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

Független komponens analízis alapú akusztikus nyalábformálás

TDK DOLGOZAT

Készítette
Tóth Máté
V. évfolyam

Konzulens
dr. Fiala Péter
egyetemi docens

2023.

Tartalomjegyzék

1. Bevezetés **1**

2. A függetlenkomponens-analízis alapjai **3**
2.1. A függetlenkomponens-analízis motivációja . . . 3
2.2. Főkomponens-analízis . . . 4
2.3. Nemnormalitás alapú ICA módszerek . . . 4
2.4. Nemnormalitás mérése . . . 5
2.4.1. Csúcsosság . . . 5
2.4.2. Negentrópia . . . 6
2.5. Az ICA korlátai . . . 6

3. A frekvenciatartományi ICA elmélete **7**
3.1. A frekvenciatartományi ICA motivációja . . . 7
3.2. Konvolutív keverési modell . . . 7
3.3. Konvolutív mixek szeparációs lehetőségei . . . 8
3.3.1. Időtartományi módszerek . . . 8
3.3.2. Frekvenciatartományi módszerek . . . 9
3.4. frekvenciatartományi ICA . . . 10

4. Komplex ICA **12**
4.1. Komplex ICA megvalósítások áttekintése . . . 12
4.2. Komplex valószínűségi alapok . . . 13
4.3. CMN ICA . . . 13
4.3.1. CMN kontrasztfüggvény . . . 14
4.3.2. A fixpontos alogoritmus . . . 14
4.3.3. Adaptable CMN . . . 15
4.3.4. Több forrás szeparációja . . . 17
4.3.5. Az ACMN algoritmus implementációja . . . 17
4.4. RobustICA . . . 17

5. A naiv FD-ICA problémái és kiküszöbölésük **19**
5.1. A naiv FD-ICA problémái . . . 19
5.2. A skálázási bizonytalanság megoldása . . . 19
5.3. Permutációs bizonytalanság megoldása . . . 20

6. FD-ICA alapú mikrofontömbös nyalábformálás **23**
6.1. Több mikrofon kihasználása . . . 24
6.2. Ismert mikrofonpozíciók kihasználása . . . 26

7. Az FD-ICA-Beamformer architektúrája **29**
7.1. Az architektúra leírása . . . 29

7.2. implementációs kérdések 30

8. Kiértékelési metodológia **31**
8.1. A mérési elrendezés bemutatása 31
8.2. Teljesítménymetrikák valós mérési adatokon 33

9. BF-FD-ICA teljesítményének értékelése **36**
9.1. Kiértékeléshez használt általános paraméterek 36
9.2. BF-FD-ICA teljesítményének általános elemzése 36
9.3. STFT ablakhossz hatása a szeparációs teljesítményre 38
9.4. Mikrofonok számának hatása a BF-FD-ICA teljesítményére 38
9.5. Komplex ICA algoritmus megválasztásának hatása 38
9.6. Permutáció korrekciós módszerek teljesítményének összehasonlítása 39
9.7. BF-FD-ICA és az irodalmi FD-ICA teljesítményének összehasonlítása . . . 39
9.8. BF-FD-ICA és klasszikus nyalábformáló algoritmusok teljesítményének vizsgálata 40
9.9. Reflektív környezet hatásának vizsgálata 40

10.Összefoglalás **46**

Köszönetnyilvánítás **47**

Ábrák jegyzéke **48**

Irodalomjegyzék **48**

HALLGATÓI NYILATKOZAT

Alulírott *Tóth Máté*, 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 autentikált 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.

Budapest, 2023. december 12.

Tóth Máté
hallgató

Kivonat

A műszaki gyakorlatban igen nagy múltra tekintő jelfeldolgozási probléma az úgynevezett vak forrásszeparálás (Blind Source Separation, továbbiakban BSS) problémaköre. Ez formálisan azt jelenti, hogy a bizonyos jeleket nincs lehetőségünk "külön-külön" szeparáltan megfigyelnünk, mindössze valamilyen ismeretlen keverő rendszer által alkotott, gyakran zajjal terhelt keveréküket tudjuk mérni, a feladat pedig ezek minél hatékonyabb szétválasztása. Mivel a probléma nagy gyakorlati jelentőséggel rendelkezik, többek között a telekommunikáció, az orvosi jelfeldolgozás, illetve az akusztika területén, az elmúlt évtizedekben jelentős szakirodalma született a BSS probléma különböző változatainak.

Adott emellett egy másik, leginkább az akusztikus és rádiós jelfeldolgozás területén felmerülő probléma, az úgynevezett nyalábformálás (beamforming). Az ezen dolgozat által részletesebben tárgyalt akusztikus nyalábformálás feladata, hogy egy ismert geometriájú mikrofon tömböt felhasználva meghatározzuk a hangtér forrásainak relatív irányát, illetve az adott forrásokra fókuszálva minél inkább szeparálni tudjuk azokat.

A föntebbi két probléma első ránézésre igen hasonlónak tűnik, így felmerülhet a kérdés, hogy a klasszikus BSS technikák mennyire jól adaptálhatók akusztikus beamforming céljára. Ezen dolgozat keretében erre a kérdésre keressük a választ, azaz egy BSS módszerre alapuló akusztikus nyalábformáló megoldás megvalósítását vizsgáljuk mely képes mikrofontömbös akusztikus mérésekből történő forráslokalizáció valamint forrásszeparáció céljára.

Megvizsgáljuk a valós akusztikus környezetekben, így a nyalábformálási probléma esetén is létrejövő, konvolutív keverékekre használható, a szakirodalom által leginkább elfogadott BSS technikát, a frekvenciatartománybéli független komponens analízist (FD-ICA), illetve az ennek alapjául szolgáló komplex független komponens analízis (complex ICA) megvalósításának lehetőségeit. Áttekintjük, hogy a mikrofontömbös mérés sajátosságai (az ismert mikrofonpozíciók, és a források számánál jelentősen nagyobb számú mikrofon) milyen többletinformációkat nyújtanak a klasszikus BSS konfigurációkkal szemben, ezeket hogyan lehetséges a klasszikus BSS módszerek gyengeségeinek kiküszöbölésére felhasználni, illetve megvizsgáljuk a "klasszikus" mikrofontömbös nyalábformálási technikák felhasználhatóságát az FD-ICA keretben.

Ezek alapján javaslunk egy új, mikrofontömbös környezetre adaptált FD-ICA alapú forrásszeparációs algoritmust, majd ezt Python környezetben implementáljuk. A javasolt megoldás szeparációs teljesítményét kiértékeljük valós mikrofontömbös méréseken. Kimérjük a sok mikrofon használatának jelentőségét, illetve összehasonlítjuk a megoldásunk teljesítményét a mikrofontömbös mérési gyakorlatban használt klasszikus beamforming technikák szeparációs teljesítményével.

1. fejezet

Bevezetés

A mérnöki gyakorlatban, a jelfeldolgozás területén gyakran előforduló feladat az ún. vak forrásszeparálás (Blind Source Separation (BSS)) problémaköre. Ez formálisan azt jelenti, hogy a feldolgozandó jeleinket gyakran közvetlenül nincs lehetőségünk megfigyelni, mindössze valamilyen (ismeretlen paraméterezésű) keverő rendszer által létrehozott, egymással, esetleg valamilyen zaj jellegű jelekkel vett keveréküket tudjuk mérni (melyek közvetlenül nyilván nem alkalmasak a feldolgozásra), feladatunk pedig ezek szeparálása. Ez szemléletesen például az úgynevezett koktélparti probléma esetén jelenik meg. Ennek lényege, hogy ha például egy zsúfolt társaságban vagy zajos bárban ülünk, és valakivel beszélgetünk, a fülünk a többi ember beszédével és egyéb háttérzajokkal keverve érzékeli a beszélgetőpartnerünk hangját, mégis igen magas zajszintig képesek vagyunk megérteni a partnerünket. Ez azért lehetséges, mert az agyunk elvégzi a szeparációs feladatot, azaz rá tudunk fókuszálni a megértendő beszédre. Ennek a mechanizmusnak az algoritmikus megvalósítása a Blind Source Separation. A probléma igen jelentős gyakorlati jelentőséggel rendelkezik, többek között a távközlési hálózatok, az orvosi képfeldolgozás, valamint a hangjel-feldolgozás területén is.

A forrásszeparációs problémára a szakma által leginkább elfogadott megoldás a függetlenkomponens-analízis (independent component analysis (ICA)) nevű statisztikai eljárás [4]. Ez egy igen mély irodalommal, valamint hatékony kész implementációkkal támogatott módszer, azonban – ahogy a korábbi saját, illetve irodalmi kutatások megmutatták – közvetlenül csak a (mikrofonos hangfeldolgozási gyakorlatban nem igazán releváns), instant módon és lineárisan kevert forrásokra működőképes.

Egy másik, az akusztikus illetve rádiós jelfeldolgozásban szintén jelentős gyakorlati vonatkozásokkal rendelkező probléma az úgynevezett nyalábformálás (beamforming) problémaköre. Az ezen dolgozat által részletesebben elemzett akusztikus nyalábformálás feladata, hogy egy ismert geometriájú, sokmikrofonos mikrofontömböt felhasználva meghatározzuk a hangtér forrásainak irányát, illetve az irányra fókuszálva minél inkább szeparálni tudjuk azokat.

Észrevehetjük, hogy az akusztikus nyalábformálás lényegében a BSS probléma egy speciális esetének tekinthető, hiszen a feladat itt is a különböző pozíciókban elhelyezett, így a források különböző keverékeit vevő mikrofonok jelei alapján az adott irányból érkező hangjeleket felerősíteni, közben az egyéb irányokból érkező források jeleit minél inkább elnyomni, azaz lényegében a forrásszeparációt megvalósítani. A mikrofontömbös nyalábformálás szakirodalmát vizsgálva azonban azt tapasztalhatjuk, hogy a beamforming algoritmusok jellemzően a mikrofontömb iránykarakterisztikájának alakításával érik el a szeparációt, az ICA módszerekkel ellentétben nem a források függetlenségére koncentrálnak. Felmerülhet tehát a kérdés: hogyan lenne felhasználható az ICA eszköztára a nyalábformálási feladat megoldására, mennyire lenne praktikus egy ICA alapú beamformer

algoritmus megvalósítása, illetve járna-e egyáltalán előnyökkel az ICA módszerek használata a klasszikus nyalábformáló módszerekkel szemben. Meglepő módon a szakirodalomban nem sok kutatással találkozhatunk ebben a témakörben. Amely cikkekben előkerül az ICA beamformingra való használatának lehetősége [25] [19] [18], azok csak az ICA ismert mikrofonpozíciók esetén fennálló beamformerként való értelmezhetőségét vizsgálják az ICA algoritmus valamilyen problémájának kiküszöbölésére, az ICA alapú módszerek sokmikrofonos beamformerként történő használhatóságát nem vizsgáljak. Ez a tény motiválta a dolgozat alaptémáját, a függetlenkomponens-analízis alapú akusztikus nyalábformálás vizsgálatát.

A dolgozat célja tehát egy ICA alapú mikrofontömbös nyalábformáló módszer megalkotása, illetve működőképességének vizsgálata. Mivel a tervezendő eljárásnak a gyakorlatban is használhatónak kell lennie (ahol a szeparálandó mixek lényegében mindig konvolutív jellegűek), az egyszerűbb klasszikus ICA használata ebben az esetben biztosan nem vezet eredményre. Szerencsére az ICA módszer adaptálható a valóságban előforduló konvolutív vak forrásszeparációs problémára is, ez az adaptáció a viszonylag jelentős irodalommal rendelkező frekvenciatartományi ICA (FD-ICA) [19].

A dolgozat második fejezetében áttekintjük a klasszikus valós értékű független komponens alapjait, majd a harmadik fejezetben az erre épülő frekvenciatartományi ICA elméleti alapjait, illetve megvalósításának lépéseit. Amint azt látni fogjuk, az FD-ICA megvalósításához komplex értékű bemeneten is működőképes ICA algoritmusra van szükség, így a negyedik fejezetben áttekintjük a komplex értékű ICA elméleti alapjait, illetve részletesen bemutatjuk a jelenleg is state-of-the-artnak tekinthető ACMN ICA [22] algoritmus működését. Az ötödik fejezetben a naiv FD-ICA megvalósítás során előkerülő, az ICA természetéből adódó problémákat tekintjük át, illetve bemutatjuk a szakirodalomban fellelhető módszereket ezek megoldására. Ezek alapján a hatodik fejezetben megalkotunk egy a gyakorlatban is működő FD-ICA architektúrát. A hatodik fejezet további részében áttekintjük, hogy a mikrofontömbös nyalábformálás milyen szempontból tekinthető az ICA speciális esetének, azaz milyen extra információk állnak a rendelkezésünkre a klasszikus ICA problémához képest (a források számánál jelentősen több mikrofon, illetve ismert mikrofonpozíciók), majd ezeket kihasználva olyan, újszerűnek tekinthető módszereket (dimenzióredukciós FD-ICA, illetve beamforming alapú permutációs módszer) mutatunk be, melyekkel a klasszikus FD-ICA szeparációs teljesítménye jelentősen javítható sokmikrofonos mikrofontömbös esetére. Ezután a hetedik fejezetben a bemutatjuk a mikrofontömbös nyalábformálásra optimalizált, általunk megalkotott BF-FD-ICA architektúrát. A nyolcadik fejezetben felvázoljuk, hogy miért lehet előnyös a beamforming algoritmusok teljesítményét valós méréseken kiértékelni, majd bemutatjuk a megalkotott BF-FD-ICA módszer értékeléséhez használt valós, 5 forrásos, illetve 24 mikrofonos mikrofontömbös mérési elrendezést. Ezután a fejezet hátralevő részében bemutatjuk a szeparációs algoritmusok valós méréseken történő kiértékelésének nehézségét, majd megalkotunk egy erre az esetre is használható kiértékelési metrikát. Végül a kilencedik fejezetben kiértékeljük a BF-FD-ICA szeparációs teljesítményét a valós méréseinken, illetve összehasonlítjuk annak teljesítményét néhány klasszikus beamforming algoritmus, valamint az irodalmi FD-ICA szeparációs teljesítményével.

2. fejezet

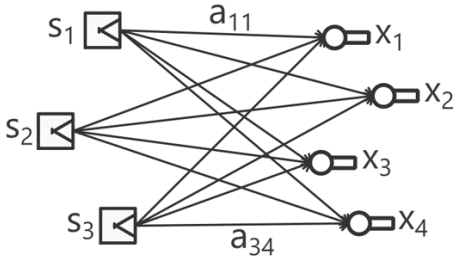
A függetlenkomponens-analízis alapjai

2.1. A függetlenkomponens-analízis motivációja

A vak forrásszeparációs probléma megoldásra talán legszélesebb körben alkalmazott módszer a függetlenkomponens-analízis (ICA) [4]. Az ICA egy statisztikai módszer, mely bizonyos feltételek teljesülése mellett képes diszkrét idejű jelek instant (késleltetésmentes) lineáris kombinációiból az eredeti jeleket becsülni. Formálisan a következőt írhatjuk fel: Legyen adott N darab független valószínűségi változó (melyek lényegében maguk a forrásainkat reprezentálják, a források adott t időpillanatban vett értékeit ezen valószínűségi változók egy-egy realizációjaként modellezzük), ezek legyenek rendre $S_1, S_2, \ldots, S_N$. A vett (ekkor már kevert, így nyilván nem független) jeleket, melyek száma a klasszikus ICA modellben megegyezik a források számával, modellezzük szintén N darab valószínűségi változóval, melyek legyenek rendre $X_1, X_2, \ldots, X_n$. Az ICA modell szerint a források illetve a megfigyelt jelek között a következő lineáris összefüggés teljesül:

$$\underline{X} = \mathbf{A}\underline{S}, \tag{2.1}$$

ahol $\underline{X} = [X_1, X_2, \ldots, X_N]^{\mathrm{T}}$, a megfigyelt jeleket modellező valós változókból álló álló vektor, $\underline{S} = [S_1, S_2, \ldots, S_N]^{\mathrm{T}}$ az ismeretlen forrásjeleket reprezentáló vektor, $\mathbf{A}$ pedig a szintén ismeretlen, konstans, valós elemekből álló keverőmátrix. Az instant keverési modellt a 2.1 ábra szemlélteti.



2.1. ábra. Az instant keverési modell

A feladat a klasszikusan $\mathbf{W}$-vel jelölt szétválasztó (demix) mátrix meghatározása, melyre az

$$\underline{S} \approx \underline{Y} = \mathbf{W}\underline{X} = \mathbf{WA}\underline{S} \tag{2.2}$$

alakban visszaállított $\underline{Y}$ források leginkább közelítik az eredeti $\underline{S}$ forrásokat, azaz $\mathbf{W} \approx \mathbf{A}^{-1}$ teljesül. A probléma nehézségét tehát az adja, hogy mind az $\mathbf{A}$ mátrix elemei, mind

az eredeti látens S_i források tetszőleges időpillanatokban mért értékei ismeretlenek, tehát a keverőmátrix direkt invertálására nincsen módunk. Az ICA módszerek ezt a problémát nem is a keverőmátrix direkt becslésére vezetik vissza, hanem azt a transzformációt keresik, mely a visszaállított jelek valamilyen metrika szerint mért függetlenségét maximalizálja. Belátható, hogy a visszaállított $\underline{Y}$ források a tökéletes függetlensége garantálja, hogy az S_i források sorrendje és konstans faktorral történő skálázásuk erejéig azok tökéletesen visszaállíthatók [10], így az ICA módszerek szükségszerűen az eredeti forrásokat becslik. Az ICA módszerek feladata tehát egy, a függetlenséget mérő költségfüggvény megalkotása, ezen a költségfüggvény optimalizálása. A szakirodalomban többféle megközelítés született a függetlenség mérésére, a lényegesebbek a nemnormalitás (non-Gaussianity), a kölcsönös információ mérése, valamint a tenzoriális illetve Maximum-Likelihood becslés alapú módszerek, melyek közül jelen dolgozatban a nemnormalitás alapú módszereket tárgyaljuk részletesebben.

2.2. Főkomponens-analízis

Az ICA módszerek jelentős része feltételezi, hogy a szeparálandó jelek fehérítettek, azaz zéró várható értékűek, egységnyi szórásúak, valamint dekorreláltak. Ennek oka, hogy a fehérített $\underline{Z}$ valószínűségi változókra igaz a következő összefüggés:

$$\mathbb{E}[\underline{Z}\underline{Z}^{\mathrm{T}}] = \mathbb{I}, \tag{2.3}$$

azaz kovariancimátrixuk egységnyi. Egyszerűen belátható, hogy ebben az esetben a függetlenítés biztosan egy ortonormált mátrixszal történő szorzással elvégezhető lineáris transzformáció. A fehérítést a főkomponens-analízis (PCA) technika használatával tudjuk végrehajtani. A PCA első lépése a bemenet centrálása:

$$\underline{\hat{X}} = \underline{X} - \mathbb{E}[\underline{X}]. \tag{2.4}$$

Ezután kiszámítjuk a centrált változók kovarianciamátrixát:

$$\mathbf{\Sigma} = \mathbb{E}[\underline{\hat{X}}\underline{\hat{X}}^{\mathrm{T}}], \tag{2.5}$$

majd meghatározzuk a $\mathbf{\Sigma}$ kovarianciamátrix sajátfelbontását ($\mathbf{\Sigma}$ négyzetes szimmetrikus mátrix, így sajátfelbontása biztosan létezik):

$$\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} = \mathbf{\Sigma}, \tag{2.6}$$

Ahol $\mathbf{Q}$ egy $\mathbb{R}^{N\times N}$-beli ortogonális mátrix, melynek i-edik oszlopa az i-edik sajátértékhez tartozó sajátvektor, $\mathbf{\Lambda}$ pedig a a sajátértékekből álló diagonális mátrix, ahol Λ_{ii} az i-edik legnagyobb abszolút értékű sajátérték. A fehérítő transzformáció ekkor a következő alakban áll elő:

$$\underline{Z} = \mathbf{\Lambda}^{\circ-\frac{1}{2}}\mathbf{Q}^{\mathrm{T}}\underline{\hat{X}} \tag{2.7}$$

Ahol $\underline{Z}$ a fehérített jelekből álló vektor, $\mathbf{\Lambda}^{\circ-\frac{1}{2}}$ pedig a $\mathbf{\Lambda}$ mátrix elemenkénti $-\frac{1}{2}$-re emelését jelöli.

2.3. Nemnormalitás alapú ICA módszerek

Az ICA módszerek legjelentősebb kérdése tehát a függetlenség mérése, illetve az ezt optimalizáló módszer megválasztása. A függetlenség egy lehetséges mértéke a forráseloszlások nemnormalitása (Non-Gaussianity). A nemnormalitás használatának alapötletét A cent-

rális határeloszlás-tétel (CHT) adja, ennek Lindeberg-féle alakja [12] ugyanis kimondja, hogy tetszőleges, független valószínűségi változók összegének eloszlása a normális eloszláshoz tart a tagszám növelésével. Vegyük észre, hogy ha a Z fehérített jelet a $W_{\text{opt}} = A^{-1}$ ideális demix mátrix egy $\underline{w}^T$ sorának α-szorosával skalárszorozzuk, a következőt kapjuk:

$$\alpha S = \alpha \underline{w}^{\mathrm{T}} \underline{Z} = \alpha \underline{w}^{\mathrm{T}} \mathbf{A} \underline{S} = \underline{q}^{\mathrm{T}} \underline{S}, \tag{2.8}$$

ahol $\underline{q}$ egy $\mathbb{R}^N$-beli „one-hot" vektor (egyik komponense α, a többi 0), S pedig az egyik független komponens. Másrészről a CHT értelmében S nemnormalitása akkor lesz biztosan maximális a Z fehérsége miatti $\|\underline{w}\| = 1$ feltétel mellett, ha a $\underline{w}^{\mathrm{T}}\mathbf{A} = \underline{q}$ one-hot, hiszen ellenkező esetben S az S_i-k lineáris kombinációja, azaz független valószínűségi változók összege lenne. (Megjegyzendő, hogy az állítás megfordítása általában nem igaz, hiszen S_i-k nemnormalitása különbözhet, valamint az $\alpha \in \mathbb{R}^+$ tényező nem elhagyható a fehérítés során végrehajtott skálázás miatt.) Mivel a nemnormalitás egy adott valószínűségi változóra (jelre) önmagában értelmezhető, ezt használva a független komponensek különkülön kinyerhetők a kevert jelből (feltéve, hogy garantálni tudjuk, hogy az így szeparált komponensek mind különböző bemeneti jelhez tartozzanak). Ezek alapján felvázolhatjuk nemnormalitás alapú ICA módszerek működését:

1. Az $\underline{X}$ bemenetet a 2.2 fejezetben tárgyalt PCA transzformáció használatával fehérítjük.

2. A kapott $\underline{Z}$-re kiszámítjuk a demix mátrix egy $\underline{w}$ sorát a következő összefüggés szerint:
$$\underline{w} = \underset{\|\underline{w}\|=1}{\operatorname{argmax}}\, G(\underline{w}^{\mathrm{T}} \underline{Z}), \tag{2.9}$$
ahol G egy nemnormalitást mérő függvény.

3. A 2-es pont N-szeri ismétlésével meghatározzuk a demix mátrix többi sorát, azon feltétel betartásával, hogy $\underline{Z}$ fehérsége miatt $\mathbf{W}$ ortogonális mátrix, ezzel biztosítva hogy $\mathbf{W}$ sorai különböző független komponenseket nyernek ki.

4. Meghatározzuk a független komponenseket a $\underline{Y} = \mathbf{W}\underline{\mathbf{Z}}$ összefüggés szerint.

2.4. Nemnormalitás mérése

A nemnormalitás alapú ICA módszerek fő kérdése tehát a nemnormalitás kvantitatív mérése. Erre az ICA gyakorlatban a leginkább elterjedt metrikák a következők:

2.4.1. Csúcsosság

Egy valószínűségi változó csúcsossága (negyedrendű kumulánsa) azt mutatja meg, hogy a valószínűségi változó sűrűségfüggvénye mennyire „lapos" a normális eloszlás sűrűségfüggvényéhez képest. A csúcsosságot zéró várható értékű (centrált) X valószínűségi változókra a következő összefüggés adja meg:

$$\operatorname{Kurt}(X) = \mathbb{E}[X^4] - 3[\mathbb{E}[X^2]]^2. \tag{2.10}$$

Fehérített Z valószínűségi változóra ez a kifejezés az egységnyi varianciát ($\mathbb{E}[Z^2] = 1$) kihasználva tovább egyszerűsödik:

$$\operatorname{Kurt}(Z) = \mathbb{E}[Z^4] - 3. \tag{2.11}$$

A csúcsosság fontos tulajdonsága, hogy értéke normális eloszlású valószínűségi változók esetére 0, az ennél nagyobb értékek a normálisnál hegyesebb sűrűségfüggvényhez tartoznak ezek az ún. szuper-gaussi eloszlások, a negatív értékek pedig a normálisnál laposabb eloszlásfüggvényt jelentenek, ezek a szub-gaussi eloszlások. Az is megfigyelhető, hogy minél hegyesebb vagy laposabb egy sűrűségfüggvény, a hozzá tartozó valószínűségi változó csúcsosságának abszolút értéke annál nagyobb, ezért a csúcsosság alkalmas a nemnormalitás mérésére. A módszer fő előnye az egyszerű és hatékony számíthatósága, (gyakorlatban az eloszlás ismerete nélkül, Egyszerűen mért minták negyedik hatványának mintaátlagával becsülhető), azonban igen érzékeny a kiugró elemekre (outlier), így valós mérések esetén néhány hibás minta nagyban torzítani tudja a becsült csúcsosságot.

2.4.2. Negentrópia

A nemnormalitás egy másik lehetséges mértéke az információelméleti hátterű negentrópia használata. A negentrópia az ún. differenciális entrópiából származtatható, mely lényegében azt mutatja meg, hogy egy X valószínűségi változó mekkora bizonytalanságot hordoz. A differenciális entrópia formális definíciója a következő:

$$H(X) = \mathbb{E}[-\log(p_X(X))] = -\int p_X(\eta)\log p_X(\eta)\,\mathrm{d}\eta. \tag{2.12}$$

Belátható, hogy adott szórású eloszlások közül a normális eloszlás differenciális entrópiája a legnagyobb. Vezessük be a negentrópiát a kövekező összefüggés szerint [12]:

$$J(X) = H(X_{\text{gauss}}) - H(X), \tag{2.13}$$

ahol X_{gauss}, egy gaussi eloszlású valószínűségi változó, melynek szórása megegyezik X szórásával. Látható, hogy J értéke pontosan akkor 0, ha X normális eloszlású, ellenkező esetben pozitív, a negentrópia tehát alkalmas a nemnormalitás mérésére. A negentrópia legnagyobb előnye a csúcsossággal szemben nagyobb robusztussága a gyakorlatban elkerülhetetlen mérési hibákkal szemben, főbb hátránya viszont nehézkes becsülhetősége, hiszen kiszámításához becsülnünk kell a mért adatok sűrűségfüggvényét.

2.5. Az ICA korlátai

Bár az ICA viszonylagosan laza feltételeket támaszt a szeparálandó kevert jelekkel szemben, az $\mathbf{A}$ keverőmátrix és az $\underline{S}$ eredeti jelek együttes ismeretlenségéből következik néhány korlátja. Legyen adott egy $\mathbf{D}$ diagonális mátrix, melynek minden főátlóbeli elemére $D_{ii} \neq 0$ teljesül, valamint legyen adott egy $\mathbf{P}$ permutáló mátrix. Ekkor az (2.1) alapján igaz a következő összefüggés:

$$\mathbf{A}\underline{\tilde{S}} = \mathbf{APD}\underline{S} = \tilde{\mathbf{A}}\underline{S} \tag{2.14}$$

Vegyük észre, hogy ha $\underline{S}$ komponensei függetlenek voltak, akkor $\underline{\tilde{S}} = \mathbf{PD}\underline{S}$ komponensei is függetlenek lesznek, hiszen azokat mindössze skáláztuk illetve megcseréltük, tehát a függetlenség vizsgálatával nem eldönthető, hogy $\underline{S}$ volt az eredeti jelek vektora, a keverőmátrix pedig $\tilde{\mathbf{A}} = \mathbf{APD}$, vagy $\underline{\tilde{S}}$ volt az eredeti jel és $\mathbf{A}$ a keverőmátrix. Ez a következő korlátokat vonja maga után az ICA modell szerint kevert jelek szeparálhatóságával kapcsolatban:

1. Az eredeti jelek csak egy jelenként különböző, ismeretlen $d_i \in (\mathbb{R}\setminus\{0\})$ skálázófaktor erejéig nyerhetők ki.

2. A szeparált jelek alapján az eredeti jelek sorrendje nem határozható meg.

3. fejezet

A frekvenciatartományi ICA elmélete

3.1. A frekvenciatartományi ICA motivációja

A 2. fejezetben áttekintettük a klasszikus ICA módszerek működésének alapjait. Magától értetődőnek tűnik a kérdés: hogyan használható ez a módszer a dolgozat eredeti céljának, azaz a mikrofontömbös nyalábformálásnak a megvalósításához? Elsőre logikus lépésnek tűnhet egy klasszikus, valós ICA módszernek, például az igen közkedvelt, valamint a népszerűbb programozási nyelveken kész implementációkkal rendelkező FastICA [9] algoritmusnak a mikrofonok által rögzített időtartományi jeleken történő közvetlen alkalmazása. Ezt kipróbálva viszont azt tapasztalhatjuk, hogy az ICA algoritmusunk jóságától függetlenül, lényegében semmilyen értelmezhető szeparációs hatást nem sikerült elérnünk. Ennek oka az ICA modellfeltevésében keresendő: Az ICA feltételezi, hogy a jeleink instant módon lettek keverve, azaz a kevert jel minden diszkrét időpillanatban az eredeti jelek adott időpillanatban vett értékének invertálható lineáris kombinációja. Ezt a 2. fejezetben formálisan úgy fogalmaztuk meg, hogy a keverés művelete az $\underline{S}$ valószínűségi változó vektor egy $\mathbf{A}$ mátrixszal történő transzformációjának eredménye, a jelek adott időpillanatbeli értéke pedig nyilván a valószínűségi változó egy konkrét realizációja, melyből a föntebbi állítás következik. A valós hangterjedésből következő keverési modell viszont nem ilyen, hiszen a különböző térbeli pozíciókban elhelyezkedő mikrofonok szinte biztosan különböző időkésleltetéssel veszik az adott forrás jelét. Ezt, a valós akusztikus környezetekben létrejövő keverési modellt konvolutív modellnek nevezzük. Ezen fejezet hátralévő részében formálisan is bevezetjük a konvolutív keverési modellt, megvizsgáljuk az ebben a modellben történő forrásszeparálás általános lehetőségeit, majd bevezetjük a frekvenciatartományi ICA-t, vagyis az ICA módszerek adaptálásának módját konvolutív környezetekre.

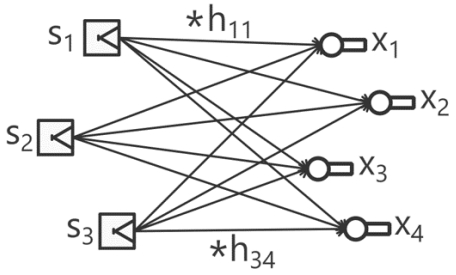
3.2. Konvolutív keverési modell

Valós akusztikus környezetekben egy adott forrás és érzékelési pont közti jelterjedést leginkább a pontok távolságából származó időkésleltetés, a többutas terjedés (reflexiók) illetve a frekvenciafüggő csillapítás határozzák meg. Ezt jelfeldolgozási gyakorlatban az adott környezet adott két pont között számított impulzusválaszával szokás megadni, amely lényegében azt mutatja meg, hogy az egyik pontban $n = 0$ időpillanatban kiadott impulzus hatására a másik pontban milyen időfüggő jelet érzékelünk. Az impulzusválasz ismeretében az X_1 mérési pontban az $[n]$ diszkrét időpillanatban az S_1 forrás jele a két függvény

diszkrét konvolúciójaként áll elő:

$$x_{1S_1}[n] = (h_{11} * s_1)[n] = \sum_{k=-\infty}^{\infty} h_{11}[n-k]s_1[k], \tag{3.1}$$

ahol x_{1S_1} az X_1 mérési pont S_1-ből származó értéke, h_{11} pedig az impulzusválasz a két pont között. A konvolutív keverési modellt a 3.1 ábra szemlélteti. Megjegyzendő, hogy a



3.1. ábra. A konvolutív keverési modell

valós akusztikus rendszerek kauzálisak ($n < 0$-ra $h[n] = 0$), illetve jó közelítéssel véges impulzusválaszúak ($\exists N \in \mathbb{N}^+$, hogy $\forall n > N$-re $h[n] = 0$), így a végtelen szumma ennek megfelelően egyszerűsödik. Ezek alapján felírható az egy mikrofon által érzékelt x_1 jel S darab forrás esetére a szuperpozíció elvét felhasználva:

$$x_1[n] = \sum_{s=1}^{S} x_{1S_s}[n] = \sum_{s=1}^{S} (h_{1s} * s_s)[n]. \tag{3.2}$$

Több mikrofon esetén a többi mikrofon által vett jelek hasonlóan írhatók fel. A [14]-ben bevezetett, FIR mátrix algebra jelölésrendszerrel felírva:

$$\begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_M[n] \end{bmatrix} = \begin{bmatrix} h_{11} & \cdots & h_{1S} \\ h_{11} & \cdots & h_{1S} \\ \vdots & \ddots & \vdots \\ h_{M1} & \cdots & h_{MS} \end{bmatrix} * \begin{bmatrix} s_1[n] \\ s_2[n] \\ \vdots \\ s_S[n] \end{bmatrix}, \tag{3.3}$$

$$\mathbf{x}[n] = \underline{\mathbf{H}} * \mathbf{s}[n], \tag{3.4}$$

ahol M a mikrofonok száma, h_{ms} pedig a rendszer impulzusválasza az s indexű forrás illetve az m indexű mérési pont között.

3.3. Konvolutív mixek szeparációs lehetőségei

Könnyen látható, hogy konvolutív modell szerint kevert jelek szeparációja jóval nagyobb kihívás mint a 2.1 fejezetben említett instant keverékek szeparációja, hiszen itt nem egy skalár értékekből álló demix mátrixot kell valamilyen módon meghatároznunk, hanem a különböző terjedési utakhoz tartozó konvolúcióknak kell valamilyen módon az inverzét meghatározni.

3.3.1. Időtartományi módszerek

A szeparációs probléma egyik lehetséges megközelítése az időtartománybéli megközelítés, azaz olyan $w_{ij}[n]$ impulzusválaszú szűrök konstruálása, melyek a következő összefüggés

szerint képesek a konvolutív keverés invertálására:

$$u_i[n] = \sum_{m=1}^{M} \sum_{k=1}^{N} w_{ij}[n] x_m[n-k]. \tag{3.5}$$

Ez a megoldás M darab FIR vagy IIR szűrő együtthatóinak optimalizációját követeli meg valamilyen függetlenségi kritérium szerint (például a már említett nemnormalitás), amely a gyakorlatban csak igen rövid impulzusválaszú rendszerekre valósítható meg hatékonyan [8], így az akusztikus forrásszeparációs gyakorlatban nem praktikus.

3.3.2. Frekvenciatartományi módszerek

A frekvenciatartományi konvolutív forrásszeparálás ötletét a Fourier-transzformáció konvolúciós tétele tétele adja, mely a következőt mondja ki:

$$\mathcal{F}\{f * g\}(\omega) = \mathcal{F}\{f\}(\omega)\mathcal{F}\{g\}(\omega). \tag{3.6}$$

Azaz ha adott két ($f(t)$ és $g(t)$) Fourier-transzformálható, folytonos idejű jel, akkor konvolúciójuk Fourier-transzformáltja a külön-külön számított Fourier-transzformáltjuk szorzataként áll elő, másképp fogalmazva a (folytonos) időtartományi konvolúció frekvenciatartományban szorzássá egyszerűsödik. Ebből, valamint a Fourier-transzformáció linearitásából az következik, hogy a konvolutív keverékeket megadó, időtartományban konvolúciók összegeként felírható (3.2) összefüggés (pontosabban annak folytonos megfelelője) frekvenciatartományban frekvenciafüggően ugyan, de skalárok szorzatának összegeként írható fel. Az időtartományban konvolutív keverési modell így frekvenciatartományban frekvenciafüggő instant keverési modellként írható fel, tehát elméletileg az instant keverékekre érvényes szeparációs módszerek használatával megoldhatóvá válik a konvolutív probléma.

Bár a frekvenciatartományi forrásszeparációs modell megvalósítása elsőre igen egyszerűnek tűnik, vegyük észre hogy annak gyakorlati megvalósítása több lényeges problémát is fölvet:

1. A konvolúciós tétel ebben a formában folytonos idejű jelekre írható fel, a digitális jelfeldolgozásban pedig nyilván diszkrét időben dolgozunk.

2. A gyakorlatban nyilván véges mintaszámú diszkrét jeleink vannak, ezekre standard diszkrét Fourier-transzformációt (DFT) végrehajtva, minden frekvenciabinre egyetlen minta jut, az instant szeparációs modellek pedig eloszlásokon dolgoznak, melyek becsléséhez minél több mintapontra (realizációra) van szükség.

3. Az időtartományban valós értékű jelek (mind folytonos, mind diszkrét esetben) frekvenciatartományban a legtöbb esetben komplex értékűek, a klasszikus instant szeparációs módszerek pedig valós értékekre működnek.

Az első probléma megoldását a diszkrét konvolúciós tétel adja, amely bár a folytonos idejű tétellel analóg módon diszkrét Fourier-transzformációra (DFT) és diszkrét konvolúcióra felírva nem igaz, levezethető, hogy cirkuláris konvolúcióra fennáll az összefüggés:

$$\mathrm{DFT}(h \circledast s)[k] = \mathrm{DFT}(h)[k]\,\mathrm{DFT}(s)[k], \tag{3.7}$$

ahol k a diszkrét frekvenciabin indexe $\circledast$ pedig a cirkuláris konvolúciót jelöli. Belátható emellett, hogy kellően nagy elemszámú DFT esetére a cirkuláris konvolúció igen jól közelíti a "normál" diszkrét konvolúciót, így a gyakorlati megvalósítás szempontjából a probléma megoldottnak tekinthető kellően nagy DFT ablak választásával.

3.4. frekvenciatartományi ICA

A frekvenciatartományi forrásszeparálás leggyakrabban használt konkrét megvalósítása a [27] által bevezetett frekvenciatartományi ICA (FD-ICA) módszer. Mivel a továbbiakban az FD-ICA-ra alapozunk a nyalábformáló algoritmus megalkotásánál, az FD-ICA pontos bevezetését nem kerülhetjük el:

Feltételezzük, hogy a források és mikrofonok M száma megegyezik. Valamint tartalmazzák a rögzített regisztrátumok a mikrofonok N diszkrét mintáit. A kevert jel ekkor egy $M \times N$-es $\mathbf{s}$ mátrixként felírható:

$$\mathbf{x} = \begin{bmatrix} \underline{x_1} \\ \underline{x_2} \\ \vdots \\ \underline{x_M} \end{bmatrix} = \begin{bmatrix} x_1[1] & \cdots & x_1[N] \\ x_2[1] & \cdots & x_2[N] \\ \vdots & \ddots & \vdots \\ x_M[1] & \cdots & x_M[N] \end{bmatrix}. \tag{3.8}$$

Alkalmazzunk F ablakhosszú Short Time Fourier Transform (STFT) transzformációt az $\mathbf{x}$ soraira. Az STFT-k kimenete egy-egy $F \times T$ méretű X_m mátrix, ahol biztosan $T > \frac{N}{F}$ (T pontos értéke az STFT átlapolási faktorától függ):

$$\mathrm{STFT}(\underline{x_m}) = \mathbf{X_m} = \begin{bmatrix} X_{11} & \cdots & X_{1T} \\ X_{21} & \cdots & X_{2T} \\ \vdots & \ddots & \vdots \\ X_{F1} & \cdots & X_{FT} \end{bmatrix} \tag{3.9}$$

Az STFT transzformáció lényegében a bemeneti vektor K hosszú, legtöbbször átfedő szegmenseinek DFT transzformációja, melyek kimenetei az X mátrix oszlopait alkotják. A mátrix sorai így az adott frekvenciabinben a különböző időszegmensekhez tartozó DFT együtthatókat tartalmazzák, Az STFT mátrix tehát az adott jel idő-frekvenciatartományi reprezentációja. Tekintsük az $\mathbf{X_m}$ mátrixok $0 < f < F$-edik soraiból mint sorvektorokból összefűzött $\mathbf{X_f}$ mátrixot:

$$\mathbf{X_f} = \begin{bmatrix} X_{1_{f1}} & \cdots & X_{1_{fT}} \\ X_{2_{f1}} & \cdots & X_{2_{fT}} \\ \vdots & \ddots & \vdots \\ X_{M_{f1}} & \cdots & X_{M_{fT}} \end{bmatrix} \tag{3.10}$$

Ahol $X_{m_{ft}}$ az $\mathbf{X_m}$ mátrix $[f, t]$ indexű eleme.

Definiáljuk $\mathbf{x}$-szel analóg módon az eredeti (látens) források mintáiból összeállított $\mathbf{s}$ mátrixot, valamint ebből a fentebbiekhez hasonlóan vezessük be az $\mathbf{S_f}$ mátrixot. Vegyük észre, hogy ha az $\mathbf{x}$ keverék az $\mathbf{s}$ források konvolutív modell szerinti keveréséből áll, illetve feltételezzük, hogy a keverőrendszer időinvariáns, a következő összefüggést írhatjuk fel:

$$\mathbf{X_f} = \begin{bmatrix} X_{1_{f1}} & \cdots & X_{1_{fT}} \\ X_{2_{f1}} & \cdots & X_{2_{fT}} \\ \vdots & \ddots & \vdots \\ X_{M_{f1}} & \cdots & X_{M_{fT}} \end{bmatrix} = \begin{bmatrix} A_{f_{11}} & \cdots & A_{f_{1M}} \\ A_{f_{21}} & \cdots & A_{f_{2M}} \\ \vdots & \ddots & \vdots \\ A_{f_{M1}} & \cdots & A_{f_{MM}} \end{bmatrix} \begin{bmatrix} S_{1_{f1}} & \cdots & S_{1_{fT}} \\ S_{2_{f1}} & \cdots & S_{2_{fT}} \\ \vdots & \ddots & \vdots \\ S_{M_{f1}} & \cdots & S_{M_{fT}} \end{bmatrix} = \mathbf{A_f S_f}, \tag{3.11}$$

ahol $\mathbf{X_f}$ oszlopai a kevert jelek adott indexű mintái, melyek az $\mathbf{S_f}$ ugyanazon indexű oszlopai által reprezentált látens mintákhoz tartoznak, $\mathbf{A_f}$ pedig az adott frekvenciabinhez tartozó $M \times S$ méretű keverőmátrix (jelen esetben feltételezzük, hogy források és mikrofonok száma megegyezik ($S = M$) tehát $\mathbf{A_f}$ $M \times M$-es).

Ez az összefüggés azért teljesül, mivel kifejtve $\mathbf{X_f}$ mátrix egy $X_{f_{mt}}$ elemét, illetve a diszkrét konvolúciós tételt felhasználva:

$$X_{f_{mt}} = \sum_{s=1}^{M} A_{f_{ms}} S_{m_{ft}} \Longleftrightarrow \underline{x_{mt}} = \sum_{s=1}^{M} \underline{h_{ms}} \circledast_F \underline{s_{s_t}}, \tag{3.12}$$

ahol $\underline{s_{s_t}}$ az s indexű látens forrás mintáinak a t indexű STFT ablakba eső szakasza, F az STFT ablakhossza.

Ez pedig egy instant modell szerinti keverésként értelmezhető, valamint az is könnyen látható, hogy a 3.3.2 alfejezetben említett problémák közül a másodikat az STFT használatával lényegében meg is oldottuk, hiszen így minden frekvenciabinhez legalább T megfigyelés tartozik, így megfelelően hosszú regisztrátum és megfelelő STFT ablakhossz esetén az X_{m_f}-ek eloszlásának paraméterei jól becsülhetők, azaz a szeparációjukhoz használható valamilyen ICA módszer. Vegyük azonban észre, hogy $\mathbf{X_f}$ elemei itt is komplex értékűek lehetnek, hiszen DFT együtthatókról beszélünk. Tegyük egyelőre fel, hogy az ICA probléma megoldására komplex bemenetekre is van módszerünk, ekkor az FD-ICA alapú szeparáció viszonylag egyszerűen végrehajtható:

1. Hajtsuk végre a komplex ICA algoritmust az $\mathbf{X_f}$ bemenetekre $\forall f \in \mathbb{N} \cap [1, M]$-re, a kapott demix mátrixokat $\forall f$-re jelölje $\mathbf{W_f}$.

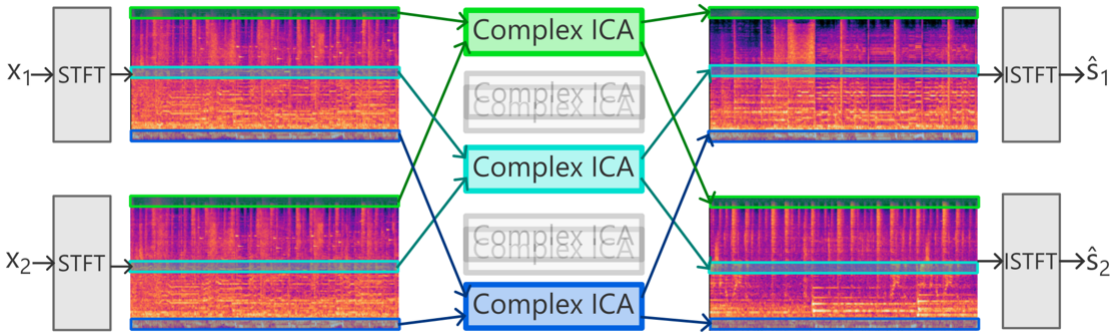
2. Számítsuk ki a $\hat{\mathbf{S}}_\mathbf{f}$ mátrixokat $\forall f$-re a következő összefüggés szerint:

$$\hat{\mathbf{S}}_\mathbf{f} = \mathbf{W_f X_f} \tag{3.13}$$

3. Az $\hat{\mathbf{S}}_\mathbf{f}$ mátrixokból állítsuk össze a $\hat{\mathbf{S}}_\mathbf{m}$ mátrixokat úgy, hogy $\forall f$ és $\forall m$-re az $\hat{\mathbf{S}}_\mathbf{m}$ $F \times T$-es mátrix f indexű sora legyen az $\hat{\mathbf{S}}_\mathbf{f}$ mátrix m indexű sora. Ezzel meghatároztuk a becsült $\underline{\hat{s}_m}$ források STFT transzformáltját.

4. Számítsuk ki $\forall m$-re az $\hat{\mathbf{S}}_\mathbf{m}$ időtartományi alakját ($\underline{\hat{s}_m}$) inverz STFT transzformáció (ISTFT) használatával.

A frekvenciatartományi ICA működését a 3.2 ábra szemlélteti.



3.2. ábra. A frekvenciatartományi ICA sematikus ábrája

4. fejezet

Komplex ICA

4.1. Komplex ICA megvalósítások áttekintése

Ahogy azt az FD-ICA bevezetésénél is láthattuk, a frekvenciatartományi szeparációs módszerek kulcsfontosságú kérdése a komplex jeleken is használható ICA algoritmus megválasztása. Mivel a komplex ICA jelentős gyakorlati alkalmazásokkal bír többek között a távközlés, radar jelfeldolgozás, arcfelismerés etc. területén, a probléma igen jelentős szakirodalmi háttérrel rendelkezik. A korai komplex ICA módszerek, mint a Fourth-Order Blind Identification (FOBI) [2] algoritmus, tipikusan negyedik momentumot (csúcsosságot) használtak függetlenségi mértékként. Bár ezek bizonyos esetekben igen jól működtek, a klasszikus ICA feltételeknél tipikusan szigorúbb feltételeket szabtak a szeparálandó forrásokra. A kumuláns-mátrix alapú Joint Approximate Diagonalization of Eigenmatrices (JADE ICA) [3] algoritmus már általános esetben is igen jól teljesített, azonban a szeparációs teljesítménye a források számának növelésével jelentősen csökken [16], így nem optimális a potenciálisan igen sok forrással dolgozó nyalábformálási feladatunkra. A valós ICA gyakorlatban igen népszerű, negentrópia alapú FastICA [9] komplex kiterjesztése, a Complex FastICA [1] igen jó konvergenciatulajdonságokkal bír, illetve népszerűsége révén hatékony kész implementációkkal rendelkezik a népszerűbb programozási nyelvekben, így gyakorlati megvalósíthatóság szempontjából optimális lenne. Hátránya viszont, hogy csak cirkuláris forrásokra működik hatékonyan, ez a tulajdonság pedig az FD-ICA esetében nem garantálható. A komplex ICA algoritmusok tanulmányozása során két modern módszerre esett a választás, melyeket ezen dolgozat során részletesebben vizsgálunk, illetve amelyeket a gyakorlatban ki is próbálunk a mikrofontömbös FD-ICA alapú beamforming esetére:

1. A Complex Maximization of Nongaussianity (CMN) algoritmuscsalád [22] a Complex FastICA továbbfejlesztésének tekinthető, ahhoz hasonlóan a CMN is negentrópia becslés alapú költségfüggvényt használ, myelynek optimalizálója fixpontos algoritmussal történik, viszont a Complex FastICA-val ellentétben mind cirkuláris, mind nemcirkuláris forrásokra hatékonyan működik. A CMN algoritmusok a negentrópiát becslő költségfüggvényben különböznek, ezen dolgozat során a legfejlettebbnek tekinthető Adaptive CMN (ACMN) [22] változatot használjuk amely adaptív, Exponential Power Family alapú becslést használ a negentrópia becslésére. Ez a módszer azért került kiválasztásra, mivel kiemelkedő szeparációs teljesítményt ígér, viszonylagosan magas számítási komplexitás árán

2. A Robust ICA [29] egy kurtózis alapú komplex ICA módszer, mely Exact Line Search algoritmust használ algebrai lépésköz számítással a költségfüggvény optimalizálása. A módszer igen jó konvergenciasebességet ígér, valamint nemcirkuláris forrásokra

is működőképes. Használatát az ACMN-nél kisebb számításigénye és jó szeparációs teljesítménye motiválta.

4.2. Komplex valószínűségi alapok

A komplex ICA módszerek tárgyalása a komplex valószínűségi változók alapvető ismertét igényli. Tekintsük ezeket át: Egy Z komplex valószínűségi változó definíciója a következő:

$$Z = Z^R + jZ^i, \tag{4.1}$$

ahol Z^R és Z^i valós értékű valószínűségi változók. A Z komplex valószínűségi változó sűrűségfüggvénye definíció szerint a valós illetve képzetes komponenseinek együttes sűrűségfüggvénye, feltéve, hogy ezek léteznek:

$$p_Z(z) = p_{\mathfrak{Re}(Z),\mathfrak{Im}(Z)}(\mathfrak{Re}(z), \mathfrak{Im}(z)). \tag{4.2}$$

A Z komplex valószínűségi változó várható értéke a következő:

$$\mathbb{E}[Z] = \mathbb{E}[\mathfrak{Re}(Z)] + j\mathbb{E}[\mathfrak{Im}(Z)], \tag{4.3}$$

feltéve, hogy mind a valós, mind a képzetes rész várható értéke létezik. Ha Z sűrűségfüggvénye létezik, a várható érték a komplex doménben is számolható:

$$\mathbb{E}[Z] = \iint_{\mathbb{C}} z f_Z(z)\,\mathrm{d}z. \tag{4.4}$$

Legyen $\underline{X}$ egy komplex valószínűségi vektorváltozó. $\underline{X}$ kovarianciamátrixa a következő:

$$\mathrm{cov}(\underline{X}) = \mathbb{E}[\underline{X} - \mathbb{E}[\underline{X}]]\mathbb{E}[\underline{X} - \mathbb{E}[\underline{X}]]^{\mathrm{H}}, \tag{4.5}$$

ahol $[]^{\mathrm{H}}$ az Hermite transzponált operátort jelöli. Az $\underline{X}$ pszeudo-kovarianciamátrixa pedig a következő:

$$\mathrm{pcov}(\underline{X}) = \mathbb{E}[\underline{X} - \mathbb{E}[\underline{X}]]\mathbb{E}[\underline{X} - \mathbb{E}[\underline{X}]]^{\mathrm{T}}. \tag{4.6}$$

A komplex valószínűségi változók fontos tulajdonsága a cirkularitás [24]. Akkor nevezünk egy Z komplex valószínűségi változót cirkulárisnak, ha $\forall\alpha \in \mathbb{R}$-re teljesül rá a következő összefüggés:

$$p_Z(z) = p_Z(e^{j\alpha}z), \tag{4.7}$$

azaz $p_Z(z)$ értéke csak $|z|$-től függ (lényegében sűrűségfüggvény az origóra forgásszimmetrikus).

4.3. CMN ICA

Az adaptív CMN a 2. fejezetben bevezetett instant ICA modellen alapul azzal a feltételezéssel, hogy itt az eredeti jelek, a kevert jelek, illetve a keverőmátrix komplex értékűek, valamint feltételezi, hogy a kevert jelek már fehérítettek. Vegyük észre, hogy ez nem jelent külön megkötést, hiszen a lineárisan kevert jelek komplex esetben is fehéríthetők a valós PCA-val teljesen analóg módon számítható komplex PCA transzformáció használatával. A 2.2. alfejezetben leírt módszer a kovarianciamátrix számítását a 4.2 alfejezetben ismertetett módon elvégezve lényegi módosítás nélkül használható. A keverési modell tehát az alábbi:

$$\underline{Z} = \mathbf{V}\underline{X} = \mathbf{VA}\underline{S} = \hat{\mathbf{A}}\underline{S} \tag{4.8}$$

ahol $\underline{S}$ az eredeti, $\underline{X}$ a kevert jelek vektora, $\mathbf{A}$ a (komplex) keverőmátrix, $\mathbf{V}$ a fehérítő mátrix, $\underline{Z}$ pedig a fehérített kevert jelek vektora ($\hat{\mathbf{A}}$ pedig a fehérítés utáni látszólagos unitér keverőmátrix). A feladat $\mathbf{W}^{\mathrm{H}} \approx \hat{\mathbf{A}}^{-1}$ unitér mátrix megtalálása, hiszen erre:

$$\mathbf{W}^{\mathrm{H}}\underline{Z} = \mathbf{W}^{\mathrm{H}}\hat{\mathbf{A}}\underline{S} = (\mathbf{W}^{\mathrm{H}}\mathbf{V})\mathbf{A}\underline{S} = \underline{\hat{S}}. \tag{4.9}$$

(Ekkor $\mathbb{E}\left[\underline{Z}\underline{Z}^{\mathrm{H}}\right] = \mathbb{I}$)

4.3.1. CMN kontrasztfüggvény

A CMN algoritmus a negentrópiamaximalizálás alapú ICA algoritmusok közé tartozik. A negentrópiamaximálizálás módszer komplex valószínűségi változókra történő egzakt levezetését megtalálhatjuk a [21] cikkben, a CMN módszer megértéséhez elegendő a J_{neg} komplex negentrópiát megadó összefüggés ismerete:

$$J_{\mathrm{neg}}(Z) = H(Z_{\mathrm{gauss}}) - H(Z), \tag{4.10}$$

ahol $H(Z_{\mathrm{gauss}})$ a Z-vel megegyező szórású komplex gaussi eloszlású valószínűségi változó, $H()$ pedig a komplex differenciális entrópia:

$$H(Z) = H\left(\mathfrak{Re}(Z), \mathfrak{Im}(Z)\right) = -\int_{-\infty}^{\infty} p_Z(z)\log(p_Z(z))\,\mathrm{d}z. \tag{4.11}$$

A CMN kontrasztfüggvény a következő:

$$J(\underline{w}) = \mathbb{E}\left[|G(\underline{w}^{\mathrm{H}}\underline{Z})|^2\right], \tag{4.12}$$

Ahol $G : \mathbb{C} \mapsto \mathbb{C}$ a (4.10) formulában definiált J_{neg} negentrópiát becslő függvény.

Megjegyzendő, hogy a gyakorlatban $\underline{Z}$ eloszlása nem ismert, mindössze z realizációi ismertek, így az optimális, negentrópiát megadó G használata nem lehetséges. Azt is érdemes megemlíteni, hogy a CMN algoritmusok legfőbb különbsége a Complex FastICA-hoz képest az, hogy ott a negentrópia becslése $|\underline{w}^{\mathrm{H}}\underline{Z}|^2$-en van értelmezve, azaz a becsült komplex jelnek csak az abszolút értékétől függ, ebből következően a módszer teljesítménye noncirkuláris források esetében igen gyenge.

Mivel Z fehérített, a $\mathbf{W}$ demix mátrix biztosan unitér [5], tehát A $\mathbf{W}$ mátrix egy $\underline{w}$ oszlopára $\|\underline{w}\|^2 = 1$ biztosan teljesül. A negentrópiát maximalizáló, egy forrást szeparáló $\underline{w_{\mathrm{opt}}}$-ra felírható tehát:

$$\underline{w_{\mathrm{opt}}} = \underset{\|\underline{w}\|^2=1}{\operatorname{argmax}}\, \mathbb{E}\left[|G(\underline{w}^{\mathrm{H}}\underline{Z})|^2\right]. \tag{4.13}$$

4.3.2. A fixpontos alogoritmus

A (4.13) egyenlet tehát egy feltételes szélsőértékkeresési feladatot fogalmaz meg. A megoldáshoz Lagrange-módszert használunk:

$$\mathcal{L}(\underline{w}, \lambda) = J(\underline{w}) + \lambda(\underline{w}^*\underline{w} - 1), \tag{4.14}$$

ahol $\mathcal{L}$ a Lagrange függvény, λ az ún. Lagrange-multiplikátor, $\underline{w}^*$ pedig $\underline{w}$ konjugáltja. Az optimum szükséges feltétele, hogy $\mathcal{L}$ parciális deriváltjai zérusok legyenek. Az optimum keresését Newton módszerrel végezzük, a [28] alapján egy skalárértékű komplex f

vektorfüggvényre felírt Newton iteráció a következő:

$$\Delta\underline{\tilde{w}} = -\left(\frac{\partial^2 f}{\partial\underline{\tilde{w}}^*\partial\underline{\tilde{w}}^{\mathrm{T}}}\right)^{-1}\frac{\partial^2 f}{\partial\underline{\tilde{w}}^*} = -\mathbf{H}_f^{-1}\hat{\nabla}_f^*, \tag{4.15}$$

ahol $\Delta\underline{\tilde{\hat{w}}} = \underline{\hat{w}}^{n+1} - \underline{\hat{w}}^n$, azaz approximáció módosító összefüggése, $\mathbf{H}$ a komplex Hesse mátrix, $\hat{\nabla}^*$ a konjugált gradiens operátor, a hullámmal jelölt vektorok a következő formában állnak elő:

$$\underline{\tilde{z}} = \left[z_1, z_1^*, z_2, z_2^*, \ldots z_N, z_N^*\right]^{\mathrm{T}}, \tag{4.16}$$

ahol $\underline{z}$ egy $\mathbb{C}^N$-beli vektor. A (4.15) Newton iterációt a (4.14) függvényre alkalmazva:

$$\Delta\underline{\tilde{w}} = -\left(\mathbf{H}_J + \lambda\mathbf{I}\right)^{-1}\left(\hat{\nabla}_J^* + \lambda\underline{\hat{w}}^n\right) \tag{4.17}$$

Átrendezve:

$$\mathbf{K}\underline{\hat{w}}^{n+1} = \left(\mathbf{H}_J + \lambda\mathbf{I}\right)\underline{\hat{w}}^{n+1} = -\hat{\nabla}_J^* + \mathbf{H}_J\underline{\hat{w}}^n \tag{4.18}$$

A (4.18) egyenletet a (4.12) kontrasztfüggvényre kifejtve az $\mathbb{E}\left[zzf(z)\right] \approx \mathbb{E}\left[zz\right]\mathbb{E}\left[f(z)\right]$ közelítést használva a következő összefüggést kapjuk [22]:

$$\mathbf{K}\underline{w}^{n+1} = -\mathbb{E}\left[G^*(Y)g(Y)\underline{Z}\right] + \mathbb{E}\left[g(Y)g^*(Y)\right]\underline{w}^n + \mathbb{E}\left[\underline{Z}\underline{Z}^{\mathrm{T}}\right]\mathbb{E}\left[G^*(Y)g'(Y)\right](\underline{w}^n)^* \tag{4.19}$$

ahol $Y = \underline{w}^{\mathrm{H}}\underline{Z}$, valamint $\mathbf{K} = (\mathbf{H}_J + \lambda\mathbf{I})$, G első deriváltja $G' = g$, második deriváltja pedig $G'' = g'$.

Megmutatható, hogy $\underline{w}$ optimumánál $\mathbf{K} = \alpha\mathbf{I}$ alakban áll elő (α valós skalár) [22], ezért a $\|\underline{w}\|^2 = 1$ kényszer miatt elhagyható (hiszen $\underline{w}$-t minden iterációban normalizáljuk). Így az iteráció a következő, végleges alakban számítható:

$$\underline{w}^{n+1} = -\mathbb{E}\left[G^*(Y)g(Y)\underline{Z}\right] + \mathbb{E}\left[g(Y)g^*(Y)\right]\underline{w}^n + \mathbb{E}\left[\underline{Z}\underline{Z}^{\mathrm{T}}\right]\mathbb{E}\left[G^*(Y)g'(Y)\right](\underline{w}^n)^* \tag{4.20}$$

A gyakorlati megvalósítás szempontjából lényeges, hogy az Y valószínűségi változót közvetlenül nem ismerjük, mindössze annak realizációit tudjuk kiszámítani az $\underline{y} = \underline{w}^{\mathrm{H}}\mathbf{z}$ összefüggés szerint, ahol $\mathbf{z} \in \mathbb{C}^{M\times N}$ a bemeneti mintákból mint oszlopvektorokból összefűzött mátrix. A (4.20) összefüggés ekkor a várható értékeket mintaátlagokkal becsülve közelíthető:

$$\underline{w}^{n+1} = \frac{1}{N}\sum_{i=1}^{N} G^*(y_i)g(y_i)\underline{z_i} + \frac{1}{N}\sum_{i=1}^{N} g(y_i)g^*(y_i)\underline{w}^n + \frac{1}{N}\sum_{i=1}^{N}\underline{z_i}\underline{z_i}^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^{N} G^*(y_i)g'(y_i)\,(\underline{w}^n)^* \tag{4.21}$$

ahol $\underline{z_i}$ a $\mathbf{z}$ mátrix i-edik oszlopvektora, és $y_i = \underline{w}^H\underline{z_i}$.

4.3.3. Adaptable CMN

A CMN algoritmusok közti lényegi különbség a negentrópiát mérő G kontrasztfüggvény megválasztása. Megmutatható, hogy a negentrópiát megadó, G_{opt} nemlinearitás a becsült S forrás sűrűségfüggvényének logaritmusa:

$$G_{\mathrm{opt}}(Y) = -\log p_S(Y) \tag{4.22}$$

Mivel ezt nem ismerjük, illetve becsülni is igen költséges, a CMN algoritmusok tipikusan valamilyen heurisztikusan meghatározott fix nemlinearitást használnak kontrasztfüggvényként [21]. Az Adaptive CMN annyiban fejlettebb, hogy itt egy parametrikus kontrasztfüggvényt használunk, amely egy Exponential Power Family (EPF) eloszlás negentrópiáját

becsüli. Az EPF eloszlás alakparaméterét iteratívan az aktuális becsült forrás varianciája alapján állítjuk be, ez alapján pedig a kontrasztfüggvény paraméterezését úgy állítjuk be hogy az EPF eloszlás aktuális negentrópiáját adja, ezzel minél jobban becsülje a forráselsozlás tényleges negentrópiáját. Az általános EPF eloszlást a következő összefüggés adja meg [17]:

$$p_{EPF}(y) = \frac{1}{2\sigma_p p^{\frac{1}{p}} \Gamma(1+\frac{1}{p})} \exp\left(\frac{-|y|^p}{p\sigma_p^p}\right), \tag{4.23}$$

ahol Γ a Gamma függvény, $\sigma_p > 0$ a variancia paraméter, $p > 0$ pedig az alakparaméter, amely 2-nél nagyobb értékekre szupergaussi, 2 értékre gaussi, 2-nél kisebb pozitív értékekre pedig szubguassi eloszlást eredményez. Levezethető, hogy egy EPF eloszlású Z valószínűségi változó negentrópiája a következő:

$$J(Z) = \frac{|Z|^{2K}}{2K} = \frac{Z^K}{\sqrt{2K}} \left(\frac{Z^K}{\sqrt{2K}}\right)^* \tag{4.24}$$

Ahol $K = p/2$ és p az EPF eloszlás alakparamétere. A (4.12) egyenlet alapján ekkor a G nemlinearitás a következő alakban írható fel:

$$G(z) = \frac{z^K}{\sqrt{2K}} \tag{4.25}$$

A feladat tehát az Y becsült forrásra illesztett EPF eloszlás p paraméterének becslése, hiszen ebből a G nemlinearitás számítható. Ehhez állítsuk az EPF szórásnégyzetét az aktuálisan becsül Y forrás szórásnégyzetének értékére (felhasználva hogy a becsült források biztosan 0 várható értékűek):

$$\mathbb{E}\left[Y^2\right] = \int_0^\infty \frac{Y^2}{\sigma_p p^{\frac{1}{p}} \Gamma\left(1+\frac{1}{p}\right)} \exp\left(\frac{-|Y|^p}{p\sigma_p^p}\right) = \frac{\Gamma\left(\frac{3}{p}\right) \sigma_p^2 p^{\frac{2}{p}-1}}{\Gamma\left(1+\frac{1}{p}\right)} \tag{4.26}$$

Vegyük észre hogy a (4.26) egyenletben σ_p értékét nem ismerjük, ennek Maximum likelihood becslése Y mintái alapján (mivel Y eloszlását nem ismerjük, csak annak mintáit tudjuk becsülni):

$$\hat{\sigma}_p = \left(\frac{\sum_{i=1}^N |y_i|^p}{N}\right)^{\frac{1}{p}} \tag{4.27}$$

Ahol N a becsült Y realizációinak száma (ez, megegyezik a bemeneti $\mathbf{z}$ mátrix oszlopainak számával azaz a kevert jel mintáinak számával), y_i-k pedig maguk a realizációk. Ekkor 4.26 egyenletbe 4.27 egyenletet helyettesítve, kihasználva, hogy a normalizálás miatt $\mathbb{E}\left[Y^2\right] = 1$, majd az egyenlet logaritmusának gyökeit Newton módszerrel keresve a következő becslést kapjuk p-re [22]:

$$p^{n+1} = p^n - \left(\frac{\partial \log\left(\mathbb{E}\left[Y^2\right]\right)}{\partial p}\right)^{-1} \log\left(\mathbb{E}\left[Y^2\right]\right) \tag{4.28}$$

Innen:

$$K^{n+1} = \frac{p^{n+1}}{2} \tag{4.29}$$

4.3.4. Több forrás szeparációja

Levezettük tehát az ACMN ICA iterációs összefüggését, valamit megadtuk az adaptív kontrasztfüggvény kiszámításának módját, ezek alapján a egy forrás szeparációja megvalósítható az ACMN algoritmussal. A gyakorlatban viszont nyilván az összes forrást ki akarjuk nyerni a keverékből. Mivel a bemeneti $\underline{Z}$ keverék fehérített, a különböző forrásokat kiválasztó $\underline{w}$ vektorok biztosan ortogonálisak egymásra ($\mathbf{W}$ demix mátrix unitér), tehát ha biztosítjuk hogy az aktuális forrást kiválasztó $\underline{w}$ vektor ortogonális legyen az összes eddigi forrást kiválasztó $\underline{w}$-re, biztosítjuk hogy új forrást szeparálunk a keverékből. Ezt pl. Gram-Schmidt módszerrel tudjuk egyszerűen megtenni. Legyenek $\{\underline{w_1}, \underline{w_2}, \dots \underline{w_m}\}$ a már szeparált forrásokat kiválasztó vektorok, $\underline{w^n}$ pedig az aktuálisan optimalizált szeparációs vektor. Az ortogonalitás biztosítható, ha minden iteráció végén kivonjuk $\underline{w^n}$-ből az eddigi szeparációs vektorok merőleges vetületét:

$$\underline{w_{ort}}^n = \underline{w^n} - \sum_{i=1}^{m} \underline{w_i} \underline{w_i}^{\mathrm{H}} \underline{w^n} \tag{4.30}$$

Majd normalizálva $\underline{w_{ort}}^n$-t biztosítjuk a $\|\underline{w}\|^2 = 1$ feltételt:

$$\underline{w_{ort}}^n = \frac{\underline{w_{ort}}^n}{\|\underline{w_{ort}}^n\|_2} \tag{4.31}$$

4.3.5. Az ACMN algoritmus implementációja

Az ACMN algoritmust tehát a következő módon írhatjuk fel: Legyen adott egy $\mathbf{z} \in \mathbb{C}^{M \times N}$ fehérített ($\frac{\mathbf{z}\mathbf{z}^{\mathrm{H}}}{N} = \mathbf{I}$) bemeneti mátrix mely a kevert jeleket reprezentálja. A mátrix sorai a kevert jelek adott időpillanatbeli értékét reprezentálják (Ezek lényegében a Z valószínűségi vektorváltozó realizációi). Az algoritmus pontos leírását az 1. algoritmus tartalmazza, ahol a dupla szögletes zárójellel jelölt azonosítók az azonos számú egyenletek jobb oldalát jelölik.

4.4. RobustICA

A RobustICA, a CMN ICA-hoz hasonlóan a 4.3 alfejezetben kifejtett komplex instant keverési modellre alapul. A robustICA a 2.4.1 alfejezetben bemutatott csúcsosság alapú ICA módszerek közé tartozik, valamint a legtöbb komplex ICA algoritmussal ellentétben képes nem előfehérített jeleken is dolgozni. A RobustICA kontrasztfüggvénye az általánosított kurtózis:

$$\mathcal{K}(\underline{w}) = \frac{\mathbb{E}\left[|Y|^4\right] - 2\mathbb{E}^2\left[|Y|^2\right] - \left|\mathbb{E}\left[Y^2\right]\right|^2}{\mathbb{E}^2\left[|Y|^2\right]}, \tag{4.32}$$

ahol:

$$Y = \underline{w}^H \underline{X}. \tag{4.33}$$

Megjegyzendő, hogy itt nem $\underline{Z}$ jelenik meg, hiszen a RobustICA nem feltételezi a fehérített jeleket, azonban mi fehérített jelekre fogjuk használni, mivel megmutatható, hogy fehérített jelekre jobb konvergenciatulajdonságokkal bír [18].

1. Algorithm ACMN ICA

Input: z

$m \leftarrow 1$
$\mathbf{W} \leftarrow Zeros(M, N)$
while $m \leq M$ **do**
$p \leftarrow p_{init}$
$\underline{w} \leftarrow NormalRand(\mu = 0, \Sigma = \mathbf{I}, N = N)$
do
$\underline{y} \leftarrow \underline{w}^{\mathrm{H}}\mathbf{z}$
$K \leftarrow \frac{p}{2}$
$\underline{w}^{n+1} \leftarrow$ [[4.21]]
$\underline{w} \leftarrow \frac{\underline{w}}{\|\underline{w}\|}$
$\hat{\sigma} \leftarrow$ [[4.27]]
$\mathbb{E}\left[Y^2\right] \leftarrow$ [[4.26]]
$p \leftarrow$ [[4.28]]
$\underline{w} \leftarrow \underline{w} - \mathbf{W}\mathbf{W}^{\mathrm{H}}\underline{w}$
$\underline{w} \leftarrow \frac{\underline{w}}{\|\underline{w}\|}$
$error \leftarrow \| |w^{n+1}| - |w^n| \|$
while $error > tolerance$
$\underline{\mathbf{W}_m^{\mathrm{T}}} \leftarrow \underline{w}$
$m \leftarrow m + 1$
end while

A RobustICA Exact Line Search módszert használ a kontrasztfüggvény optimalizálására. Ehhez először meghatározzuk a gradienst:

$$\nabla\mathcal{K}(\underline{w}) = \frac{4}{\mathbb{E}^2\left[|Y|^2\right]}\left[\mathbb{E}\left[|Y|^2 Y^* \underline{X}\right] - \mathbb{E}\left[Y\underline{X}\right]\mathbb{E}\left[Y^{*^2}\right] - \frac{\left(\mathbb{E}\left[|Y|^4\right] - \left|\mathbb{E}\left[Y^2\right]\right|^2\right)\mathbb{E}\left[Y^*\underline{X}\right]}{\mathbb{E}\left[|Y|^2\right]}\right], \tag{4.34}$$

majd a gradiens mentén optimalizáljuk a csúcsosságot:

$$\mu_{opt} = \underset{\mu}{\operatorname{argmax}} \left|\mathcal{K}(\underline{w} + \mu\nabla\mathcal{K}(\underline{w}))\right|. \tag{4.35}$$

Bár ennek a kifejezésnek az optimalizálása klasszikus numerikus optimalizációs módszerekkel igen költséges, megmutatható, hogy ebben az esetben az optimális μ_{opt} meghatározása visszavezethető egy negyedfokú polinom gyökeinek kiszámítására, ezt pedig Ferrari-formulát használva analitikusan is megoldható. A RobustICA algoritmus pontos levezetését illetve implementációjának lépéseit megtalálhatjuk [29]-ben, az algoritmus implementálása ez alapján egyszerűen megvalósítható.

5. fejezet

A naiv FD-ICA problémái és kiküszöbölésük

5.1. A naiv FD-ICA problémái

A 3. fejezetben láthattuk, hogy hogyan valósítható meg a konvolutív keverékek szeparációja az FD-ICA módszer szerint, valamint a 4. fejezetben áttekintettük, hogy hogyan valósítható meg az ehhez szükséges komplex értékű ICA. Ezek alapján egyszerűen implementálhatunk egy naiv FD-ICA algoritmust, azonban ezt akár szimulált, akár valós akusztikus környezetben mért kevert jeleken próbáljuk futtatni, azt tapasztalhatjuk hogy az algoritmus szeparációs teljesítménye igen gyenge. Ennek oka pedig az, hogy a 2.5 alfejezetben áttekintett limitációk természetesen a komplex ICA-ra is vonatkoznak, hiszen a keverési modell formálisan megegyezik. (a különbség nyilván annyi, hogy a komplex esetben a skálázófaktor is komplex értékű) Ezek a limitációk az ICA modell legtöbb alkalmazásánál nem jelentenek problémát, az FD-ICA keretben viszont igen hiszen itt a frekvenciabinenkénti ICA által behozott ismeretlen skálázás, a binek közt fáziseltolást, és a jel energiaspektrumának random torzulását okozza, ez pedig nyilván azt jelenti hogy a szeparált jelek jelentősen különbözhetnek az eredeti független jelektől. A permutációs bizonytalanság hatása ennél is jelentőseb, hiszen hiába szeparáljuk szét frekvenciabinenként a kevert jeleket, az ismeretlen permutáció hatására, a szeparált STFT spektrumok frekvenciabinenként véletlenszerűen tartalmazhatják a különböző források jeleit, így az ISTF transzformációt végrehajtva lényegében ismét összekevertük a jeleket. A skálázási valamint a permutációs bizonytalanság megoldása tehát kulcsfontosságú a hatékony FD-ICA alapú szeparációhoz.

5.2. A skálázási bizonytalanság megoldása

A skálázási problémára többféle megoldási ötletet olvashatunk a szakirodalomban. A korábbi megoldások többek között a demix mátrix normalizálásával [27], illetve annak diagonális elemeinek 1 értékűre korlátozásával [23] próbálták megoldani a skálázási bizonytalanságot. Ezen dolgozatban a leginkább elfogadott, megfigyelési térbe vetítés [19] módszerét használjuk $M \times M$-es esetre általánosítva: Vegyük a klasszikus instant ICA modellt:

$$\underline{X} = \mathbf{A}\underline{S}, \tag{5.1}$$

valamint egyelőre feltételezzük, hogy a permutációs probléma már megoldott és a becsült $\mathbf{W}$ demix mátrix a következő:

$$\mathbf{W} = \left(\hat{\mathbf{A}}\mathbf{\Lambda}\right)^{-1} = \mathbf{\Lambda}^{-1}\hat{\mathbf{A}}^{-1}, \tag{5.2}$$

ahol $\hat{\mathbf{A}}$ a keverőmátrix közelítése, $\mathbf{\Lambda} = \mathrm{Diag}(\Lambda_1, \Lambda_2 \dots \Lambda_M)$ pedig diagonális, ismeretlen skálázó mátrix. Ekkor a becsült források:

$$\underline{\hat{S}} = \mathbf{W}\underline{X} = \mathbf{\Lambda}^{-1}\hat{\mathbf{A}}^{-1}\mathbf{A}\underline{S} \approx \mathbf{\Lambda}^{-1}\underline{S}. \tag{5.3}$$

Tekintsük ekkor $\mathbf{W}$ inverzét:

$$\mathbf{W}^{-1} = \hat{\mathbf{A}}\mathbf{\Lambda} = \begin{bmatrix} \hat{A}_{11}\Lambda_1 & \hat{A}_{12}\Lambda_2 & \cdots & \hat{A}_{1M}\Lambda_M \\ \hat{A}_{21}\Lambda_1 & \hat{A}_{22}\Lambda_2 & \cdots & \hat{A}_{2M}\Lambda_M \\ \vdots & \vdots & \ddots & \vdots \\ \hat{A}_{M1}\Lambda_1 & \hat{A}_{M2}\Lambda_2 & \cdots & \hat{A}_{MM}\Lambda_M \end{bmatrix} \tag{5.4}$$

Vegyük észre, hogy $\mathbf{W}^{-1}$ oszlopaiban megjelennek a Λ^{-1} skálázófaktorok multiplikatív inverzei, tehát ha konstruálunk egy diagonális mátrixot melynek minden elme a $\mathbf{W}^{-1}$ azonos sorának egy eleme, a skálázást ki tudjuk küszöbölni (az helyette megjelenő $\hat{A}_{mn}$ faktorok hatását később tárgyaljuk). Vegyük tehát a $\mathbf{W}_{Diag}$ mátrixot, melyet úgy kapunk hogy $\mathbf{W}^{-1}$ diagonális elemein kívül mindent 0-ra állítunk

$$\mathbf{W}_{Diag} = \begin{bmatrix} \hat{A}_{11}\Lambda_1 & 0 & \cdots & 0 \\ 0 & \hat{A}_{22}\Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{A}_{MM}\Lambda_M \end{bmatrix} = \hat{\mathbf{A}}_D\mathbf{\Lambda}, \tag{5.5}$$

ahol $\mathbf{A}_D = \mathrm{Diag}(\hat{A}_{11}, \hat{A}_{22} \dots \hat{A}_{MM})$. Ekkor:

$$\underline{\hat{S}_{\mathrm{obs}}} = \mathbf{W}_D\underline{\hat{S}} = \hat{\mathbf{A}}_D\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\hat{\mathbf{A}}^{-1}\mathbf{A}\underline{S} \approx \hat{\mathbf{A}}_D\underline{S}. \tag{5.6}$$

Tehát a módszer használatával valóban sikerült a $\mathbf{\Lambda}$ skálázást kiküszöbölni, viszont így az S_m forrást csak egy A_{mm} faktor erejéig tudjuk meghatározni. Vegyük észre viszont hogy ha minden frekvenciabinre a becsült forrás $A_{mm}S_m$, $\mathbf{A}$ pedig az adott frekvenciasávban a keverőmátrix, a konvolúciós tétel értelmében a következő összefüggést írhatjuk fel:

$$\hat{S}_{\mathrm{obs}_i}[n] = A_{ii}[f]S_i[f] \Longleftrightarrow \hat{s}_{\mathrm{obs}_i}[n] = h_{ii}[n] \circledast s_i[n] \approx h_{ii}[n] * s_i[n], \tag{5.7}$$

ahol h_{ii} az i-edik forrás és az i-edik mikrofon közt számított impulzusválasz. Tehát a becsült források lényegében az eredeti forrásoknak a mikrofonpozíciókban mért jelei lesznek ezt a módszert használva (innen a megfigyelési térbe vetítés elnevezés), ez pedig nem túlzottan reflektív környezetet feltételezve igen jól közelíti az eredeti források jelét. Megjegyzendő, hogy a levezetés során permutációs bizonytalanság nélküli esetre vizsgáltuk ugyan a módszert, könnyen belátható azonban [19], hogy a permutációs bizonytalanság nem befolyásolja a működését.

5.3. Permutációs bizonytalanság megoldása

A permutációs bizonytalanság megoldására alapvetően két alapvető megközelítéssel találkozhatunk a szakirodalomban. A forrás modellezés alapú módszerek azt használják ki,

hogy a valós hangforrások jelei frekvenciatartományban tipikusan jól csatoltak azaz a szomszédos frekvenciabinek energiája valóságos jelekre a legtöbbször igen hasonló. Ilyen módszer például a [15]-féle időtartományi forrásmodellezés. A másik megközelítés a csatornamodellezés, mely arra épül hogy a források és a mikrofonok közti átviteli függvény közeli frekvenciabinekre igen hasonló, tehát a demix mátrixoknak is hasonlónak kell lenniük a szomszédos frekvenciabinekben [27]. Ebben a dolgozatban a [18] féle Reduced likelihood ratio jump (RLRJ) módszert használjuk, mely [19]-ben bevezetett Likelihood Ratio Jump módszeren alapul. Az RLRJ egy idő-frekvenciatartományi forrásmodellezés alapú módszer. Gyakorlati tapasztalatok alapján megállapítható, hogy a valóságos hangforrások a frekvenciatartományban erősen szupergaussiak (csúcsosságuk igen nagy), eloszlásuk jól közelíthető komplex Laplace-eloszlással (míg időtartományban kevésbé jól közelíthetők valamilyen jellegzetes eloszlással), így a forrásmodellezés frekvenciatartományi megközelítése célszerűbbnek tűnik. A tisztán frekvenciatartományi forrásmodellezés fő problémája viszont a források időfüggő energiaeloszlásának (energy envelope) modellezhetőségének a hiánya. A probléma megoldása az idő-frekvenciatartományi megközelítés használata, azaz az ablakozott ferkvenciatartományi megközelítés használata. Az LRJ által használt forrásmodell a következő:

$$p(s_i[f,t]) \propto \frac{1}{\beta_i[t]} \exp\left|-h\frac{s_i[f,t]}{\beta_i[t]}\right|, \tag{5.8}$$

ahol $s_i[f,t]$ az i-edik forrás STFT transzformáltjának f frekvenciabinjének t időindexű blokkja, h a forrás struktúráját leíró függvény, $\beta_i[t]$ pedig a forrás időfüggő energiatartalmát írja le. Szemléletesen a modell azt írja le, hogy a forráseloszlás egy adott STFT ablakban frekvenciatartományban egy 0 várható értékű komplex Laplace eloszlás melynek skálaparamétere (így szórása is), az adott STFT ablak energiatartalmával (hangosságával) arányos. Levezethető, hogy $\beta_i[t]$ a következő alakban számítható [19]:

$$\beta_i[t] = \frac{1}{F}\sum_{f=1}^{F}|s_i[f,t]|, \tag{5.9}$$

ahol F az STFT ablakhossz. Belátható az is, hogy az (5.8) modell alapján maximum likelihood becslést használva [19] annak valószínűsége, hogy az f frekvenciabinben a jelenleg i indexű forrás jelenlegi STFT spektrum szerinti tényleges indexe j:

$$\gamma_{ij} = \frac{1}{T}\sum_{t=1}^{T}\frac{|s_i[f,t]|}{\beta_j[t]}. \tag{5.10}$$

Legyen $\underline{p^f} \in \mathbb{R}^M$ egy permutációs vektor, melynek komponensei az $\{1,2,\ldots,M\}$ halmaz elemeinek egy permutációja. Jelölje P_p^f annak a valószínűségét, hogy a többi frekvenciabin permutációját fixálva az f frekvenciabin optimális permutációját $\underline{p^f}$ adja meg (az s_i forrás optimális indexe p_{f_i}). Ekkor:

$$P_p^f = \prod_{m=1}^{M}\gamma_{mp_m^f}. \tag{5.11}$$

Az optimális permutációt pedig P_p^f maximalizálásával érjük el:

$$\underline{p_{\text{opt}}^f} = \underset{\underline{p^f}}{\operatorname{argmax}}\, P_p^f. \tag{5.12}$$

Vegyük azonban észre, hogy az optimum meghatározásához $M!$ esetet kell végig próbálnunk minden frekvenciabinre, valamint a mivel a $\beta_i[t]$ értékei kezdetben biztosan nem a tényleges szeparált jelek időfüggő energiaprofilját adják meg (hiszen ekkor még frekvencia-

binenként ismeretlen módon permutálva vannak a jelek), csak iteratív módon, az optimális permutációkat mindig az adott β-khoz kiszámítva várhatjuk, hogy az algoritmus a globálisan optimális permutációkhoz konvergáljon. Emiatt az algoritmus számításigénye igen gyorsan növekszik a források számának növelésével, így az FD-ICA alapú beamformerhez az RLJ nem optimális. Megmutatható azonban, hogy az algoritmus számítási komplexitása nagymértékben csökkenthető azzal, ha az adott iterációban az adott frekvenciabinben a cserék számát legfeljebb 1-re korlátozzuk (a permutációs vektort úgy definiáljuk, hogy az annak elmei $[1, 2, \ldots, M]$ vektor elemeiből legfeljebb kettőt megcserélve kaphatók meg), eközben a konvergenciatulajdonságok nem romlanak jelentősen. Ez, az úgynevezett RLRJ algoritmus, mely a következő módon vezethető be:

1. Határozzuk meg a lehetséges $\underline{p}$ permutációs vektorok P_2 halmazát az $[1, 2, \ldots, M]$ vektor pontosan két elemének összes lehetséges módon történő megcserélésével.

2. Számítsuk ki az aktuális STFT spektrumok alapján $\beta_i[t]$-t az (5.9) összefüggés használatával.

3. Minden frekvenciabinben $\forall \underline{p} \in P_2$-re számítsuk ki P_p^f értékét az (5.11) összefüggés szerint, valamint számítsuk a jelenlegi $\underline{p_{\text{curr}}} = [1, 2 \ldots, M]$ permutáció valószínűségét ($P_{p_{\text{curr}}}^f$), ha bármelyik $\underline{p} \in P_2$-re $P_p^f > P_{p_{\text{curr}}}^f$, akkor permutáljuk az aktuális frekvenciabin forrásait az adott $\underline{p}$ vektornak megfelelően

4. Ha az előző lépésben bármelyik frekvenciabinben $P_p^f > P_{p_{\text{curr}}}^f$ teljesült, ugorjunk a 2. lépéshez

6. fejezet

FD-ICA alapú mikrofontömbös nyalábformálás

A permutációs valamint a skálázási bizonytalanság kiküszöbölésével tehát alkothatunk egy elméletileg működőképes FD-ICA megoldást, melynek magas szintű működését a 6.1. ábra szemlélteti.

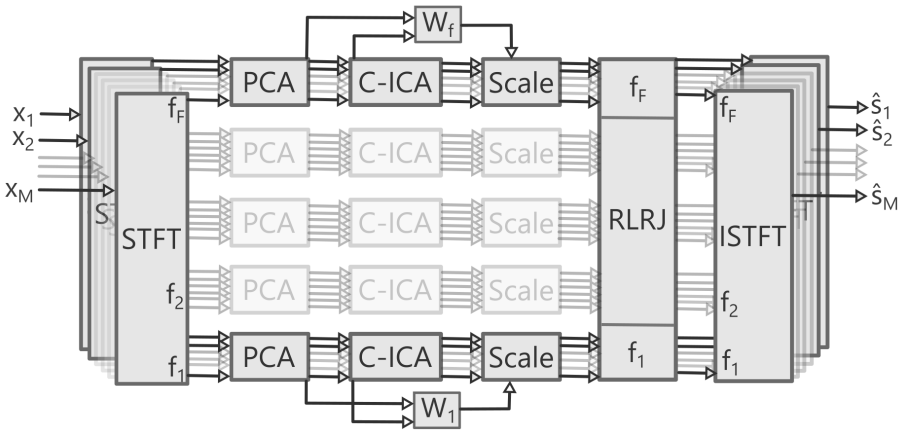
Az $(M \times M)$-es (azaz azonos, M számú forrást és mikrofont feltételező) FD-ICA algoritmus működése a következő:

1. Az $\underline{x_m} \in \mathbb{R}^N$ bemeneti jelvektorokra F ablakhosszú STFT transzformációt végrehajtva megkapjuk a kevert jelek idő-frekvenciatartományi reprezentációját tartalmazó $\mathbf{X_m}$ mátrixokat.

2. Ezekből a 3.4. alfejezetben leírt módon kiszámítjuk az adott frekvenciabinhez tartozó DFT együtthatók (lényegében mint idősorok) által alkotott vektorokból összefűzött $\mathbf{X_f}$ mátrixokat.

3. Minden frekvenciabinre (komplex) PCA használatával a 2.2. alfejezetben leírt módon meghatározzuk a fehérített $\mathbf{Z_f}$ jeleket, valamint a $\mathbf{V_f}$ fehérítő mátrixokat.

4. Minden frekvenciabinre a fehérített $\mathbf{Z_f}$ jelekből komplex ICA (pl. ACMN ICA, vagy RobustICA) használatával kiszámítjuk a szeparált $\hat{\mathbf{S}}_\mathbf{f}$ jeleket, valamint a fehérített jelekre vonatkozó $\mathbf{W_f^{white}}$ demix mátrixokat.



6.1. ábra. Az FD-ICA blokkvázlata

5. Minden frekvenciabinre a szeparált $\hat{\mathbf{S}}_\mathbf{f}$ jelek, illetve a az eredeti jelekre vonatkozó $\mathbf{W_f} = \mathbf{W_f^{white}}\mathbf{V_f}$ szeparációs mátrixok alapján a megfigyelési térbe vetítés technikát használva meghatározzuk a skálázott $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{s}$ jeleket.

6. A $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{s}$ jelek permutációját az RLRJ módszerrel korrigáljuk, így meghatározzuk a szeparált, skálázott, valamint permutáció-korrigált $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{sp}$ jeleket.

7. Az $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{sp}$ mátrixok alapján minden $0 \leq m < M$-re kiszámítjuk az $\hat{\mathbf{S}}_\mathbf{m}$ mátrixokat olyan módon, hogy minden $0 \leq f < F$-re: $\underline{\hat{S}_{m_f}} = \underline{\hat{S}^{sp}_{f_m}}$, Ezzel megkapjuk a a szeparált források STFT trnaszformáltját reprezentáló $\hat{\mathbf{S}}_\mathbf{m}$ mátrixokat.

8. $\hat{\mathbf{S}}_\mathbf{m}$ mátrixokból inverz STFT transzformáció használatával kiszámítjuk A szeparált időtartományi $\underline{\hat{s}_m}$ jeleket

Ezt az FD-ICA architektúrát megvalósítva már a gyakorlatban is azt tapasztalhatjuk, hogy az $(M \times M)$-es esetre igen jó szeparációs teljesítménnyel működik (lásd eredmények), viszont a beamformer-ként való használatra még mindig nem optimális, mivel a beamformereknél tipikusan fennálló $M \gg S$ sajátosságot, azaz hogy a mikrofonjaink száma jelentősen nagyobb, mint a források száma, nem tudjuk kihasználni. Természetesen az technikailag működik, ha éppen S darab mikrofon kimenetén futtatjuk a beamformert, de ekkor a többi mikrofont, melyek jele potenciálisan extra információt hordoz, nem használjuk ki. A másik jelentős extra információ a mikrofontömbös bemenetben a klasszikus FD-ICA-hoz képest az, hogy itt a mikrofonok pozíciói ismertek, ez a geometriai információ pedig (ahogy majd látni fogjuk) a permutációs bizonytalanság megoldásához nyújthat segítséget.

6.1. Több mikrofon kihasználása

Az extra mikrofonok által nyújtott többletinformáció kihasználásával többféle módon próbálkozhatunk. Egy egyszerű módszer a FD-ICA $(M \times M)$-es módban futtatása kevesebb forrás esetére is abban reménykedve, hogy az FD-ICA megtalálja a keverékben a tényleges forrásokat, a többi kimeneti csatornán pedig a háttérzaj jelenik meg (M a mikrofonok száma). Ezeket a zajcsatornákat utófeldolgozásként eltávolítjuk. A módszert kipróbálva azonban azt tapasztalhatjuk, hogy a szeparációs teljesítmény a több mikrofon hozzáadásával jobb esetben nem romlik drasztikusan, rosszabb esetben pedig lényegében egyáltalán nem szeparál a módszer. Ennek oka valószínűsíthetően a fehérítési lépésben keresendő. Vegyük észre ugyanis, hogy az $M > S$-es esetben nem minden főkomponenshez fog tényleges forrás tartozni, hiszen az ún. jeltér biztosan alacsonyabb dimenziós ($\mathbb{R}^S$-beli) lesz, mint az M mikrofon jelének ($\mathbb{R}^M$-beli) tere, így a jeltéren felüli főkomponensek biztosan a (jeltérre ortogonális) zajtérbe fognak tartozni. (Megjegyzendő, hogy ez lényegében a MUSIC nyalábformáló algoritmus [7] alapötlete is.) Ez megmagyarázza, hogy az említett esetben miért teljesít rosszul az FD-ICA, hiszen ekkor az zajtér (alapvetően jóval kisebb energiájú) komponenseit is egységvarianciára hozzuk, épp mint a jelteret, így az ICA algoritmusnak egy olyan keverékből kellene kiválasztani a valós forrásokat, melybe extra zajkomponenseket illesztettünk. Vegyük észre azonban, hogy ezt a problémát igen egyszerűen megoldhatjuk: használjuk a PCA-t dimenzióredukciós módszerként. Ezt a következő módon tehetjük meg:

Adott az $\mathbf{X_f} \in \mathbb{C}^{M \times T}$ f frekvenciabinhez tartozó idő-frekvenciatartományi jelek mátrixa, (feltételezzük, hogy a jeleket már centráltuk). A komplex kovarianciamátrix ekkor

$$\mathbf{\Sigma} = \mathbf{X_f}{\mathbf{X_f}}^H. \tag{6.1}$$

Jelölje $\mathbf{\Sigma}$ sajátértékeit λ_i, hozzájuk tartozó sajátvektorait pedig $\underline{v_i}$, úgy, hogy λ_i az i-edik legnagyobb abszolút értékű sajátérték. Feltételezve, hogy a forrásaink jelentősen "hangosabbak", mint a háttérzaj, egyszerűen belátható, hogy a tényleges forrásokhoz tartozó sajátvektorok sajátértékének abszolút értéke is jelentősen nagyobb a zajtérbe tartozó sajátértékeknél. Ez tehát azt jelenti, hogy ha kiválasztjuk az S legnagyobb sajátértékhez tartozó $\{\underline{v_1}, \underline{v_2}, \ldots, \underline{v_S}\}$ sajátvektorokat, és a hozzájuk tartozó $\{\lambda_1 \lambda_2 \ldots, \lambda_S\}$ sajátértékeket, a dimenzióredukciót, illetve fehérítést végrehajtó $\mathbf{V^D} \in \mathbb{C}^{S \times M}$ PCA mátrix a következő alakban adható meg:

$$\mathbf{V^D} = \mathbf{\Lambda^{S^{\circ -\frac{1}{2}}}} \mathbf{Q^S}, \tag{6.2}$$

ahol

$$\mathbf{Q^S} = \begin{bmatrix} \underline{v_1} \\ \underline{v_2} \\ \vdots \\ \underline{v_S} \end{bmatrix}, \quad \mathbf{\Lambda^S} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_S \end{bmatrix}. \tag{6.3}$$

Ezzel a módszerrel tehát már a fehérítési lépés után a tényleges források terére vetítve (és nyilván egység kovarianciára hozva) kapjuk meg a kevert forrásokat, ezzel a FD-ICA fehérítést követő lépései lényegében módosítás nélkül úgy hajthatók végre, mintha $(S \times S)$-es esetre alkalmaznánk a módszert. Eközben a források terére vetítéstől azt várhatjuk, hogy a több mikrofon használatával a vetített jelek jel-zaj viszonya javulni, fog, ezzel pedig a szeparáció minősége is javul. Megjegyzendő, hogy a források S számát nem szükséges "kézzel" meghatároznunk, azt pl. a MUSIC algoritmust [7] használva meghatározhatjuk előfeldolgozási lépésként a mikrofonok időtartományi jeleiből. Vegyük azonban észre, hogy (a PCA-t követő lépésekben egyedül) a skálázási bizonytalanság kiküszöbölésénél $\mathbf{W_{Diag}}$ konstruálásánál felhasználtuk, hogy $\mathbf{W}$ invertálható, a dimenzióredukciós esetben pedig:

$$\mathbf{W} = \mathbf{V^D W^{white}}. \tag{6.4}$$

Tehát $\mathbf{W} \in \mathbb{C}^{S \times M}$, azaz biztosan nem invertálható. Szerencsére a módszer minimális módosítás után ebben az esetben is használható. Tekintsük az S forrás, M mikrofon instant keverési modellt:

$$\underline{X} = \mathbf{A}\underline{S}, \tag{6.5}$$

ahol $\underline{X} \in \mathbb{C}^M$, $\underline{S} \in \mathbb{C}^S$, $\mathbf{A} \in \mathbb{C}^{M \times S}$ Feltéve hogy a permutációs probléma fennáll, $\mathbf{W}$ demix mátrix a következő alakú:

$$\mathbf{W} = \left(\hat{\mathbf{A}}\mathbf{\Lambda}\right)^+ = \mathbf{\Lambda}^+ \hat{\mathbf{A}}^+ = \mathbf{\Lambda}^{-1} \hat{\mathbf{A}}^+, \tag{6.6}$$

ahol $\mathbf{A}^+$ egy $\mathbf{A}$ mátrix Moore-Penrose féle pszeudoinverze, illetve kihasználtuk, hogy invertálható mátrixokra a pszeudoinverz és az inverz megegyezik.

Ekkor kihasználva, hogy lineárisan független oszlopvektorokból álló $\mathbf{A}$ mátrixra $\mathbf{A}\mathbf{A}^+ = \mathbf{I}$, valamint feltételezve, hogy keverőmátrix oszlopai függetlenek (Ez a gyakorlatban szinte mindig teljesül ha forráspozíciók különbözőek, hiszen különböző pontok között az átviteli függvények szinte mindig különbözőek):

$$\hat{\underline{S}} = \mathbf{W}\underline{X} = \mathbf{\Lambda}^{-1} \hat{\mathbf{A}}^+ \mathbf{A}\underline{S} \approx \mathbf{\Lambda}^{-1}\underline{S}. \tag{6.7}$$

Ekkor $\mathbf{W}$ pszeudoinverzét tekintve, kihasználva hogy $\mathbf{A}^{++} = \mathbf{A}$:

$$\mathbf{W}^{+} = \hat{\mathbf{A}}\mathbf{\Lambda} = \begin{bmatrix} \hat{A}_{11}\Lambda_1 & \hat{A}_{12}\Lambda_2 & \cdots & \hat{A}_{1S}\Lambda_S \\ \hat{A}_{21}\Lambda_1 & \hat{A}_{22}\Lambda_2 & \cdots & \hat{A}_{2S}\Lambda_S \\ \vdots & \vdots & \ddots & \vdots \\ \hat{A}_{M1}\Lambda_1 & \hat{A}_{M2}\Lambda_2 & \cdots & \hat{A}_{MS}\Lambda_S \end{bmatrix}. \tag{6.8}$$

Vegyük észre, hogy ekkor $\mathbf{W}^{+}$ főátlójának elemeiből egy $\mathbf{W_{Diag}} = \mathrm{Diag}(\hat{A}_{11}\Lambda_1, \hat{A}_{22}\Lambda_2\hat{A}_{SS}\Lambda_S)$ mátrixot konstruálva épp az 5.2. alfejezetben levezetett skálázómátrixot kapjuk, az FD-ICA dimenzióredukció utáni komponensei által használt látszólagos $S \times S$-es esetre.

6.2. Ismert mikrofonpozíciók kihasználása

A klasszikus FD-ICA algoritmusokhoz képest a másik "többletinformáció" a mikrofonpozíciók ismert relatív pozíciója az ICA alapú beamformer megvalósításánál. Az ismert mikrofonpozíciók kihasználására az FD-ICA szakirodalmában több módszert is kidolgoztak, tipikusan a permutációs bizonytalanság megoldására. Felmerülhet a kérdés, hogy miért akarjuk beamforming alapon is megoldani a permutációs problémát, hiszen azt az RLRJ-vel már úgyis megoldottuk. Ennek oka az, hogy az RLRJ a kezdeti permutációktól függően bizonyos esetekben képes erősen nem optimális megoldáshoz konvergálni. Ezt elkerülhetjük, ha a kezdeti permutációk már nagyjából jók, illetve megmutatható általánosságban is, hogy jobb kezdeti permutációk alapján gyakran a tényleges permutációkhoz közelebbi megoldáshoz konvergál az RLRJ [18]. A beamforming alapú permutációs módszer célja tehát előfeldolgozásként szolgálni az RLRJ algoritmushoz.

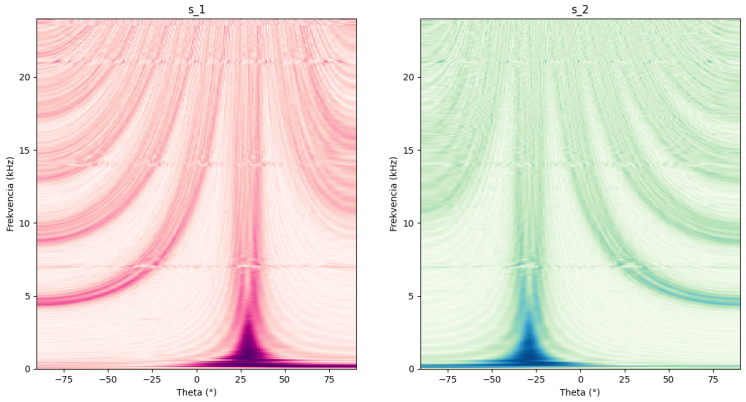
A klasszikus mikrofontömbös nyalábformálás alapfeladata adott irányokból érő jelek felerősítése, közben a nemkívánatos irányokból érkező jeleket elnyomjuk. Ezeket az irányokat az ún. Direction of Arrival (DOA) írja le, mely lineáris mikrofontömb esetén azt mutatja meg (2 dimenezióban), hogy az adott jel forrásiránya milyen szöget zár be a mikrofontömb által meghatározott szakasz felezőmerőlegesével. Az adott irányok elnyomását, illetve felerősítését a klasszikus beamforming algoritmusok általában a mikrofonok kimenetének megfelelő impulzusválaszú szűrőkkel szűrésével, majd a szűrt jelek összeadásával érik el. (A beamforming algoritmusok fő kihívása így szűrőarchitekúrák megválasztása, illetve a szűrőegyüttható optimalizálása valamilyen kritérium szerint). Egy beamformer fontos jellemzője az iránykarakterisztikája, mely azt mutatja meg, hogy egy adott frekvencián a beamformer erősítése hogyan alakul a DOA függvényében. Megmutatható, hogy egy lineáris mikrofontömb alapú beamformer iránykarakterisztikáját a következő összefüggés adja meg [11]:

$$F\left(f, \underline{w^f}, \theta\right) = \sum_{m=1}^{M} \tilde{w}_m^f \exp\left(j2\pi f(k-1)d\frac{\sin(\theta)}{c}\right), \tag{6.9}$$

ahol d a mikrofonok távolsága, θ a DOA, c a hangsebesség, $w_m^{ph}(f)$ pedig:

$$\tilde{w}_m^f = \frac{w_m^f}{|w_m^f|}, \tag{6.10}$$

ahol $w_m(f)$ az m indexű mikrofon szűrőjének átviteli karakterisztikája az f frekvencián. Vegyük észre, hogy az FD-ICA frekvenciafüggő $\mathbf{W_f}$ demix mátrixának elemei értelmezhetők ilyen átviteli karakterisztika értékeként az adott frekvencián, így az FD-ICA alapú szeparációs módszerre is meghatározhatjuk az az adott frekvenciákon vett iránykarakte-



6.2. ábra. FD-ICA iránykarakterisztikák abszolút értéke

risztikákat. Ehhez mindössze az adott frekvencián az $1 \leq i < S$ forráshoz tartozó $\underline{w_i}(f)$ vektorokat kell meghatároznunk, melyek a $\mathbf{W_f}$ mátrix i-edik sorvektoraként állnak elő. Erre alapuló permutációs módszert dolgozott ki [19] valamint [25] is. Ezek a módszerek azt használják ki, hogy ahhoz hogy egy forrást szeparálni tudjunk, a többi forrás irányába az iránykarakterisztika közel 0 értéket kell hogy felvegyen minden frekvenciabinben (ez az úgynevezett null-beamformer megközelítés). Az általunk megvalósított dimenzióredukciós FD-ICA módszer iránykarakterisztikáit vizsgálva $M \gg S$-re viszont azt tapasztaljuk, hogy azok a kisebb frekvenciabinekben majdnem minden esetben egy irány közelében vesznek fel nagyobb értéket, ez pedig egy valódi forráshoz tartozó irány (nagyobb frekvenciákon a mikrofonok térbeli távolsága sérti a térbeli mintavételi törvényt így aliasing effektusok jelennek meg az iránykarakterisztikán). A 6.2. ábrán egy ilyen iránykarakterisztika abszolút értéke látható, melyet két forrás dimred. FD-ICA módszerünk demix mátrixai alapján generáltunk (megjegyzendő, hogy itt a szemléltetés kedvéért a permutációs problémát már megoldottuk). Az ábrán a sötétebb színek nagyobb értékeket jelölnek.

Vegyük észre, hogy ezeket az iránykarakterisztikákat felhasználhatjuk a permutációs probléma megoldására: permutáljuk úgy a különböző frekvenciabinekben a forrásokat, hogy a hozzájuk tartozó iránykarakterisztikák értéke az adott forráshoz minden frekvenciabinben ugyanazon DOA környékén legyen maximális. A beamforming alapú permutációs algoritmus tehát a következő módon valósítható meg:

1. Határozzuk meg a tényleges θ_i DOA-kat például MUSIC [7] algoritmus használatával, legyen ezek vektora $\underline{\theta} = [\theta_1, \theta_2, \dots \theta_S]$

2. Számítsuk ki a minden becsült forrás esetén minden frekvenciabinre az iránykarakterisztika abszolút értékének integrálját a valós forrásirányok környezetében (nyilván a gyakorlatban numerikus integrálást alkalmazva). Legyen $F^f_{s\theta_s}$ az f frekvenciabinben a jelenlegi permutációban s indexű forrás iránykarakterisztikájának abszolút integrálja a θ_i irány környezetében:

$$F^f_{s\theta_s} = \int_{\theta min}^{\theta max} F\left(f, \underline{W^f_s}, \theta\right) d\theta \tag{6.11}$$

$$\theta_{min} = \max\left(-\frac{\pi}{2}, \theta_s - \min\left(\frac{\Delta\theta}{2}, \Delta\theta_0\right)\right) \tag{6.12}$$

$$\theta_{max} = \min\left(\frac{\pi}{2}, \theta_s + \min\left(\frac{\Delta\theta}{2}, \Delta\theta_0\right)\right), \tag{6.13}$$

ahol $\underline{W_s^f}$ a $\mathbf{W^f}$ demix mátrix s-edik sorvektora, $\Delta\theta$ a $\underline{\theta}$ elemeinek páronkénti különbségének minimuma, $\Delta\theta_0$ pedig a megengedett maximális "szélessége" a θ_i környezetének (ezt a mérések során 5°-ra állítottuk).

3. Konstruáljuk meg az $\mathbf{F^f}$ mátrixot minden frekvenciabinre a következő módon:

$$\mathbf{F^f} = \begin{bmatrix} F^f_{1\theta_1} & F^f_{1\theta_2} & \cdots & F^f_{1\theta_S} \\ F^f_{2\theta_1} & F^f_{2\theta_2} & \cdots & F^f_{2\theta_S} \\ \vdots & \vdots & \ddots & \vdots \\ F^f_{S\theta_1} & F^f_{S\theta_2} & \cdots & F^f_{S\theta_S} \end{bmatrix}. \tag{6.14}$$

4. Az optimális permutációt ekkor egy f frekvenciabinben a $\mathbf{F^f}$ szomszédossági mátrix által reprezentált páros gráf legnagyobb összsúlyú teljes párosítása adja meg (vagy ezzel ekvivalensen az $\mathbf{F^f}$ mátrixban a legnagyobb összsúlyú bástyaelrendezés). Ezt az ún. magyar módszerrel [13] határozhatjuk meg.

5. Legyen az optimális párosítás egy i éle $v_i = [s_k, \theta_l]$ $(1 \leq i, k, l \leq S)$, ekkor az optimális permutációt az adott f binben a jelenlegi k indexű forrás l indexre "mozgatásával" érjük el minden i élre.

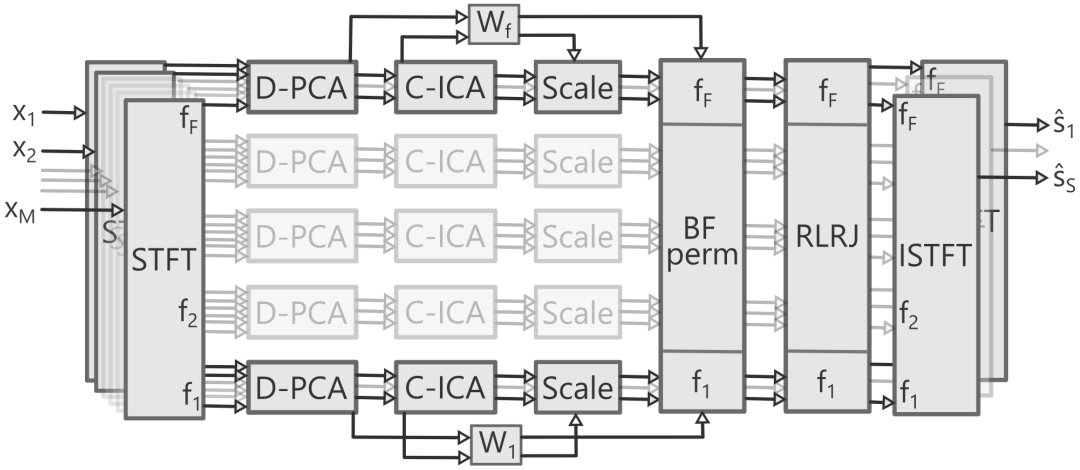
6. A párosítást illetve a permutációkat minden frekvenciabinre elvégezve megkapjuk az iránykarakterisztikák szerint optimális permutációt.

7. fejezet

Az FD-ICA-Beamformer architektúrája

7.1. Az architektúra leírása

A mikrofontömbös sajátosságokat kihasználva a 6. fejezetben bemutatott FD-ICA architektúra módosításával megalkothatjuk a végleges, mikrofontömbös jelfeldolgozásra optimalizált FD-ICA alapú szeparációs architektúrát, hívjuk ezt a továbbiakban BF-FD-ICA-nak. Az architektúra magas szintű áttekintését a 7.1 ábra szemlélteti. A BF-FD-ICA architek-



7.1. ábra. BF-FD-ICA blokkvázlata

túra működése tehát a következő:

1. Az $\underline{x_m} \in \mathbb{R}^N$, $1 \leq m \leq M$ bemeneti jelvektorokra F ablakhosszú STFT transzformációt végrehajtva megkapjuk az kevert jelek idő-frekvenciatartományi reprezentációját tartalmazó $\mathbf{X_m}$ mátrixokat.

2. Ezekből a 3.4 alfejezetben leírt módon kiszámítjuk az adott frekvenciabinhez tartozó DFT együtthatók (lényegében mint idősorok) által alkotott vektorokból összefűzött $\mathbf{X_f} \in \mathbb{C}^{M \times T}$ mátrixokat

3. Minden frekvenciabinre végrehajtjuk a dimenzióredukciót illetve fehérítést megvalósító dimenzióredukciós PCA algoritmust (D-PCA), 6.1 alfejezetben leírt módon, ezzel megkapjuk vetített, fehérített $\mathbf{Z_f} \in \mathbb{C}^{S \times T}$ jeleket, valamint a $\mathbf{V_f^D} \in \mathbb{C}^{S \times M}$ fehérítő mátrixokat. (az S értékét akár manuálisan, akár MUSIC algoritmussal határozhatjuk meg)

4. Minden frekvenciabinre a fehérített $\mathbf{Z_f}$ jelekből komplex ICA (pl. ACMN ICA, vagy RobustICA) használatával kiszámítjuk a szeparált $\hat{\mathbf{S}}_\mathbf{f} \in \mathbb{C}^{S \times T}$ jeleket, valamint a $\mathbf{W_f^{white}} \in \mathbb{C}^{S \times S}$, fehérített jelekre vonatkozó demix mátrixokat.

5. Minden frekvenciabinre a szeparált $\hat{\mathbf{S}}_\mathbf{f}$ jelek, illetve a az eredeti jelekre vonatkozó $\mathbf{W_f} = \mathbf{W_f^{white}} \mathbf{V_f^D}$ szeparációs mátrixok alapján az 6.1 alfejezetben levezetett módosított megfigyelési térbe vetítés technikát használva meghatározzuk a skálázott $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{s}$ jeleket.

6. A $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{s}$ jelek optimális permutációját a 6.2 alfejezetben levezetett beamformer alapú permutációs módszerrel a $\mathbf{W_f}$ mátrixokból számított iránykarakterisztikák alapján meghatározzuk, majd kiszámítjuk az eszerint permutált $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{sb}$ jeleket

7. A $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{sb}$ jelek permutációját az RLRJ módszerrel finomítjuk, így meghatározzuk a szeparált, skálázott, valamint permutáció korrigált $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{sbp}$ jeleket.

8. Az $\hat{\mathbf{S}}_\mathbf{f}^\mathbf{sbp}$ mátrixok alapján minden $0 \leq s \leq S$-re kiszámítjuk az $\hat{\mathbf{S}}_\mathbf{s}$ mátrixokat olyan módon, hogy minden $0 \leq f \leq F$-re: $\underline{\hat{S}_{s_f}} = \underline{\hat{S}_{f_s}^{sbp}}$. Ezzel megkapjuk a a szeparált források STFT transzformáltját.

9. $\hat{\mathbf{S}}_\mathbf{s}$ mátrixokból inverz STFT transzformáció használatával kiszámítjuk A szeparált időtartományi $\underline{\hat{s}_s}$ jeleket

7.2. implementációs kérdések

Az előző alfejezetben bemutatott architektúra implementációját Python környezetben végeztük, Numpy matematikai könyvtár használatával.

8. fejezet

Kiértékelési metodológia

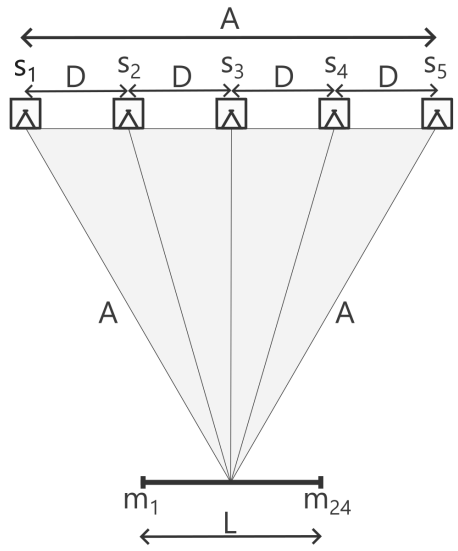
Az implementáció után az egyetlen hátralévő feladat a megvalósított BF-FD-ICA módszer működőképességének vizsgálata. A mikrofontömbös nyalábformálás illetve az ICA szakirodalmában is azzal találkozhatunk, hogy a szeparációs módszerek teljesítményét nagyrészt szimulált bemeneteken kipróbálva vizsgálják. Ez olyan szempontból igen előnyös megoldás, hogy a szimulált esetekre viszonylag egyszerűen számolhatunk különböző teljesítmény metrikákat, hiszen a források közti átviteli függvényeket valamint az elvárt szeparált kimeneteket is könnyedén meg tudjuk határozni. A szimulált jeleken történő kiértékelésnek azonban az a jelentős hátránya a mi esetünkben, hogy a rendelkezésünkre álló szimulációs keretrendszerekkel, csak tökéletes, végtelen nyílt térben számított, reflexiómentes, konstans abszolút értékű átviteli karakterisztikájú jelterjedéseket tudunk szimulálni, ez pedig a gyakorlatban létrejövő akusztikus környezeteknél jelentősen egyszerűbb terjedési modellt hoz magával. Ennek eredménye az, hogy abból, hogy a megoldásunk egyszerű szimulált bemenetre jól működik, nem feltétlenül következik, hogy módszer valós bemenetekre is jól fog működni. Emiatt, hogy valós képet kapjunk a megoldásunk teljesítményéről, valós, mért adatokon értékeljük ki a működését. Ez természetesen külön kihívásokkal jár hiszen valós akusztikus környezetben a terjedési paramétereket pontosan nem tudjuk meghatározni, illetve az elvárt kimenetek meghatározása sem triviális. A fejezet további részében így bemutatjuk a használt mérési elrendezést, valamint megmutatjuk, hogy milyen metrikákat használhatunk szeparációs teljesítmény valós mérésekre történő kiértékelésére.

8.1. A mérési elrendezés bemutatása

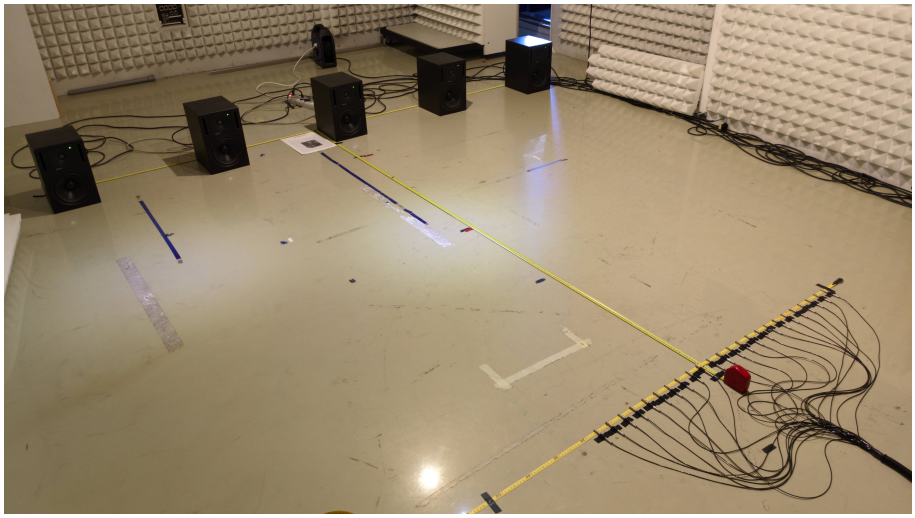
A BF-FD-ICA módszer értékelését tehát valós mért jeleken végeztük. A méréshez egy *Gfai mcdRec* típusú mérési adatgyűjtőt használtunk 24 darab omnidirekcionális, közel konstans frekvenciamenetű mikrofonból álló mikrofontömbbel. A mikrofonokat egy egyenes mentén, lineáris mikrofontömbként helyeztük el a padlóra, állandó $d = 0.05$ m távolságra egymástól. A teljes mikrofontömb hossza így $L = 1.15$ m. A méréshez 5 darab forrást használtunk, melyek *Genelec 1030A* típusú monitorhangfalak voltak. A hangfalakat úgy helyeztük el a padlóra egy egyenes mentén, hogy a két szélső hangfal és a mikrofontömb középpontja egyenlő oldalú háromszöget alkosson, valamint a mikrofontömb és a hangfalak által alkotott egyenesek párhuzamosak legyenek egymással. A hangfalak távolságát $D = 0.7$ m-re állítottuk, így a szabályos háromszög egy oldala $A = 2.8$ m hosszú. A mérést egy terem a teremben elven épült félsüket mérőhelyiségben (A padló szinte tökéletesen reflektív, a többi felület hangelnyelő borítású) végeztük, így mivel a hangforrásokat és a mikrofonokat is a padlóra helyeztük, így a teremben reflexiók csak igen kis mértékben léptek fel (A teremben voltak egyéb tárgyak, azokról természetesen verődtek vissza hanghullámok). Könnyen lát-

ható, hogy a DOA vektor ebben a mérési elrendezésben: $\underline{\theta} \approx [-30°, -16.1°, 0°, 16.1°, 30°]$. A mérési elrendezést a 8.1. ábra, illetve a 8.2. fotó szemlélteti.

A mérést kontrollált, reflektív környezetben is megismételtük, ehhez a fentebb ismertetett mérési elrendezést annyiban módosítottuk, hogy a mikrofontömb mögé, azzal párhuzamosan $R = 2$ m távolságra, a középpontját a középső hangszóró és a mikrofontömb középpontja által meghatározott egyenesre illesztve bekerült egy erősen reflektív falemez. A falemez függőlegesen megdöntésre került a mikrofontömb felé, így a mikrofontömbnél megjelent egy igen jelentős intenzitású reflekció, megközelítőleg $\tau = \frac{2R}{c} \approx 0.012$ s időkésleltetéssel. A reflektív mérési elrendezést a 8.3 fotó szemlélteti.

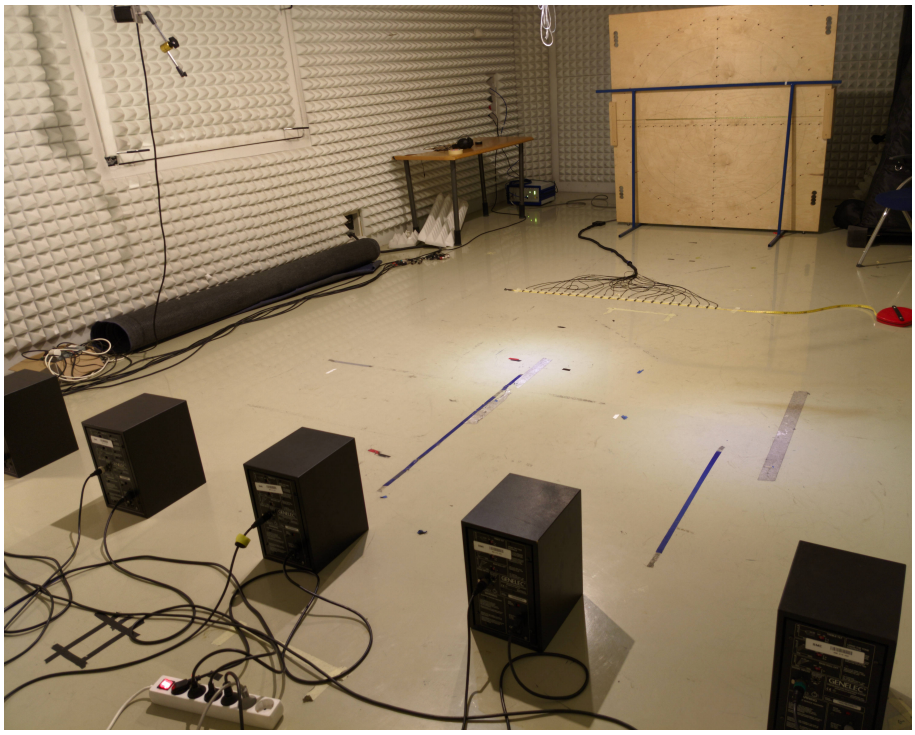


8.1. ábra. A használt mérési elrendezés



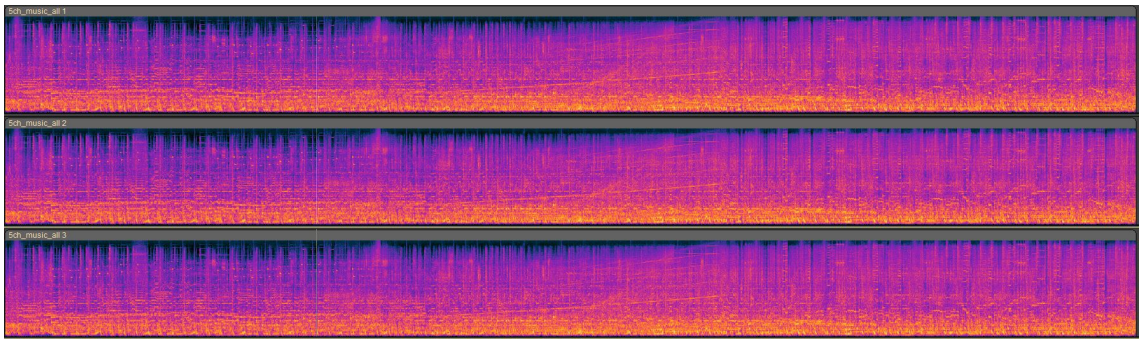
8.2. ábra. Mérési elrendezés a valóságban

A méréshez spektrálisan, és időtartományban is igen komplex elektronikus zenei részleteket használtunk. Azok hangerejét normalizáltuk, és 1 perces szegmenseket vágtunk ki belőlük, melyekből összeállítottuk a mérőjeleket. A szeparációs teljesítmény értékeléséhez használt referenciajeleket, a mérőjeleket külön-külön kijátszva (ugyanazon a hangszórón mint ahol az egyszerre játszott kevert esetben szólt), majd a mikrofontömb egyik középső



8.3. ábra. Reflektív mérési elrendezés a gyakorlatban

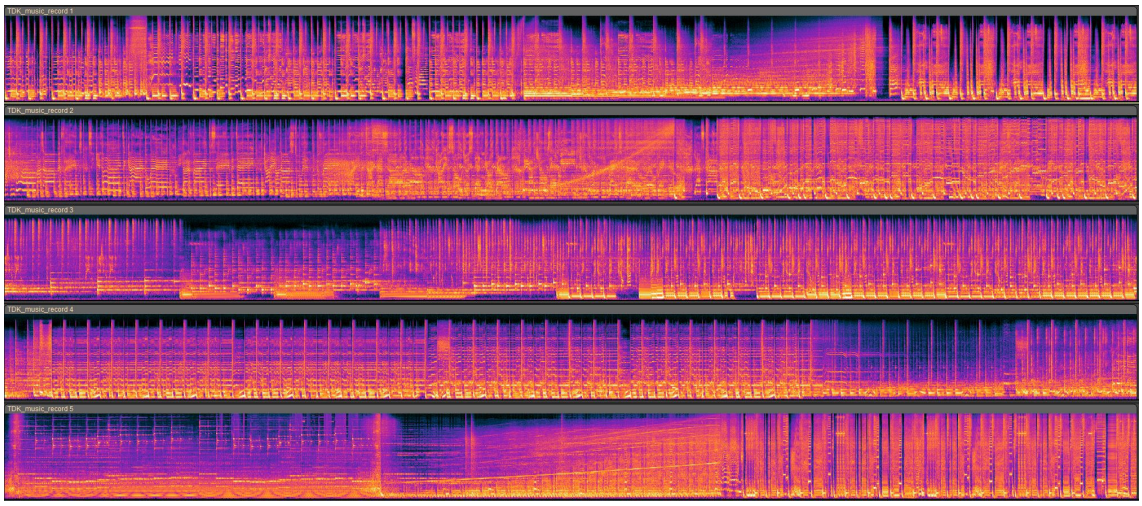
(12-es) indexű mikrofonjával felvéve hoztuk létre. A mért kevert jelek közül néhányat a 8.4 ábra, a referenciajeleket a 8.5 ábra szemlélteti.



8.4. ábra. Kevert jelek spektogramjai

8.2. Teljesítménymetrikák valós mérési adatokon

A szeparációs teljesítmény szubjektív értékelése triviális feladat, akár a szeparált források meghallgatása alapján, vagy akár az STFT spektogramokra tekintve, akár a generált frekvenciafüggő iránykarakterisztikák alapján. Ezt azért tudjuk ilyen egyszerűen megtenni, mert az ember jelentős priori ismerettel bír azzal kapcsolatban, hogy a szeparált jeleknek milyen jellegűnek kell lennie az adott reprezentációban. Ezt azonban nem tudjuk kvantitatívan leírni, így a kiértékelési metrika származtatásánál egy módszer marad: valamilyen hibát számolni az elvárt kimenethez képest. Vegyük észre azonban, hogy valós mérések esetére az elvárt kimenetek meghatározása sem triviális, mivel az eredeti, hangszórókon kijátszott jelek a hangszóró valamint a hangszóró és a mikrofonok közti akusztikus környezet átvitelének megfelelően torzulnak, így az eredeti jelekhez számított hibában ezen



8.5. ábra. Referenciajelek spektogramjai

torzulások hatása is benne lesz, ez pedig nyilván nem kívánatos. Ez azonban még viszonylag egyszerűen megoldható probléma a jelek külön-külön kijátszásával, majd felvételével (ezekre referenciajelekként hivatkozunk a továbbiakban). A jelentősebb kérdés a hibametrika meghatározása. A pusztán időtartományban számított (pl. $\mathbf{L}_2$) hiba használata nem célravezető, hiszen az nem veszi figyelembe a szeparációs módszerek teljesítményének frekvenciafüggését. Vegyük észre, hogy a tisztán frekvenciatartományi hibaképzés sem célszerű, hiszen ez a jelek időfüggő energiaprofilját nem veszi figyelembe. A hiba számítását ezért az általunk javasolt módszerben idő-frekvenciatartományban végezzük, hiszen így az időbeli illetve frekvenciatartományi viselkedést is figyelembe tudjuk venni. Ezek alapján az ebben a dolgozatban használt SNR-ként hivatkozott hibametrikát a következő módon számoltuk:

1. Legyenek a szeparált időtartományi jelek mint sorvektorok $\underline{\hat{s}_i}$, a referenciajelekből pedig $\underline{s_i^{ref}}$ minden $0 < i \leq S$-re. Feltételezzük hogy a szeparált, valamint a referenciajelek sorrendje megegyezik. (Ezt akár manuálisan meghatározva biztosíthatjuk, akár a forráspozíciók ismeretében az iránykarakterisztikák rendezésével)

2. Normalizáljuk a szeparált valamint a referencia jelek jelszintjét, egy adott frekvenciasávban mért RMS teljesítményük alapján. Erre a különbségképzés alapú hiba miatt van szükség. A jelszint számítást azért végezzük egy adott frekvenciasávban (a gyakorlatban az $[1000\ \text{Hz}, 5000\ \text{Hz}]$ frekvenciasávot használtuk), mivel a normalizálást olyan sávban érdemes végezni ahol a legjobb a szeparációs teljesítmény (hiszen ekkor normalizálunk legnagyobb a tényleges jelteljesítmény alapján).

3. Korreláltassuk időtartományban az összetartozó $\underline{\hat{s}_i}$ illetve $\underline{s_i^{ref}}$ jeleket. A korreláció maximumhelye adja meg a szinkronizációhoz szükséges időeltolást az egyes jelek között. Ezzel korrigálva megkapjuk a normalizált és szinkronizált $\underline{\hat{sn}_i}$ és $\underline{sn_i^{ref}}$ jeleket.

4. Számítsuk $\underline{\hat{sn}_i}$ és $\underline{sn_i^{ref}}$ jelek STFT spektrumának abszolút értékét. Legyenek ezek:

$$\mathbf{S_i} = \left|\text{STFT}(\underline{\hat{sn}_i})\right| \tag{8.1}$$

$$\mathbf{S_i^{ref}} = \left|\text{STFT}(\underline{sn_i^{ref}})\right| \tag{8.2}$$

5. Számítsuk ki az idő-frekvenciatartományi $\mathbf{S_i^{err}}$ hibát, ezt tekintjük a szeparált jel zajkomponensének, ez alapján megadható a kívánatos jelkomponens $\mathbf{S_i^{sig}}$ is:

$$\mathbf{S_i^{err}} = \left| \mathbf{S_i^{ref}} - \mathbf{S_i} \right| \tag{8.3}$$

$$\mathbf{S_i^{sig}} = \mathbf{S_i} - \mathbf{S_i^{err}} \tag{8.4}$$

6. Majd határozzuk meg a frekvenciafüggő jel illetve zajamplitúdók RMS (Root Mean Square) értékét:

$$err_i[f] = \sqrt{\sum_{t=1}^{T} S_i^{err}[f,t]^2} \tag{8.5}$$

$$sig_i[f] = \sqrt{\sum_{t=1}^{T} S_i^{sig}[f,t]^2} \tag{8.6}$$

   Ahol T az STFT időszegmenseinek a száma.

7. Innen a frekvenciafüggő jel-zaj amplitúdó arányok meghatározhatók az i rekonstruált forrás esetére:

$$SNR_i[f] = \frac{sig_i[f]}{err_i[f]} \tag{8.7}$$

   ez decibel skálán megadva:

$$SNR_{i_{dB}}[f] = 20 \log_{10} \left(\frac{sig_i[f]}{err_i[f]} \right) \tag{8.8}$$

8. Ismételjük megy az SNR számítást $\forall \underline{\hat{s}_i}$-re majd számítsuk ki az átlagos SNR-t a következő összefüggés szerint:

$$SNR_{avg_{dB}}[f] = \sum_{i=1}^{S} SNR_{i_{dB}}[f] \tag{8.9}$$

9. fejezet

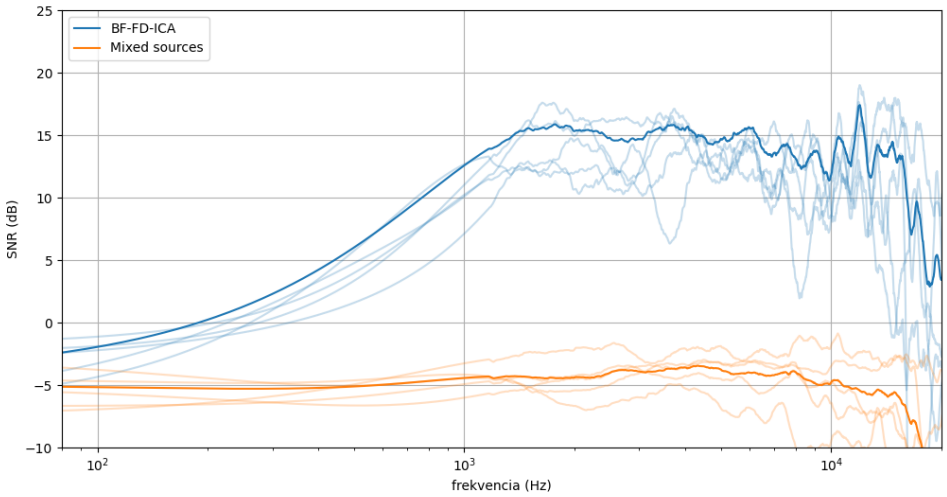
BF-FD-ICA teljesítményének értékelése

9.1. Kiértékeléshez használt általános paraméterek

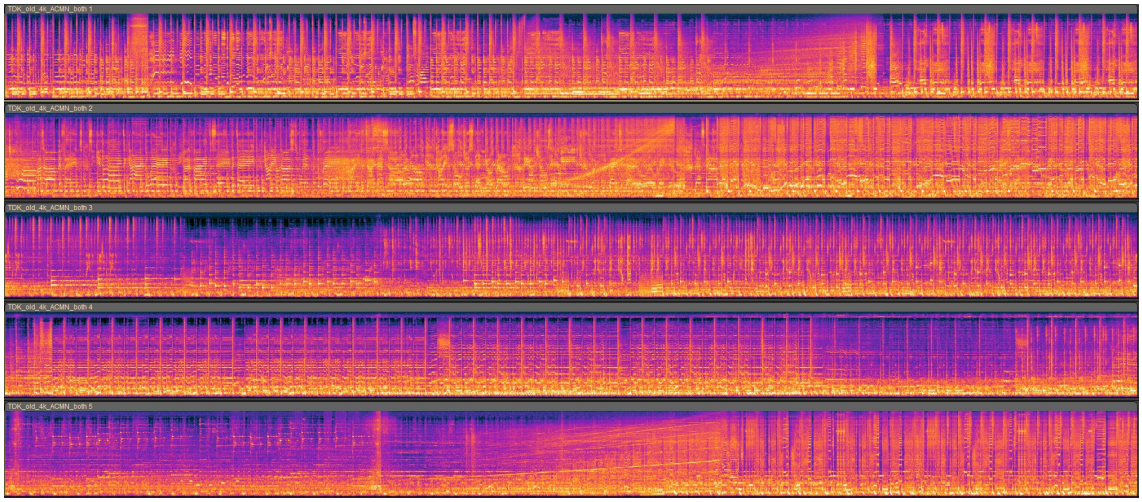
A kiértékeléshez, ahol ezt külön nem jelöltük, minden esetben az összes mikrofont használtuk a reflexiómentes elrendezésben a BF-FD-ICA bemeneteként, a szeparálandó források számát pedig "kézzel" 5-re állítottuk. Az algoritmus STFT ablakhosszát $F = 4096$-ra állítottuk, valamint az általában jobban teljesítő ACMN algoritmust használtuk komplex ICA-ként. A szeparációs teljesítmény értékelésére a 8.2 alfejezetben bevezetett relatív hibát mérő SNR metrikát használtuk. A frekvenciafüggő SNR-t Savitzky-Golay szűrővel [26] történő simító szűrés után decibel skálán ábrázoltuk halvány vonallal az 5 forrás esetére külön-külön, illetve "normál" vonallal az 5 forrásra számított átlagos értéket.

9.2. BF-FD-ICA teljesítményének általános elemzése

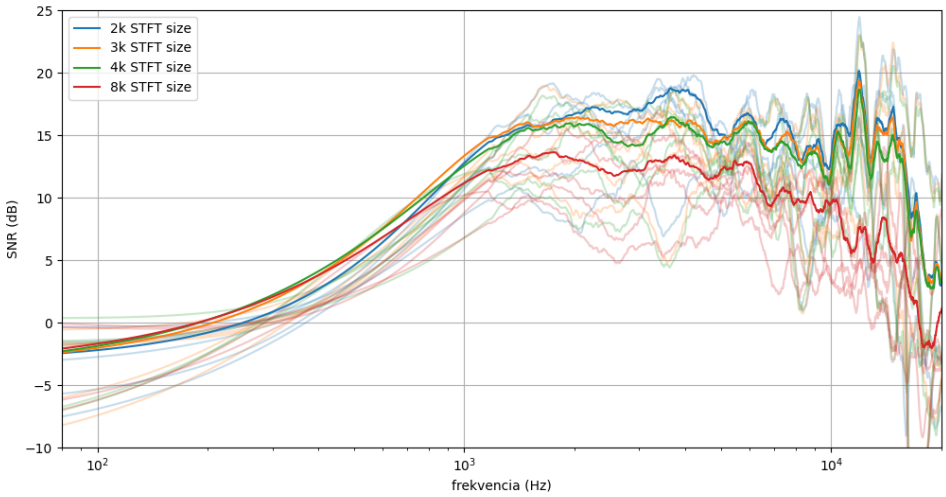
Az 5 forrásra a 9.1 alfejezetben leírt paraméterekkel futtatva, majd kiértékelve az algoritmus teljesítményét, valamint a kevert források közül 5-öt kiválasztva és arra SNR-t számítva a 9.1 ábrán látható SNR görbét kapjuk. Ez alapján azt mondhatjuk, hogy a módszer igen átlagosan igen jónak mondható 15 dB relatív hibájú szeparációt valósít meg széles, 1000 Hz - 15000 Hz-es sávban. Ez nagyjából 20 dB-es javulást jelent jel-zaj viszonyban a szeparálatlan esethez képest, hiszen ott az SNR nagyjából -4 dB és -6 dB közötti. Megjegyzendő, hogy mivel a források RMS teljesítménye normalizálásra került a mérés során, valamint egy adott forrás jelét tekintve a többi 4 forrás jele zajként jelenik meg, így elméletileg -6 dB körüli SNR-t várhatunk a szeparálatlan esetben, ezt a számított értékek jól közelítik. Alacsonyabb frekvenciákon a teljesítmény romlik, melynek oka az lehet, hogy a kisebb frekvenciákon a nagyobb hullámhossz miatti kis relatív fáziskülönbség hangfalak kevésbé irányított lesugárzása miatt a frekvenciatartományban közel szinguláris keverőmátrixok jönnek létre (kis frekvenciákon a mikrofonok közel ugyanazt hallják), így itt nyilván a nehezebb feladat a hatékony szeparáció. Megállapítható azonban, hogy a szeparációs teljesítmény kisebb frekvenciákon sem esik be nagyon hirtelen, még 500 Hz környékén is 6 dB környéki a relatív hiba, azaz 12 dB körüli a szeparációs teljesítmény. A szeparált jelek spektogramjait a 9.2 ábra mutatja, ez alapján megbizonyosodhatunk hogy a szeparáció valóban jól működött. (Látható, hogy a szeparált spektogramok nagymértékben hasonlítanak a 8.5 ábrán látható referencia spektogramokra.)



9.1. ábra. A BF-FD-ICA szeparációs teljesítménye



9.2. ábra. A BF-FD-ICA által szeparált jelek spektogramja



9.3. ábra. A szeparáció teljesítménye az STFT ablakhossz függvényében

9.3. STFT ablakhossz hatása a szeparációs teljesítményre

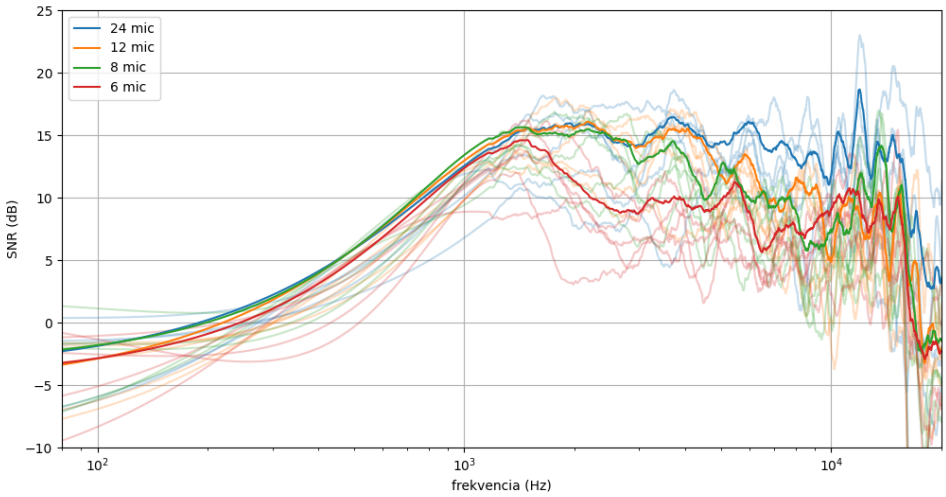
A szeparációs teljesítmény STFT ablakhossztól való függését a 9.3 ábra mutatja. Az ábrán azt láthatjuk, hogy az ablakhossz csökkentésével némileg javul a szeparációs teljesítmény. Ez elsőre nem tűnik logikusnak a 3.3.2 alfejezetben leírt nagyobb STFT ablakokra a normál konvolúció cirkuláris konvolúcióval történő jobb közelíthetősége miatt. A látott eredményeknek itt az lehet az oka, hogy az igen gyengén reverbáns környezet miatt a mikrofonok és a források közti impulzusválaszok igen rövidek így rövid STFT ablak is elég, viszont a hosszabb STFT ablak azt jelenti, hogy egy frekvenciabinbe kevesebb minta fog esni, így a komplex ICA algoritmusnak is kevesebb adatból kell becsülni a független komponeseket. Megjegyzendő, hogy reverbánsabb környezetben valószínűleg a nagyobb STFT ablakokra jobb teljesítményt láthatnánk.

9.4. Mikrofonok számának hatása a BF-FD-ICA teljesítményére

A szeparációs teljesítmény használt mikrofonok számától való függését a 9.4 ábra mutatja. A mikrofonok számának csökkentését úgy valósítottuk meg, hogy a 24 csatornás mikrofontömbből rendre csak minden második (12 mikrofon esete), harmadik (8 mikrofon esete), illetve negyedik (6 mikrofon esete), mikrofon jelét vettük figyelembe a BF-FD-ICA futtatásánál. A csökkentett mikrofonszámokkal futtatott algoritmus teljesítménye főképp a magasabb frekvenciákon romlik. Ennek egyik oka az lehet, hogy a kevesebb mikrofonnal járó nagyobb mikrofontáv miatt a beamfromer alapú permutációs algoritmus kevésbé jól működik. A másik oka az hogy a 6.1 alfejezetben bemutatott dimenzióredukciós módszer ténylegesen működni látszik.

9.5. Komplex ICA algoritmus megválasztásának hatása

A szeparációs teljesítménynek a használt komplex ICA algoritmustól való függését a 9.5 ábra mutatja. Azt tapasztalhatjuk, hogy az ACMN algoritmus, bár működési elvét valamint



9.4. ábra. A szeparáció teljesítménye az mikrofonszám függvényében

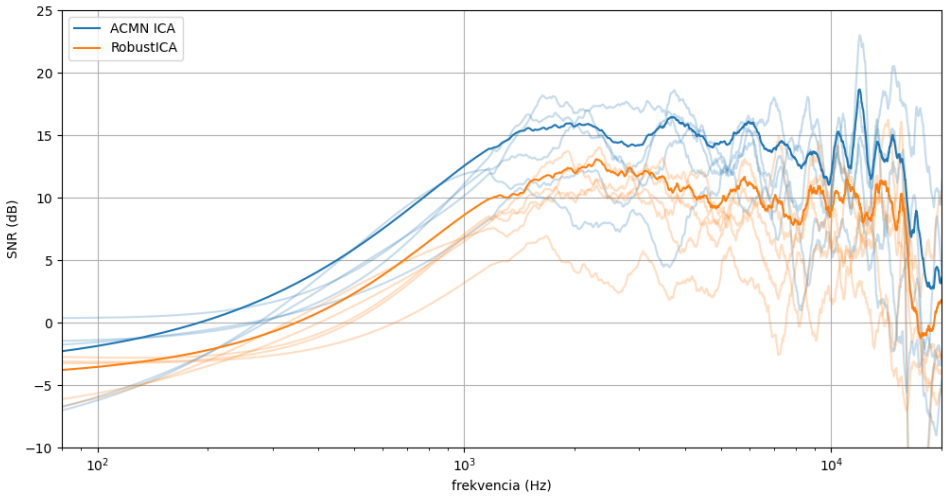
számításigényét tekintve is komplexebb, lényegében a teljes frekvenciasávban $4-5$ dB-el jobb szeparációs teljesítményt nyújt mint a RobustICA. Megjegyzendő azonban, hogy az ábrán látható teljesítményt a RobustICA $3-4$-szer rövidebb számítási idővel nyújtotta, így sebességkritikus esetekben megfontolandó lehet a használata.

9.6. Permutáció korrekciós módszerek teljesítményének összehasonlítása

A szeparációs teljesítménynek a használt permutáció korrekciós algoritmustól való függését a 9.6 ábra mutatja. Látható, hogy a permutációs probléma megoldása nélkül minimálisan, vagy egyáltalán nem szeparál az algoritmus így a permutáció korrekció kritikus fontosságú. A permutációs algoritmusokat összevetve azt tapasztaljuk, hogy a beamforming alapú permutáció korrekciós módszer kisebb frekvenciákon igen jól működik, azonban magasabb frekvenciákon, a térbeli mintavételi tétel megsértéséből adódó, az iránykarakterisztikán megjelenő melléknyalábok miatt az algoritmus teljesítménye leromlik. Az RLRJ módszer szélesebb frekvenciatartományban működik azonban az átlagos teljesítménye rosszabb mint a beamformer alapú megoldásé. Két módszer kombinálása, az elvártaknak megfelelően igen jó szeparációs teljesítményhez vezet, amely lényegében minden frekvenciasávban jobb eredményt ad mint az egy módszer használata. A permutációs módszerek vizsgálatához érdemes az lehet a frekvenciafüggő iránykarakterisztikákat is elemezni, ezeket a 9.7, 9.8, 9.9, 9.10 ábrák mutatják.

9.7. BF-FD-ICA és az irodalmi FD-ICA teljesítményének összehasonlítása

Az 9.11 ábra a 24 mikrofonos, dimenzióredukciót, illetve RLRJ és Beamforming alapú permutáció korrekciót alkalmazó BF-FD-ICA, illetve a az 5 mikrofonos RLRJ permutációs módszert használó módszer szeparációs teljesítményét hasonlítja össze. Az FD-ICA-t a mikrofontömbből 5 mikrofon jelét kiválasztva futtattuk. Látható hogy a BF-FD-ICA a



9.5. ábra. A szeparáció teljesítménye a komplex ICA algoritmus függvényében

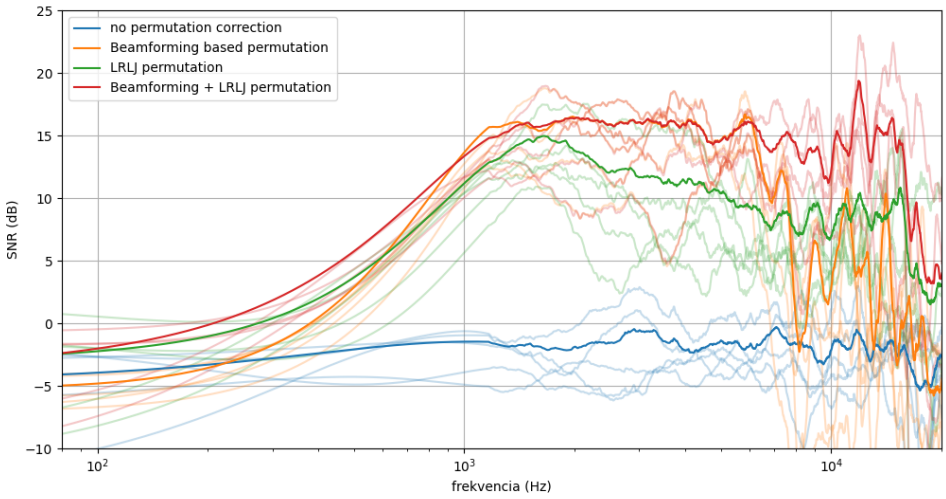
teljes frekvenciatartományban jobban teljesít 1500 Hz frekvencia felett nagyjából $6-8$ dB-el, így a 6. fejezetben bevezetett módosítások valóban működnek erre az esetre.

9.8. BF-FD-ICA és klasszikus nyalábformáló algoritmusok teljesítményének vizsgálata

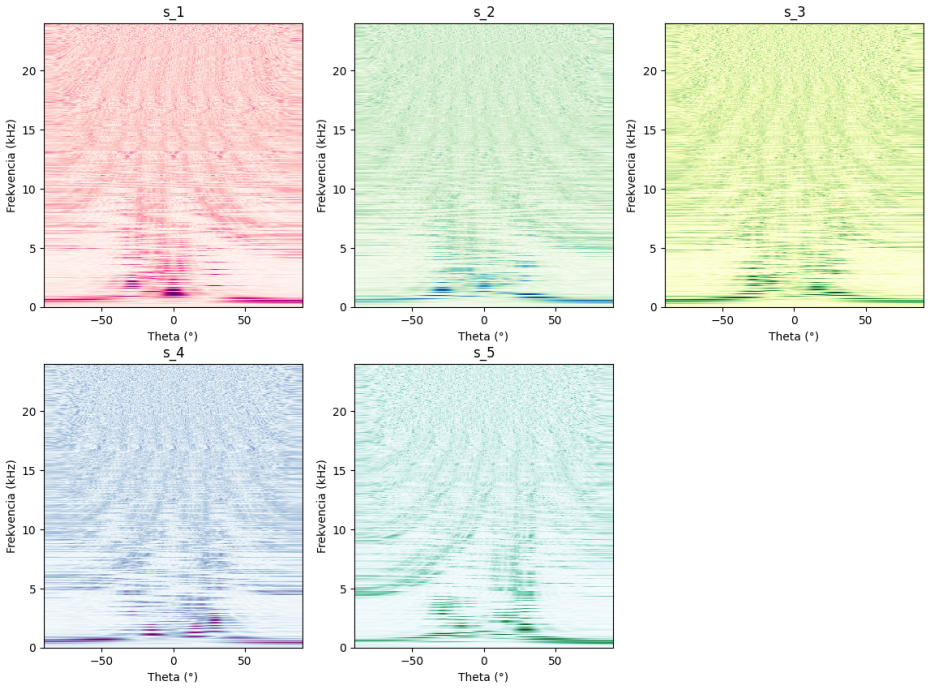
Azt is megvizsgáltuk, hogyan viszonyul a BF-FD-ICA teljesítménye néhány klasszikus, mikrofontömbös beamformer algoritmus teljesítményéhez. Az összehasonlításhoz a jól ismert Delay and Sum (DAS) beamformer algoritmust [20], illetve a fejlettebb Frost beamformert [6] használtuk, mindkettőt a MATLAB Phased array toolbox féle implementációban. A klasszikus beamformerek főnyaláb irányaként a mérési elrendezésből következő tényleges DOA-kat adtuk meg, valamint bemenetként ugyanazt a 24 mikrofonos felvételt, melyen a BF-FD-ICA-t is futtattuk. A beamformerek, illetve a BF-FD-ICA szeparációs teljesítményének összevetését a 9.12 ábra mutatja. Látható hogy az ICA alapú megoldásunk nagyjából 2500 Hz-ig megközelítőleg $8-10$ dB-el felülmúlja a DAS beamformer teljesítményét, magasabb frekvenciákon pedig a DAS lényegében egyáltalán nem szeparál így ebben az esetben a kijelenthető, hogy az ICA alapú megoldásunk szeparációs teljesítménye egyértelműen felülmúlja a DAS beamformerét. A Frost beamformer gyenge teljesítménye igen meglepő, attól ugyanis azt várnánk, hogy a DAS beamformernél jobb legyen a szeparációs teljesítménye. Megjegyzendő, hogy végtelen nyílt térre szimulált jeleken ugyanerre az elrendezésre kipróbálva jól működött a Frost algoritmus, így ebben az esetben a gyenge szeparációs teljesítmény okaként a valós környezetben fellépő minimális reflexiókra gyanakodhatunk.

9.9. Reflektív környezet hatásának vizsgálata

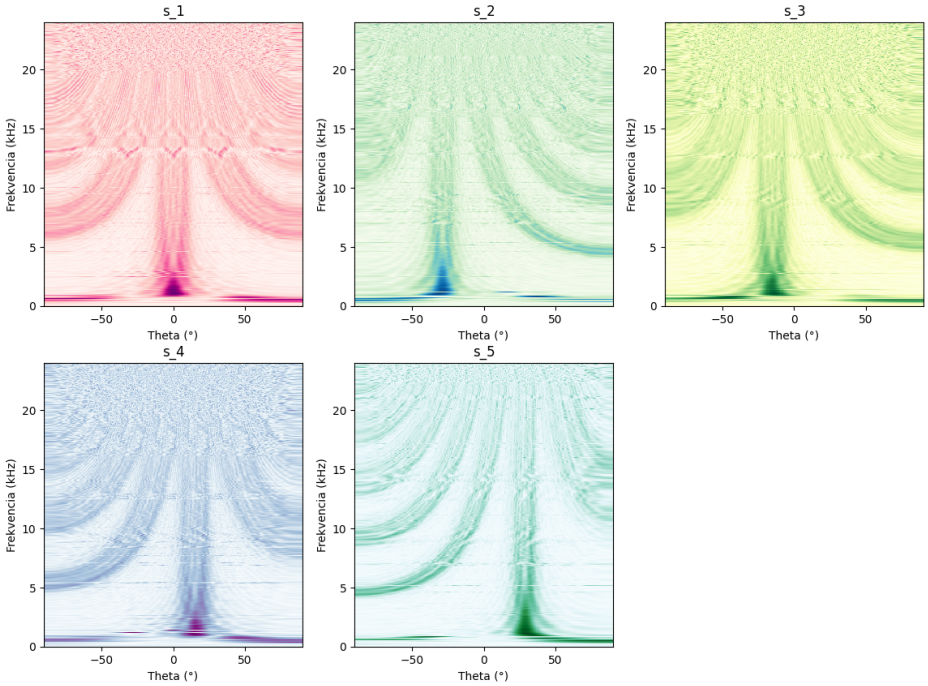
Végül a BF-FD-ICA teljesítményét a 8.1 alfejezetben ismertetett reflektív elrendezésben is megvizsgáltuk. Komplex ICA algoritmusként itt is az ACMN-t használtuk, valamint a szeparációs teljesítményt itt is többféle STFT ablakhossz esetére értékeltük ki. Ez azért volt indokolt, mert a reflektív környezetben létrejövő hosszabb impulzusválaszok miatt



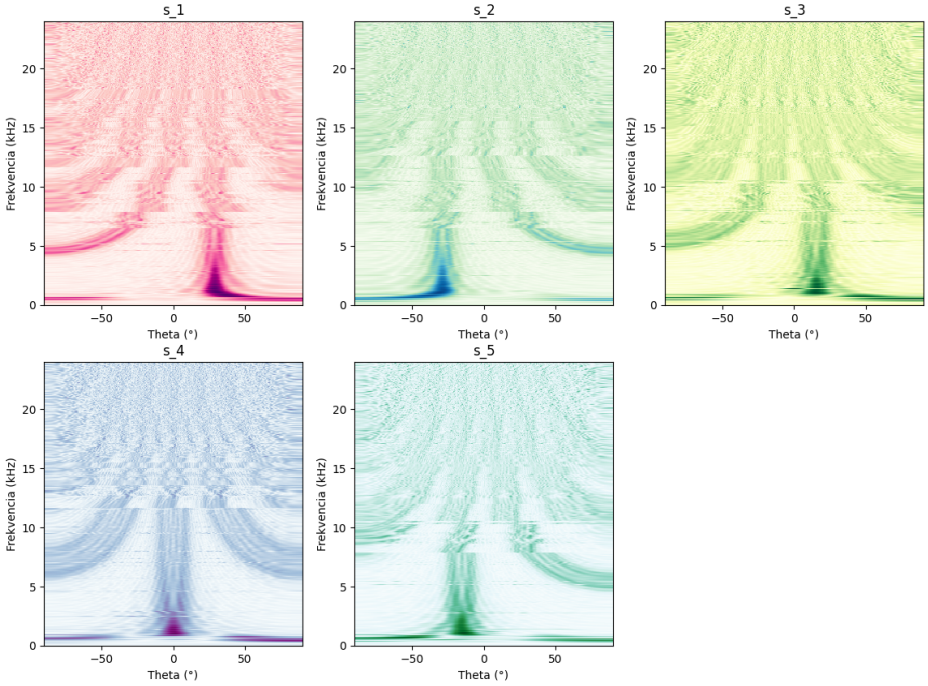
9.6. ábra. A szeparáció teljesítménye a használt permutáció korrekciós algoritmus függvényében



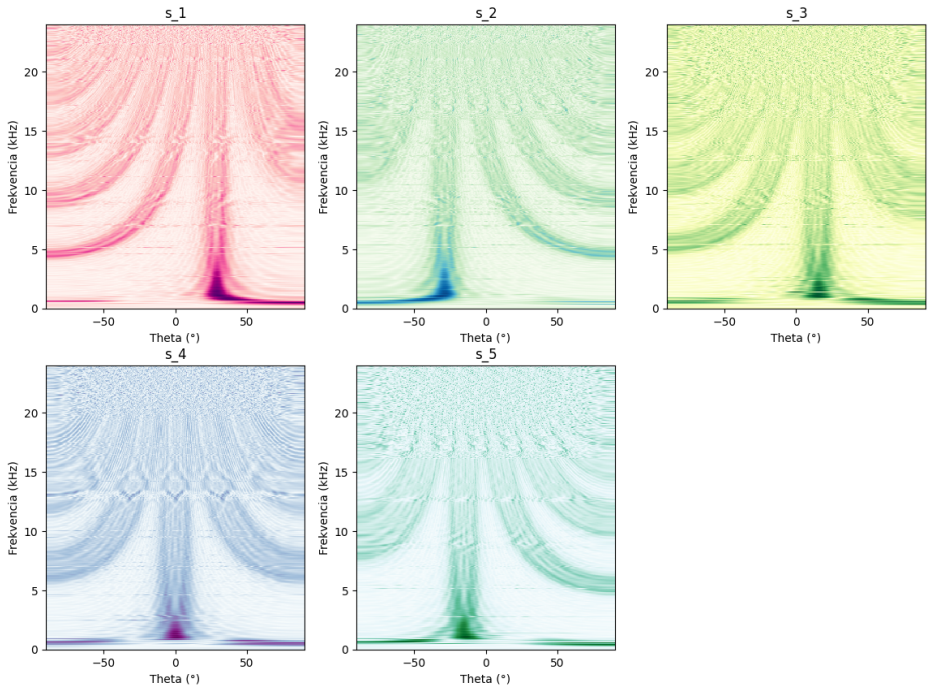
9.7. ábra. Frekvenciafüggő iránykarakterisztikák permutáció korrekció nélkül



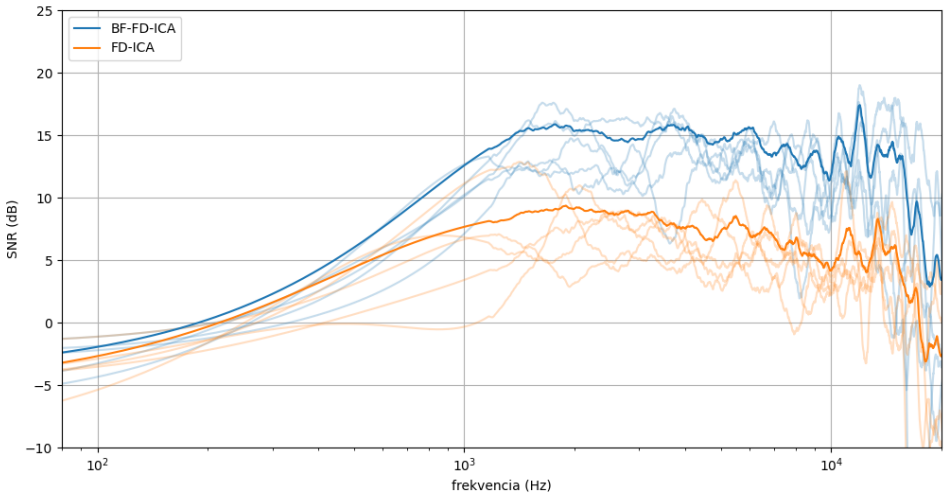
9.8. ábra. Frekvenciafüggő iránykarakterisztikák RLRJ használatával



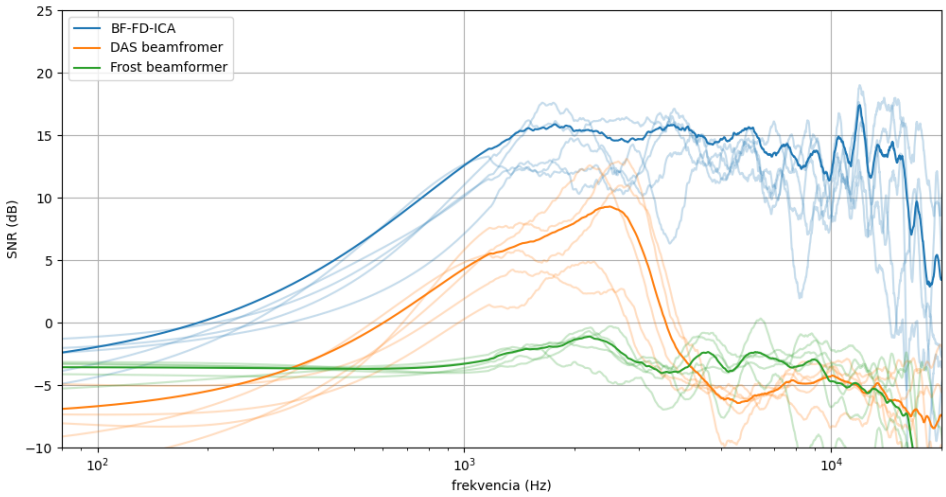
9.9. ábra. Frekvenciafüggő iránykarakterisztikák beamforming alapú permutációs módszer használatával



9.10. ábra. Frekvenciafüggő iránykarakterisztikák RLRJ és beamformer permutációs módszer használatával

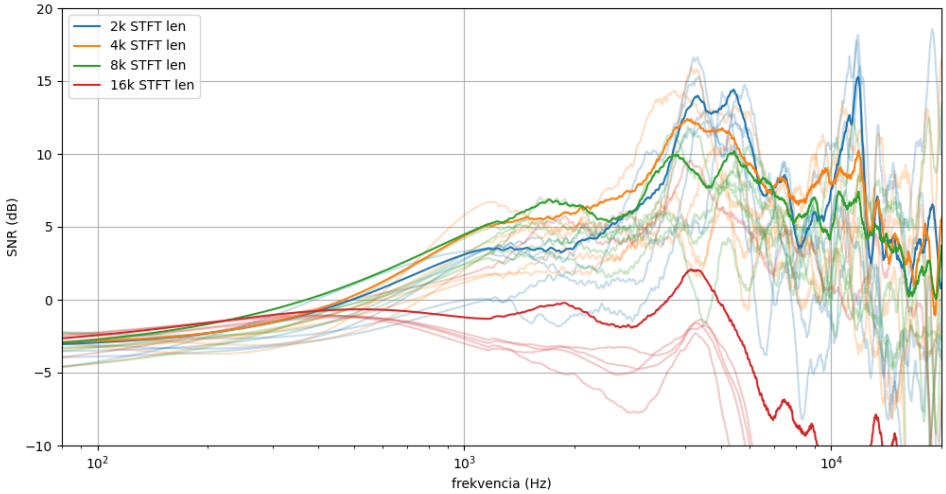


9.11. ábra. BF-FD-ICA és klasszikus FD-ICA teljesítményének összehasonlítása

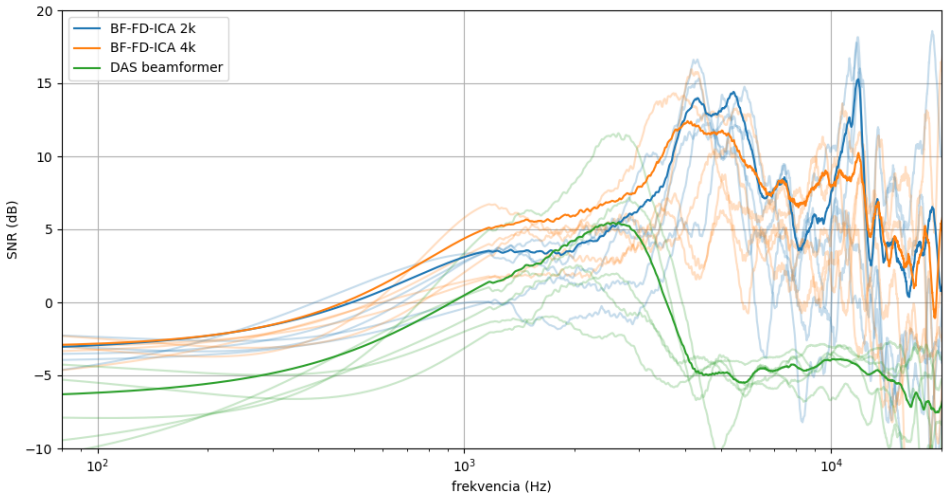


9.12. ábra. A DAS, a Frost, és a BF-FD-ICA algoritmus szeparációs teljesítményének összevetése

a rövidebb STFT ablak nem biztosan jelent jobb szeparációs teljesítményt. A reflektív esetre számított szeparációs teljesítményt az STFT ablakhossz függvényében a 9.13 ábra szemlélteti. Amint arra számíthattunk, a szeparációs teljesítmény a reflektív esetben jelentősen (megközelítőleg $5-10$ dB-el) gyengébb mint a reflexiómentes esetben. Ennek oka a jelentősen komplexebb impulzusválasz a reflexiómentes esethez képest (reflekciómentes esetben az impulzusválasz lényegében egy egyszerű időkésleltetés, reflektív esetben viszont megjelennek a többutas terjedés hatásai). Az is megfigyelhető, hogy a rövidebb STFT hossz valóban nem javítja egyértelműen a szeparációs teljesítményt ebben az esetben, viszont a túlzottan hosszú ablakhossz választás sem célszerű (a 16000-es STFT ablak esetére valószínűleg az okozza a gyengébb teljesítményt, hogy ekkor egy frekvenciabinre túlságosan kevés minta esik a komplex ICA algoritmus hatékony működéséhez, valamint a több frekvenciabin miatt a permutáció optimalizációs módszernek is nehezebb dolga van). Mindezek mellett ebben az esetben is elmondható, hogy a BF-FD-ICA 1 kHz fölött megközelítőleg $10-15$ dB-es szeparációt valósít meg, ez pedig még mindig igen jónak mondható. A klasszikus beamformer módszerekkel összehasonlítva (9.14 ábra) a reflektív esetben azt tapasztalhatjuk, hogy a FROST gyakorlatilag egyáltalán nem szeparál (így ezt az ábrán fel sem tüntettük), valamint a DAS beamformer teljesítménye is romlik a reflekciómentes esethez képest ($\approx$ 5 dB-el), így a BF-FD-ICA ebben az esetben is jelentősen jobb szeparációs teljesítménnyel rendelkezik, mint a hagyományos beamformer algoritmusok.



9.13. ábra. A szeparáció teljesítménye az STFT ablakhossz függvényében a reflektív elrendezésben



9.14. ábra. A DAS és a BF-FD-ICA algoritmus szeparációs teljesítményének összevetése reflektív esetben

10. fejezet

Összefoglalás

A dolgozatban áttekintettük az ICA, valamint az FD-ICA, illetve az ehhez szükséges komplex ICA szakirodalmát. Megvizsgáltuk a naiv FD-ICA problémáinak a szakirodalomban fellelhető megoldási lehetőségeit, majd bemutattuk, hogy ezek alapján hogyan lehet egy gyakorlatban is működő FD-ICA architektúrát megvalósítani. Ezután áttekintettük a beamforming probléma milyen extra információkkal szolgál a klasszikus ICA problémához képest (a források számánál jelentősen több mikrofon, illetve ismert mikrofonpozíciók), majd ezek kihasználására újszerű módszereket vezettünk be (dimenzióredukciós FD-ICA, illetve beamforming alapú permutációs módszer), ez alapján pedig megalkottuk, a mikrofontömbös környezetre optimalizált BF-FD-ICA architektúrát. A megoldásunk teljesítményét valós, 5 forrásos, 24 mikrofonos mikrofontömbös mérésre értékeltük ki, komplex elektronikus zenei hangmintákat használva. A kiértékeléshez bevezettük a valós mérésekre is használható idő-frekvenciatartományi hibán alapuló jel-zaj viszony számítási módszert.

A módszer szeparációs teljesítményének értékelése során megállapítottuk, hogy a használt valós mérési elrendezésben, széles közel 1500 Hz - 15000 Hz-es frekvenciasávban, hozzávetőleg 20 decibeles szeparációs teljesítményt értünk el, valamint a szeparációs teljesítmény kisebb frekvenciákra sem csökken drasztikus módon, még 500 Hz-en is 10 decibel környéki szeparációs teljesítményről beszélhetünk. A klasszikus FD-ICA algoritmushoz hasonlítva a módszerünket, azt tapasztaltuk, hogy a szeparációs teljesítmény nagyjából 1000 Hz frekvencia felett jelentősen javul az általunk javasolt BF-FD-ICA használatával. A módszerünk teljesítményét természetesen hagyományos beamforming algoritmusokkal (delay and sum, illetve Frost) is összevetettük. Itt is azt tapasztaltuk, hogy a BF-FD-ICA módszer szeparációs teljesítménye jelentősen felülmúlja az erősen limitált frekvenciasávban hatékony DAS illetve az erre az esetre szinte egyáltalán nem működő Frost beamformer teljesítményét. Végezetül a BF-FD-ICA teljesítményét reflektív környezetben is megvizsgáltuk. Ekkor azt tapasztaltuk, hogy a módszer szeparációs teljesítménye, bár a reflexiómentes esethez képest számottevően romlott, ekkor is felülmúlja a klasszikus beamformer algoritmusok teljesítményét.

Összességében elmondhatjuk tehát, hogy a kutatás célját, azaz egy ICA alapú, valóságban is működő beamformer algoritmus megvalósítását teljes mértékben sikerült elérnünk, így a bevezetőben feltett kérdésre, azaz hogy érdemes-e a beamforming problémát függetlenkomponens-analízis alapon megközelíteni, egyértelműen igennel válaszolhatunk.

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani a konzulensemnek, Dr. Fiala Péternek, aki ezen diplomaterv elkészültéhez nagyban hozzájárult a témajavaslattal, a kutatás során adott tanácsaival, illetve a dolgozat megírásában nyújtott segítségével.

Szeretném megköszönni emellett az Akusztikai és Stúdiótechnikai Laboratórium munkatársainak a mérésekhez nyújtott helyszínt és eszközöket, melyek nélkül ezen diplomaterv ebben a formában nem készülhetett volna el.

Ábrák jegyzéke

2.1. Az instant keverési modell 3

3.1. A konvolutív keverési modell 8
3.2. A frekvenciatartományi ICA sematikus ábrája 11

6.1. Az FD-ICA blokkvázlata 23
6.2. FD-ICA iránykarakterisztikák abszolút értéke 27

7.1. BF-FD-ICA blokkvázlata 29

8.1. A használt mérési elrendezés 32
8.2. Mérési elrendezés a valóságban 32
8.3. Reflektív mérési elrendezés a gyakorlatban 33
8.4. Kevert jelek spektogramjai 33
8.5. Referenciajelek spektogramjai 34

9.1. A BF-FD-ICA szeparációs teljesítménye 37
9.2. A BF-FD-ICA által szeparált jelek spektogramja 37
9.3. A szeparáció teljesítménye az STFT ablakhossz függvényében 38
9.4. A szeparáció teljesítménye az mikrofonszám függvényében 39
9.5. A szeparáció teljesítménye a komplex ICA algoritmus függvényében 40
9.6. A szeparáció teljesítménye a használt permutáció korrekciós algoritmus függvényében 41
9.7. Frekvenciafüggő iránykarakterisztikák permutáció korrekció nélkül 41
9.8. Frekvenciafüggő iránykarakterisztikák RLRJ használatával 42
9.9. Frekvenciafüggő iránykarakterisztikák beamforming alapú permutációs módszer használatával 42
9.10. Frekvenciafüggő iránykarakterisztikák RLRJ és beamformer permutációs módszer használatával 43
9.11. BF-FD-ICA és klasszikus FD-ICA teljesítményének összehasonlítása 43
9.12. A DAS, a Frost, és a BF-FD-ICA algoritmus szeparációs teljesítményének összevetése 44
9.13. A szeparáció teljesítménye az STFT ablakhossz függvényében a reflektív elrendezésben 45
9.14. A DAS és a BF-FD-ICA algoritmus szeparációs teljesítményének összevetése reflektív esetben 45

Irodalomjegyzék

[1] Ella Bingham – Aapo Hyvärinen: A fast fixed-point algorithm for independent component analysis of complex valued signals. *International journal of neural systems*, 10. évf. (2000. 03), 1–8. p.

[2] J-F Cardoso: Source separation using higher order moments. In *International Conference on Acoustics, Speech, and Signal Processing,* (konferenciaanyag). 1989, IEEE, 2109–2112. p.

[3] Jean-François Cardoso – Antoine Souloumiac: Blind beamforming for non-gaussian signals. In *IEE proceedings F (radar and signal processing)* (konferenciaanyag), 6. köt. 1993, IET, 362–370. p.

[4] Pierre Comon: Independent component analysis, a new concept? *Signal Processing*, 36. évf. (1994) 3. sz., 287–314. p. ISSN 0165-1684. URL `https://www.sciencedirect.com/science/article/pii/0165168494900299`. Higher Order Statistics.

[5] Jan Eriksson – Visa Koivunen: Complex-valued ica using second order statistics. In *Proceedings of the 2004 14th IEEE Signal Processing Society Workshop Machine Learning for Signal Processing, 2004.* (konferenciaanyag). 2004, IEEE, 183–192. p.

[6] Otis Lamont Frost: An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60. évf. (1972) 8. sz., 926–935. p.

[7] Monson H Hayes: *Statistical digital signal processing and modeling*. 1996, John Wiley & Sons.

[8] Simon Haykin: *Adaptive Filter Theory (3rd Ed.)*. USA, 1996, Prentice-Hall, Inc. ISBN 013322760X.

[9] A. Hyvarinen: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10. évf. (1999) 3. sz., 626–634. p.

[10] Aapo Hyvärinen – Juha Karhunen – Erkki Oja: *Independent Component Analysis*. 26. köt. 2001. 06. ISBN 9780471405405.

[11] Don H Johnson: Array signal processing. *concepts and techniques*, 1993.

[12] Walter Kramer: Probability & Measure : Patrick Billingsley (1995): (3rd ed.). New York : Wiley, ISBN 0-471-0071-02, pp 593, [pound sign] 49.95. *Computational Statistics & Data Analysis*, 20. évf. (1995. December) 6. sz., 702–703. p.
URL `https://ideas.repec.org/a/eee/csdana/v20y1995i6p703-702.html`.

[13] Harold W Kuhn: The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2. évf. (1955) 1-2. sz., 83–97. p.

[14] Russell Howard Lambert: *Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures*. 1996, University of Southern California.

[15] Te-Won Lee – Anthony Bell – Russell Lambert: Blind separation of delayed and convolved sources. *Advances in neural information processing systems*, 9. évf. (1996).

[16] Hualiang Li – Tulay Adali: Gradient and fixed-point complex ica algorithms based on kurtosis maximization. In *2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing* (konferenciaanyag). 2006, IEEE, 85–90. p.

[17] G Lunetta: Di una generalizzazione dello schema della curva normale. *Annali della Facolta di Economia e Commercio di Palermo*, 17. évf. (1963) 2. sz., 237–244. p.

[18] Dimitrios Mallis – Thomas Sgouros – Nikolaos Mitianoudis: Convolutive audio source separation using robust ica and an intelligent evolving permutation ambiguity solution. *Evolving Systems*, 9. évf. (2018), 315–329. p.

[19] Nikolaos Mitianoudis: *Audio Source Separation using Independent Component Analysis*. PhD értekezés (University of London). 2004.

[20] Ronald Mucci: A comparison of efficient beamforming algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32. évf. (1984) 3. sz., 548–558. p.

[21] Michael Novey – Tulay Adali: Complex ica by negentropy maximization. *IEEE Transactions on Neural Networks*, 19. évf. (2008) 4. sz., 596–609. p.

[22] Mike Novey – Tulay Adali: Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm. In *2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing* (konferenciaanyag). 2006, IEEE, 79–84. p.

[23] Lucas Parra – Clay Spence: Convolutive blind separation of non-stationary sources. *IEEE transactions on Speech and Audio Processing*, 8. évf. (2000) 3. sz., 320–327. p.

[24] Bernard Picinbono: On circularity. *IEEE Transactions on signal processing*, 42. évf. (1994) 12. sz., 3473–3482. p.

[25] Hiroshi Saruwatari – Satoshi Kurita – Kazuya Takeda – Fumitada Itakura – Tsuyoki Nishikawa – Kiyohiro Shikano: Blind source separation combining independent component analysis and beamforming. *EURASIP Journal on Advances in Signal Processing*, 2003. évf. (2003), 1–12. p.

[26] Abraham Savitzky – Marcel JE Golay: Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36. évf. (1964) 8. sz., 1627–1639. p.

[27] Paris Smaragdis: Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22. évf. (1998) 1-3. sz., 21–34. p.

[28] A Van Den Bos: Complex gradient and hessian. *IEE Proceedings-Vision, Image and Signal Processing*, 141. évf. (1994) 6. sz., 380–382. p.

[29] Vicente Zarzoso – Pierre Comon: Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *IEEE Transactions on neural networks*, 21. évf. (2009) 2. sz., 248–261. p.