesis-souped-up-surround-sound/ (2016. december)

[11] Michael Kunkes: *Three-Dimensional Sound? Iosono Installs Immersive Audio Process at Todd-AO*, Motion Picture Editors Guild https://www.editorsguild.com/fromtheguild.cfm?FromTheGuildid=125 (2016. december)

[12] Dan Barry, Bob Lawlor and Eugene Coyle: *Real-time Sound Source Separation: Azimuth Discrimination and Resynthesis*, AES 117th Convention, San Francisco, 28-31 October 2004

[13] Salil Apte: *Time-Varying Azimuth Discrimination And Resynthesis: A New Method For Music Repurposing*, Thesis, Department of Computer Science, Brown University, September 2005

[14] Keunwoo Choi, Tae Jin Park, Jeongil Seo, and Kyeongok Kang: *Multichannel-To-Wave Field Synthesis Upmixing Technique Based On Sound Source Separation*, AES 52nd International Conference, Guildford, 2-4 September 2003

[15] Nikolaos Mitianoudis and Tania Stathaki: *Overcomplete source separation using Laplacian Mixture Models*, IEEE Signal Processing Letters, Vol. 1, No. 11, June 2004

[16] Nagy László: *A Fourier-analízis elmélete és gyakorlata*, Diplomamunka, 2007

[17] A.Wims Magdalene Mary, Anto Prem Kumar and Anish Abraham Chacko: *Blind Source Separation Using Wavelets*, IEEE International Conference on Computational Intelligence and Computing Research, 2010

[18] C. M. Leavey, M. N. James, J. Summerscales and R. Sutton: *An introduction to wavelet transforms: a tutorial approach*, Insight, Vol. 45, No. 5, pages 344-353, May 2003

[19] C. Valens: *A Really Friendly Guide to Wavelets*, 1999 http://agl.cs.unm.edu/~williams/cs530/arfgtw.pdf (2016. december)

[20] MathWorks: *Documentation - scal2frq (Scale to frequency) function* https://www.mathworks.com/help/wavelet/ref/scal2frq.html (2016. december)