



M Ű E G Y E T E M 1 7 8 2

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

Csóka Bence

**HANGFORRÁSOK
LOKALIZÁCIÓJA
MIKROFONRENDSZEREKKEL**

KONZULENS

Dr. Fiala Péter

BUDAPEST, 2021



M Ű E G Y E T E M 1 7 8 2

DIPLOMATERV FELADAT

Csóka Bence

szigorló villamosmérnök hallgató részére

Forráskövetés akusztikai nyalábformálással

Az akusztikai nyalábformálás olyan akusztikai forráslokalizációs eljárás, mely egy mikrofontömb jeleinek alkalmas erősítésével és késleltetésével képes meghatározni a beérkező hang irányát vagy a források pozícióit. A forráslokalizáció pontosságát erősen befolyásolja a mikrofontömb mérete és térbeli felbontása, az analízis frekvenciatartománya és természetesen az akusztikai tér jel-zaj-viszonya. Így, mozgó források esetén különösen fontos olyan forráskövető algoritmusok alkalmazása, melyek a nyalábformálás aktuális eredményét a korábbi pozíciókra alapozott mozgásmodellel, arra épülő forráskövetéssel finomítják.

A hallgató feladata a forráskövető algoritmus megvalósítása. A megoldásnak az alábbiakra kell kiterjednie:

1. Tekintse át az akusztikai fókuszálás irodalmát, mutassa be a fókuszálás alapelveit.
2. Ismertesse az akusztikai nyalábformálás legfontosabb algoritmusait (CBF, MUSIC), kitérve azok hibájára, és a belőlük származtatott lokalizáció szórására.
3. Ismerkedjen meg olyan Kalman-szűrő algoritmussal, mely a forráslokalizáció aktuális pozíció- és szóráseredménye alapján a pályamodellel becslésére alkalmas.
4. Implementálja a forráskövető algoritmust.
5. Drónokon végzett valós mérések segítségével értékelje az algoritmus pontosságát és megbízhatóságát.

Tanszéki konzulens: Dr. Fiala Péter, BME Hálózati Rendszerek és Szolgáltatások Tanszék
Ipari konzulens:

Budapest, 2020. 03. 15.

Dr. Imre Sándor

egyetemi tanár
tanszékvezető

Konzulensi vélemények:

Tanszéki konzulens: | Beadható, | Nem beadható, dátum:

aláírás:

Külső konzulens:

aláírás:

Tartalomjegyzék

Kivonat.....	4
Abstract.....	5
1 Bevezetés	6
1.1 Motiváció, cél	6
1.2 Korábbi kutatások a témában.....	6
2 Nyalábformáló algoritmusok	9
2.1 A nyalábformálás alapjai	9
2.1.1 Fókuszálás és forráslokalizáció	9
2.1.2 A mikrofonok és a vászon elhelyezkedése	10
2.1.3 A vizsgálati frekvencia	12
2.2 Conventional Beamforming.....	13
2.3 Multiple Signal Classification	14
2.4 Compressive Beamforming	14
2.4.1 Compressive Beamforming zajmentes esetben	15
2.4.2 Compressive Beamforming zajos esetben	15
2.4.3 Compressive Beamforming pontszórásfüggvényekkel	17
2.5 Távolságbecslés az akusztikus vászon kiterjesztésével	17
3 Mozgó forrás követése Kálmán-szűrővel.....	21
3.1 Optimális lineáris állapotbecslés	21
3.2 Unscented Kalman Filter	23
3.3 Egynél több hangforrás követése	24
4 Szimulációk.....	28
4.1 A szimulációs keretrendszer	28
4.2 Szimulációk két dimenzióban	30
4.2.1 Conventional Beamforming.....	31
4.2.2 Multiple Signal Classification	32
4.2.3 Compressive Beamforming	33
4.2.4 Mozgó hangforrás követése Kálmán szűrővel.....	36
4.3 Szimulációk három dimenzióban.....	37
4.3.1 Conventional Beamforming.....	38
4.3.2 Multiple Signal Classification	39

4.3.3 Compressive Beamforming	40
4.3.4 Mozgó hangforrás követése Kálmán-szűrővel	42
4.3.5 Két hangforrás irány- és távolságbecslése	43
5 Mérések, eredmények.....	47
5.1 Szabadtéri mérések	47
5.1.1 A Kálmán-szűrő paraméterfüggő működése	48
5.1.2 Secopx8.....	52
5.1.3 Tarot680.....	53
5.1.4 Drónok irány- és távolságbecslése.....	55
5.2 Továbbfejlesztési lehetőségek	57
6 Összefoglalás.....	58
Köszönetnyilvánítás	60
Irodalomjegyzék.....	61

Kivonat

Jelen dolgozat témája hangforrások pozíciójának vagy pályájának automatizált meghatározása mikrofonrendszerekkel történő akusztikai mérésekkel. A mikrofonrendszer által a felvett hangjelek megfelelő erősítésével és késleltetésével egy adott irányba fókuszálhatunk különböző nyalábformáló algoritmusok segítségével.

A dolgozat három nyalábformáló algoritmus részletezésével foglalkozik, ezek közül kettő a szakirodalom által részletesen tárgyalt. A CBF (Conventional Beamforming) módszer a három közül a legegyszerűbb, ugyanakkor a legkevésbé pontos is. A MUSIC (Multiple Signal Classification) algoritmus a vett jelek kereszt spektrum-mátrixának a jel- és zajterekre való felbontásán alapul.

Újdonságként bemutatásra kerül a CS (Compressive Sensing) elven alapuló Compressive Beamforming algoritmus. A Compressive Beamforming a forráslokálizálást konvex optimalizálási problémára vezeti vissza, és egy ritkasági kényszerfeltételnek köszönhetően pontosabb eredményt ad, mint a lineáris algebrai elven működő algoritmusok.

Tárgyalásra kerül a mozgó hangforrások követésének problémája is, ami például Kálmán-szűrő segítségével oldható meg. A Kálmán-szűrő hagyományos változata optimális állapotbecslést ad lineáris, mozgó vagy változó állapotú rendszerekre, illetve kiterjeszhető, hogy nemlineáris rendszerek esetében is működjön.

A különböző nyalábformáló algoritmusokat MATLAB környezetben elvégzett szimulációk segítségével zajmentes és zajos esetekben is összehasonlítjuk. Méréseket mutatunk be, melyekkel a gyakorlatban is értékelhetjük a nyalábformáló algoritmusok működését, ezzel is igazolva használhatóságukat különböző hangforrások (például drónok) lokalizációjához. Mindezek mellett a Kálmán-szűrő működését is megvizsgáljuk a mérési eredményeken, hogy milyen előnyöket és többletinformációt nyújt a nyalábformáló algoritmusokhoz képest, és hogy a különböző paraméterei hogyan befolyásolják a működését.

Abstract

The topic of this thesis is the automatized determination of the location and trajectory of sound sources by acoustic measurements with microphone systems. It is possible to focus on a specific direction with the microphone system by amplifying and delaying the received signals in a suitable manner, with the help of beamforming algorithms.

The thesis details three types of beamforming algorithms, two of which have already been discussed extensively in the scientific literature. CBF (Conventional Beamforming) is the simplest of the three, but it is also the least accurate. MUSIC (Multiple Signal Classification) is based on the separation of the received signal's cross-spectral matrix to signal and noise subspaces.

As a novelty, the CS (compressive sensing) based Compressive Beamforming is also studied. This method localizes sources by solving a convex optimization problem, and it offers a more accurate solution compared to linear algebraic algorithms thanks to a sparsity condition.

The problem of tracking moving sound sources is also discussed which can be solved with the help of a Kalman-filter. The traditional version of the Kalman-filter gives an optimal estimation for the state of linear, moving systems, and it can be extended to work on nonlinear systems.

These beamforming algorithms are comparable through simulations in the MATLAB environment, both in noiseless and noisy conditions. We can also evaluate the performance of these methods in practice by performing measurements and prove their usability in localizing sound sources. Moreover, the Kalman-filter is also tested on measurements, to deduce its advantages over beamforming algorithms, and how its different parameters affect its performance.

1 Bevezetés

1.1 Motiváció, cél

Napjainkban sokszor felmerül az igény arra, hogy akusztikai kameraként használt mikrofonrendszerek segítségével, akusztikai módszerekkel tudjuk meghatározni egy hangforrás pozícióját. Ez a módszer jól használható alternatívát jelent olyan esetekben, amikor különböző okokból nem lehet vagy nem érdemes szenzorokat elhelyezni a hangot lesugárzó objektumon. A mikrofonok által vett jelek nyalábformáló algoritmusokkal feldolgozva felhasználhatók arra, hogy megbecsüljük a forráseloszlást, és ez alapján amplitúdótérképeket készítsünk. Mindezt álló és mozgó hangforrás esetén is megtehetjük, utóbbinál az idő függvényében is vizsgálhatjuk a forráseloszlás és a rá végzett becslések változását.

Munkám során a céloom két nyalábformáló algoritmus megismerése és tesztelése, ezek a Conventional Beamforming (CBF) és a Multiple Signal Classification (MUSIC). Ezeket összehasonlítom egy újszerű módszerrel, a Compressive Sensing (CS) alapú Compressive Beamforming-gal, ami jelenleg is nemzetközileg kutatás alatt áll. Mindezek mellett mozgó hangforrás Kálmán-szűrővel való követésének problémájával is foglalkozom. Az algoritmusokat MATLAB környezetben implementálva felhasználom szimulációkhoz és mérési eredmények feldolgozásához, ezáltal kipróbálom a pontosságukat és a megbízhatóságukat különböző hangforrások és körülmények esetén.

1.2 Korábbi kutatások a témában

Ebben a szakaszban röviden megemlítek néhány újabb kutatási eredményt a nyalábformáló algoritmusok és a forráskövetés témakörében a teljesség igénye nélkül, és ezeket összevetem a saját munkámmal, a legfőbb hasonlóságokat és különbségeket keresve.

A hagyományos nyalábformálás [1], [2], [3] és a MUSIC [4], [5], [6] a szakirodalom által részletesen tárgyalt algoritmusok. Xenaki, Gerstoft és Mosegaard 2014-es cikke többek között a CBF-et, a MUSIC-ot és a CS-t tárgyalja, különös figyelmet fordítva ez utóbbira [7]. Az algoritmusok bemutatása mellett kétdimenziós

szimulációkkal hasonlítják össze őket. Konklúzióként levonják, hogy a CS nagyfelbontású iránybecslést tud adni más módszerekhez képest.

Coldrey és Viberg a hagyományos nyalábformálás kiterjesztését végzik el úgy, hogy figyelembe veszik a hangforrás (és egyben maga a hang, mint jel) térbeli kiterjedését [8]. Yang egy dekonvolvált CBF változatot javasol, amivel vékonyabb főnyalábokat és alacsonyabb melléknyaláb-szinteket lehet elérni a módszer robusztusságának megtartása mellett [9]. Gupta és Kar a MUSIC algoritmusnak egy olyan változatát dolgozták ki, amely koherens források iránybecslésére is alkalmas [10]. Yaning, Juntao, Xinghao és Le a számítási komplexitást csökkentik az algoritmus javításával [11].

Xenaki és Gerstoft egy másik, 2015-ös cikke [12], illetve Park, Choo és Seong egy 2018-ban megjelent publikációja [13] olyan CS alapú módszert tárgyal, amely a hagyományosan alkalmazott diszkrét szögtartomány helyett folytonos szögtartományt használ az iránybecslésben, ezáltal kiküszöbölik a módszer bizonyos hátrányait. Mindkét cikkben kétdimenziós szimulációkkal igazolják az eljárásuk működését, az utóbbiban ehhez több pillanatképet együttesen is felhasználnak.

A nyalábformáló algoritmusok, illetve a forráskövető és lokalizáló módszerek a gyakorlatban használhatóak bányászásban [2], víz alatti környezetben gyenge jelek detektálására [9], teremakusztikában és telekonferencia-rendszerekben [14], hallókészülékekben, megfigyelő rendszerekben [15] és navigációs rendszerekben. A CS maga is számos területen alkalmazható, például az orvostudományban [16], ultrahangos képalkotásban [17], csatornakódolás során hibajavításban [18], szeizmikus képalkotásban [19] és képrekonstrukcióban [20]. A nyalábformálás akusztikai módszereken kívül rádióhullámokkal is történhet, ezt alkalmazzák például radarokban és vezeték nélküli kommunikációban is.

Jelen dolgozatban bemutatom a nyalábformálás alapelvét, majd ehhez kapcsolódóan két egyszerű algoritmust, a CBF-et és a MUSIC algoritmus hagyományos formáját. Ezeket összehasonlítom a Compressive Beamforming több különböző változatával. A becslésekhez diszkrét szögtartományokat használok. Tárgyalom a Kálmán-szűrőt és annak nemlineáris kiterjesztését, illetve azt, hogy hogyan használható mozgó hangforrás (vagy hangforrások) állapotbecslésére. A szakirodalomhoz képest újdonságként foglalkozom azzal is, hogy a nyalábformáló algoritmusok hogyan használhatók irány- mellett távolságbecslésre is. Ezeket az algoritmusokat kétdimenziós

szimulációk mellett három dimenzióban is tesztelem. Az eddigi kutatásokhoz képest újdonságként végül pilóta nélküli légi járművek, mint mozgó hangforrások segítségével kapott valós mérési eredményeken is megvizsgálom a működésüket, és értékelem a gyakorlati használhatóságukat. A közelmúltban két TDK dolgozatot írtam a témáról, jelenlegi munkám ezekre is építkezik [21].

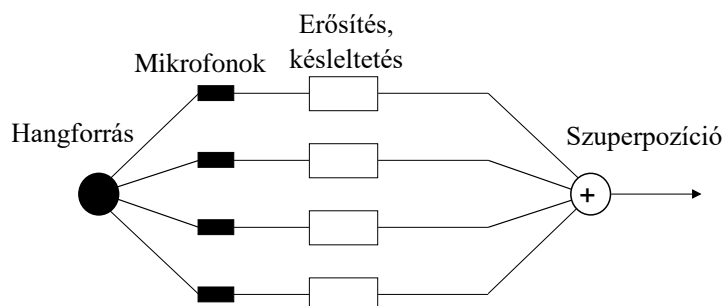
2 Nyalábformáló algoritmusok

Ebben a fejezetben a nyalábformálás elvét és a két vizsgált algoritmust részletezem. A 2.1 szakaszban bemutatom a használt Delay and Sum módszert, ennek véges és végtelen fókusz távolságú eseteit, a mikrofonok és a mérési pontok elhelyezkedésének hatását, és a választott mérési frekvencia alsó és felső határait. A 2.2, 2.3 és 2.4 szakaszok tárgyalják a Conventional Beamforming, a Multiple Signal Classification és a Compressive Beamforming elvét, illetve ez utóbbinak néhány kiegészítését. Végül a 2.5 szakaszban mutatom be a távolságbecslés elvét.

2.1 A nyalábformálás alapjai

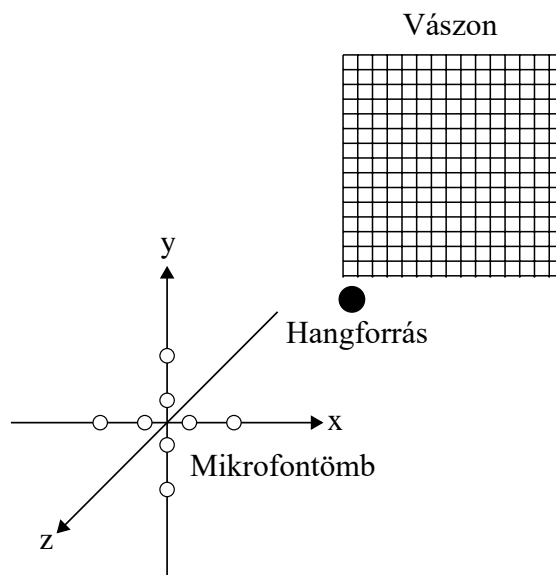
2.1.1 Fókuszálás és forráslokalizáció

A nyalábformáló algoritmusok használatával és a mikrofonrendszerrel két fő feladatot kell elvégeznünk, ezek az akusztikai fókuszálás és a forráslokalizáció. Az akusztikai fókuszálás a Delay and Sum nevű módszeren alapszik, aminek lényege, hogy gömbkarakterisztikájú mikrofonokat többféle különböző elrendezésben elhelyezhetünk, majd ezek vett jelét külön-külön erősítjük és késleltetjük, majd ezeket a módosított jeleket szuperponáljuk [22]. A jelek egymáshoz képesti fázisviszonyai megfelelő beavatkozás esetén azt eredményezik, hogy a tér egy adott irányából érkező hangot kiemeljük velük, más irányokból érkezőket pedig elnyomjuk. A mikrofonok térbeli elrendezésével és a vett jelek erősítésével és késleltetésével lehet kialakítani egy számunkra megfelelő iránykarakterisztikát, a késleltetések további befolyásolásával pedig hasonló hatást érhetünk el, mintha elforgatnánk a mikrofontömböt, ezáltal többféle irányra lehet fókuszálni.



1. ábra: A Delay and Sum módszer. A mikrofonok vett jeleit külön-külön erősítjük és késleltetjük, majd szuperponáljuk, ezzel egy adott irányra fókuszálunk.

A másik fő feladat, a forráslokalizáció (vagy nyálábformálás) során a térben lévő hangforrások irányát és teljesítményét becsüljük meg a nyálábformáló algoritmusok segítségével. A térben kijelölünk egy ponthalmazt, amik közösen alkotják az úgynevezett akusztikai vásznat, és az ezeken a pontokon elhelyezett virtuális források közül azokat választjuk ki, amelyek által létrehozott hangtér a leginkább hasonlít a mikrofontömb által vett valós hangjelekhez. A két feladat (fókuszálás és nyálábformálás) egymástól függetlenül is végrehajtható, de jelen munka során mindkettőt együttesen használva hajthatjuk végre az iránybecslést és a forrásjel kiemelését.



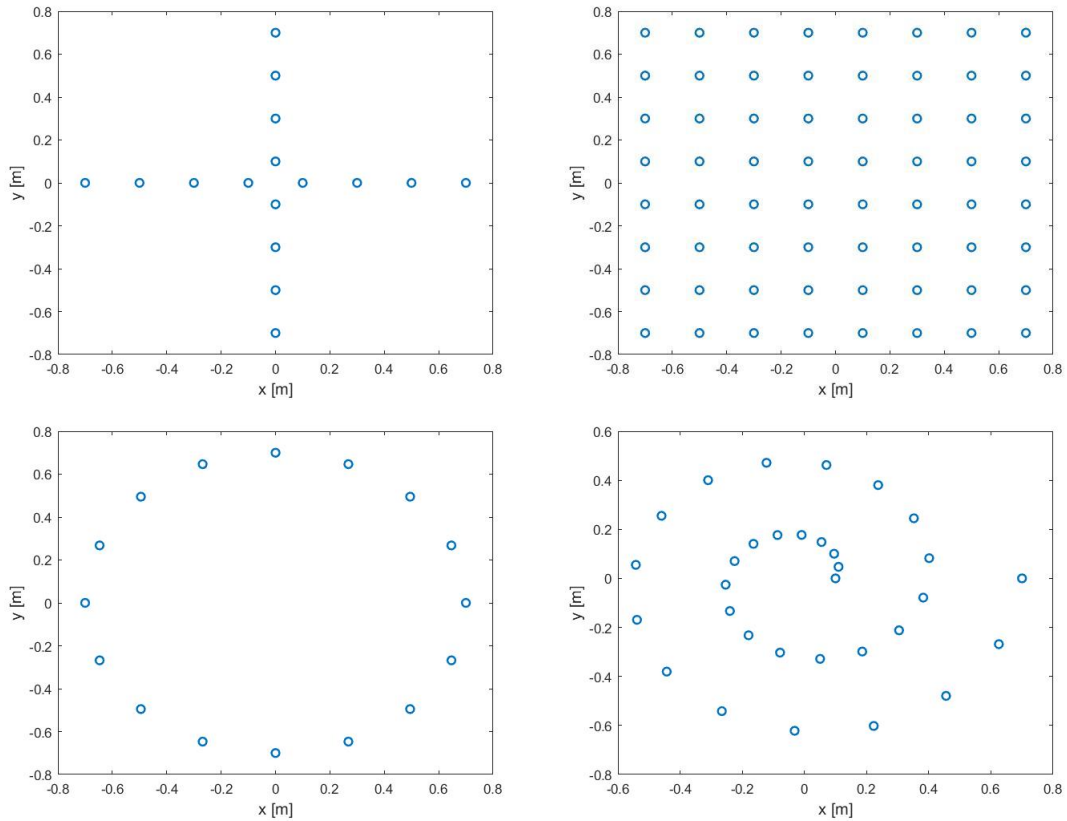
2. ábra: Általános mérési elrendezés.

2.1.2 A mikrofonok és a vászon elhelyezkedése

A mikrofonokat többféle elrendezésben helyezhetjük el, például vonal, kereszt, téglalap, négyzetrács, kör vagy spirál formában. Az elrendezés formája és az egyes mikrofonok pozíciói is meghatározzák a vett jeleket módosító beavatkozó függvényt, illetve a mikrofontömb iránykarakterisztikáját is. Két dimenzióban, ahol csak egy síkon belül, egy szögparaméterrel leírható irányokra fókuszálunk, a vonal elrendezést használjuk. Három dimenzióban, ahol az egész térben, két szögparaméterrel leírható irányokra fókuszálhatunk, a drágán megvalósítható négyzetrács elrendezés helyett érdemes téglalapot vagy keresztet (esetleg kört vagy spirált) alkalmazni.

Az akusztikai vászon pontjai/irányai általában egy egyenesen/körön, vagy egy síkban/gömbfelületen helyezkednek el egyenletesen. A hangnak a vászon és a

mikrofonrendszer közötti terjedésében figyelembe kell venni a levegő, mint közvetítő közeg bizonyos tulajdonságait, mint például a légnyomást és a hőmérsékletet, amik alapján kiszámolható a hangsebesség.

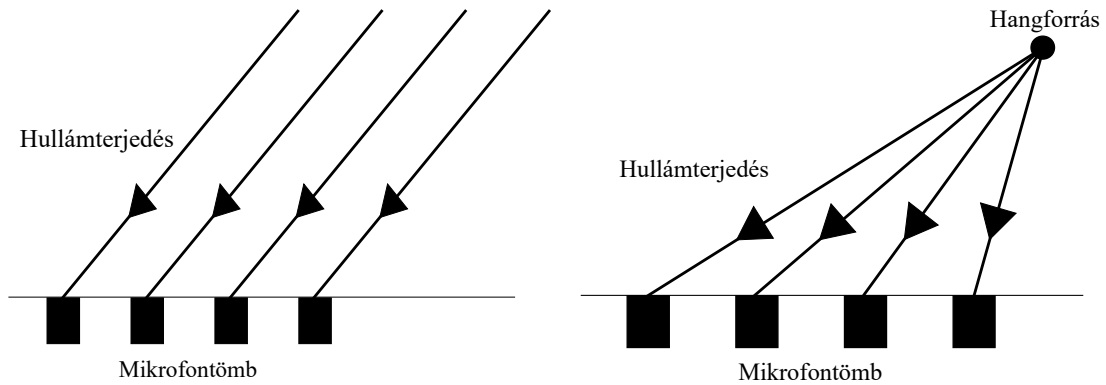


3. ábra: Kereszt, négyzetrács, kör és spirál elrendezés.

A nyalábformálást kétféleképpen is elvégezhetjük attól függően, hogy a forrás milyen messze helyezkedik el a mikrofontömbtől. Ha ehhez a távolsághoz képest elhanyagolható a tömb mérete, akkor alkalmazhatjuk a végtelen fókusz távolságú nyalábformálást. Ebben az esetben a forrás által kibocsátott hang hullámfrontja síkfelületnek tekinthető, és a mikrofonokhoz közel azonos amplitúdójú és beesési szögű hullámok érkeznek, de különböző fázisban. Itt a nyalábformálás során csak a beesési szöget kell figyelembe venni, a forrás és a mikrofon távolságát nem, és a vászon definiálásakor is elég csak irányokat (szögeket) megadni.

Ha a mikrofonok és a forrás távolsága összemérhető a mikrofonrendszer méretével, akkor véges fókusz távolságú nyalábformálást kell használnunk. Ilyenkor a hullámfront egy gömbfelület, azaz a mikrofonok által vett jelek különböző amplitúdójúak és fázisúak, és a beesési szög is különböző minden mikrofontól. A nyalábformálás során a beesési szöget és a távolságot is figyelembe kell venni, a vásznat

pontok koordinátáinak összességéként kell definiálni. Érdeemes megvizsgálni azt is, milyen mérési és szimulációs eredmények adódnak akkor, ha a forrás távolsága és a vászon távolsága a mikrofontömbtől különbözik.



4. ábra: Végtelen és véges fókusz távolságú nyalábformálás.

2.1.3 A vizsgálati frekvencia

A nyalábformáló algoritmusokat frekvenciatartományban alkalmazzuk, ami ez esetben azt jelenti, hogy egy adott pillanatban (rövid időintervallumban) minden mikrofon vett jelének egy adott frekvenciájú komponensét használjuk fel a becsléshez. A megfelelő frekvencia választása kulcsfontosságú, be kell tartanunk a térbeli mintavételezés elvét, ugyanis bizonyos határfrekvenciák által behatárolt tartományon kívül a mérési elvbe hiba csúszik.

Amennyiben a frekvencia megfelelően lett kiválasztva, a forrás irányára fókuszálva a mikrofonok vett jelét késleltetés után összeadva azok erősítik egymást, mert fázisban vannak, más irányokra pedig gyengítik, vagy akár ki is oltják egymást a fáziskülönbségek miatt. Ha a frekvencia túl nagy (a hullámhossz túl kicsi), akkor térbeli átlapolódás történik, a mikrofonok jele olyan irányra fókuszálva is fázisban lehet, ahol valójában nincs forrásunk. Ha a frekvencia túl kicsi (a hullámhossz túl nagy), akkor ugyan nem látunk szellemforrásokat, de rossz irányra fókuszálva sem lesz nagy a fáziskülönbség a jelek között, és csak nagyon durva becslést tudunk adni [23]. A felső határfrekvencia az (1) összefüggéssel számítható ki:

$$f < \frac{c}{2d}, \quad (1)$$

ahol c a hangsebesség, d pedig a mikrofonok egymástól való távolsága (egyenletes elrendezést feltételezünk). Minél közelebb vannak egymáshoz a mikrofonok, annál nagyobb a felső határfrekvencia. Az alsó határfrekvenciára nincs konkrét összefüggés,

minél kisebb a frekvencia, úgy fokozatosan annál elmosódottabb lesz az amplitúdótérkép.

2.2 Conventional Beamforming

A Delay and Sum módszerben a mikrofonok vett jelének az erősítését és késleltetését minden mikrofonhoz egy külön beavatkozó függvénnyel tehetjük meg. A beavatkozó függvényt úgy is fel lehet fogni, hogy valódi forráseloszlás-vektor (\mathbf{x}) és a mikrofonok által vett jelek vektora (\mathbf{y}) között a kapcsolatot egy \mathbf{A} érzékelési mátrix valósítja meg. Ennek a mátrixnak a segítségével tudunk becslést adni a vett jelekből a forráseloszlásra a nyálábformáló algoritmusokkal. Az érzékelési mátrix a mikrofonok és a vászon elhelyezkedése, illetve a választott mérési frekvencia alapján határozható meg. Az \mathbf{x} és \mathbf{y} vektorok frekvenciatartománybeli információt hordoznak, azaz azt mondják meg, hogy a kisugárzott/vett jeleknek a kiválasztott vizsgálati frekvencián mekkora amplitúdójú komponensük van egy rövid időszakason.

Az érzékelési mátrix elemei a következő formulák alapján határozhatók meg, a (2)-es egyenlet kétdimenziós, a (3)-as pedig háromdimenziós esetre vonatkozik [7], [8]:

$$\mathbf{A}(i, j) = \frac{1}{\sqrt{M}} e^{jk\alpha_i \sin \eta_j}, \quad (2)$$

$$\mathbf{A}(i, j) = \frac{1}{\sqrt{M}} e^{jk(\alpha_i \sin \theta_j \sin \varphi_j + \beta_i \cos \theta_j \sin \varphi_j)}, \quad (3)$$

ahol M a mikrofonok száma ($1/\sqrt{M}$ -mel normalizálunk), k pedig a hullámszám (a vizsgálati körfrekvencia és a hangsebesség hányadosa). α_i és β_i az i -dik mikrofon koordinátái derékszögű koordináta-rendszerben az α és a β tengelyek mentén (a mikrofonok az α - β síkban helyezkednek el), η_j , θ_j és φ_j pedig a vászon j -dik pontjának az irányát azonosítják, végtelen fókusztávolság esetén (η az α tengellyel bezárt szög mínusz 90° , φ az irány α - β síkra levetítésének az α tengellyel bezárt szöge, θ pedig a γ tengellyel bezárt szög, a koordináta-rendszer középpontja egyben a mikrofonrendszer geometriai középpontja is). Véges fókusztávolság esetén természetesen minden mikrofon-vászonpont párhoz külön-külön beesési szögek tartoznak. Az \mathbf{A} mátrixnak összesen annyi oszlopa van, ahány pontból áll a vászon, és annyi sora, ahány mikrofonból áll a mikrofontömb. Az érzékelési mátrix az alábbi módon hozza létre a kapcsolatot a forrás (\mathbf{x}) és a vett jelek (\mathbf{y}) között:

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \quad (4)$$

Az egyik legegyszerűbb nyálábformálási módszer a CBF, ami során az \mathbf{y} -t balról szorozzuk az \mathbf{A} mátrix konjugált transzponáltjával:

$$\hat{\mathbf{x}} = \mathbf{A}^H \mathbf{y}. \quad (5)$$

Ennek a módszernek az előnye az egyszerűsége és kis számításigénye, hátránya azonban a pontatlansága, ugyanis az antenna iránykarakterisztikájának a főnyalábja véges, de nullánál nagyobb szélességű, és a valódi forráspozíció közvetlen közelében is zérustól jelentősen eltérő értékeket becslünk, az amplitúdótérkép elmosódott lesz.

2.3 Multiple Signal Classification

A MUSIC algoritmus egy lineáris algebrai módszer [24], amelynek során a vett jelekből képzett spektrális keresztkorrelációs mátrixot szétválasztjuk jel és zaj alterekre, majd a zaj sajátvektorainak és az érzékelési mátrixnak a segítségével becslünk.

A keresztkorrelációs mátrixot (\mathbf{R}) az alábbiak szerint határozhatjuk meg:

$$\mathbf{R} = \mathbf{y}\mathbf{y}^H. \quad (6)$$

A mátrix sajátvektorait sorba rendezzük a hozzájuk tartozó sajátértékek szerint növekvő sorrendbe. Ezek közül a K legnagyobb sajátértékhez tartozó sajátvektor alkotja a jel alterét (\mathbf{U}_s), a többi pedig a zajét (\mathbf{U}_n), ahol K a források becsült száma (ezt például úgy is megbecsülhetjük, hogy megnézzük, hogy a sorba rendezett sajátértékek közül melyik két szomszédos hányadosa a legnagyobb, és ott lesz a jel és a zaj alterek határa). Az érzékelési mátrixszal és a zaj alterrel a következő formula szerint végezzük a becslést [4], [5], [7], [10]:

$$\mathbf{P}_{MUSIC} = \frac{1}{\mathbf{A}^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{A}}, \quad (7)$$

vagyis az $\mathbf{A}^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{A}$ mátrix minden elemének vesszük a reciprokát.

A módszer hátránya, hogy előre meg kell becsülni a források számát, illetve az, hogy csak inkoherens forrásokra működik. Előnye azonban a hagyományos nyálábformáláshoz képest nagyobb pontosság, illetve a számításigénye viszonylag alacsony, valós idejű alkalmazásokban is használható.

2.4 Compressive Beamforming

A Compressive Beamforming a Compressive Sensing (CS) eljárás alapul. Ez a módszer a forráslokalizációt egy konvex optimalizálási problémára visszavezetve oldja meg, azonban a forrásnak meg kell felelnie egy ritkasági kényszerfeltételnek annak érdekében, hogy a módszer pontos becslést adhasson [1], [25], [26]. A

Compressive Beamforming előnye a pontossága a kényszerfeltétel teljesülése esetén, hátránya viszont a nagy számításigény, ezért valós időben ritkán használható.

2.4.1 Compressive Beamforming zajmentes esetben

A ritkasági kényszerfeltétel azt jelenti, hogy kevés forrásunk van a vásznon pontjainak számához képest (amik elég sűrűn helyezkednek el ahhoz, hogy ne adódjon probléma a források rossz illeszthetősége miatt). Ezt matematikailag úgy is megfogalmazhatjuk, hogy a forrást leíró vektornak (vagy mátrixnak) kevés nem-nulla eleme van. Ha ez a feltétel teljesül, akkor a célunk az, hogy a forrásra végzett becslésünknek minél kevesebb zérustól különböző értéke legyen, miközben $\mathbf{Ax} = \mathbf{y}$ teljesül:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_0 \text{ úgy, hogy } \mathbf{Ax} - \mathbf{y} = 0, \quad (8)$$

ahol \mathbf{x} a becslésünk a forrásra, \mathbf{y} a vett jelek mátrixa, \mathbf{A} az előre meghatározott és fix (\mathbf{x} -től és \mathbf{y} -tól nem függő) érzékelési mátrix, $\|\mathbf{x}\|_0$ pedig az \mathbf{x} vektor l_0 -normája. Egy adott vektor l_p -normája általánosan:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad (9)$$

ami $p=0$ -ban éppen a nem-zérus elemek számát adja meg:

$$\|\mathbf{x}\|_0 = \sum_{i=1}^n 1_{x_i \neq 0}. \quad (10)$$

A (8) probléma megoldására hatékony algoritmus nem létezik, azonban ha az l_0 -norma szerinti feltételt relaxáljuk l_1 -normára, akkor egy jóval könnyebben megoldható problémához jutunk:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_1 \text{ úgy, hogy } \mathbf{Ax} - \mathbf{y} = 0. \quad (11)$$

A Compressive Sensing módszerek területén nagy áttörés volt annak a bebizonyítása, hogy megfelelően ritka forrás esetén (jóval kevesebb forrás van, mint vizsgálati pont a vásznon) a (11) ekvivalens a (8) problémával, és felhasználható nyálábformáló algoritmusokban [27], [28], [29].

2.4.2 Compressive Beamforming zajos esetben

A gyakorlatban a mikrofonok által vett jelre zaj is rakódik, amit a nyálábformáló algoritmusban figyelembe kell venni. Additív zajt feltételezve:

$$(\mathbf{y} = \mathbf{Ax} + \mathbf{n}) \quad (12)$$

a konvex optimalizálási problémánk a következőképpen módosul [30]:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_1 \text{ úgy, hogy } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \varepsilon. \quad (13)$$

Az ε a zaj miatt bevezetett hibahatár a becslésre, amire teljesülnie kell az $\|\mathbf{n}\|_2 \leq \varepsilon$ feltételnek.

Ha a zajszint ismeretlen, nekünk kell becslést adnunk rá, azonban felmerül a kérdés, hogy az alul- vagy a felülbecslés közül melyik a kedvezőbb. A zajszint alulbecslése esetén előfordulhat, hogy a megoldás kevésbé lesz ritka a valóságnál, azaz olyan helyeken is forrást látunk, ahol valójában nincsenek. Felülbecsléskor ennek az ellenkezője történik, azaz lehet, hogy a megoldás túl ritka lesz, és a kisebb erősségű forrásokat nem találjuk meg. A kettő közül az alulbecslés a kedvezőbb, ugyanis a ritkaság csökkenése kiküszöbölhető újrasúlyozott l_1 minimalizálással [31], és szintkorrekcióval.

Az újrasúlyozott l_1 minimalizálás során először egy kezdeti becslést kell végrehajtani a (13) formula segítségével. Ezután iteratíván megoldjuk az alábbi súlyozott problémát:

$$\min \|\mathbf{Wx}\|_1 \text{ úgy, hogy } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \varepsilon, \quad (14)$$

ahol \mathbf{W} egy diagonális mátrix, az elemeinek értéke a kezdeti becslés előtt 1, utána pedig a legutóbbi becslés alapján adhatóak meg:

$$w_i = \frac{1}{\hat{x}_i + \xi}. \quad (15)$$

Itt ξ egy nullánál nagyobb szám a legkisebb várható amplitúdó nagyságrendjében, ami biztosítja, hogy a súlyozás egyik eleme se legyen végtelen, \hat{x}_i a forrásra végzett legutóbbi becslés i -dik eleme. Az iterációt akkor hagyjuk abba, amikor az utolsó két becslés közel megegyezik, ez általában két iterációval megvalósul:

$$|w_i x_i|^{k+1} = \begin{cases} \frac{\hat{x}_i}{\hat{x}_i + \xi} \approx 1, & \text{ha } |x_i| \geq 0, \\ 0, & \text{egyébként.} \end{cases} \quad (16)$$

Az újrasúlyozott megoldás tovább javítható szintkorrekció alkalmazásával. Ha az érzékelési mátrixnak csak azokat az oszlopait hagyjuk meg, amelyek a megoldásban csúcserősségnek felelnek meg, a többinek az elemeit pedig nullával helyettesítjük, akkor a kapott \mathbf{A}_a mátrixot felhasználva oldjuk meg a következő egyenletrendszer [32]:

$$\hat{\mathbf{x}}_a = \mathbf{A}_a^+ \mathbf{y}, \quad (17)$$

ahol \mathbf{A}_a^+ a mátrix pszeudo inverze.

2.4.3 Compressive Beamforming pontszórásfüggvényekkel

A Compressive Beamforming módszernek bevezethetjük egy olyan változatát, amelynek során egy másfajta konvex optimalizálási problémát oldunk meg a pontszórás-függvényeket felhasználásával. A pontszórás-függvényeket minden virtuális forráspozícióra definiálhatjuk úgy, hogy feltételezzünk egy egységnyi teljesítményű forrást az adott pontban, majd pedig megnézzük, hogy a vászon összes pontjára mekkora forrás-teljesítményt becsülhetünk hagyományos nyalábformálással (CBF). A pontszórás-függvényekből létrehozhatunk egy mátrixot, aminek annyi oszlopa (és sora) van, ahány virtuális forráspozíció, és a k -edik oszlopa a k -edik ponthoz tartozó függvény vektora. Ez a mátrix az érzékelési mátrixból így számítható ki:

$$\mathbf{A}_{PSF} = \mathbf{A}^H \mathbf{Y} = \mathbf{A}^H \mathbf{A} \mathbf{E} = \mathbf{A}^H \mathbf{A}. \quad (18)$$

A módosított konvex optimalizálási probléma során a hagyományos nyalábformálás eredményét próbáljuk előállítani a pontszórás-függvények lineáris kombinációjaként:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_1 \text{ úgy, hogy } \|\mathbf{A}_{PSF} \mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon \text{ és } \mathbf{x} \geq 0, \quad (19)$$

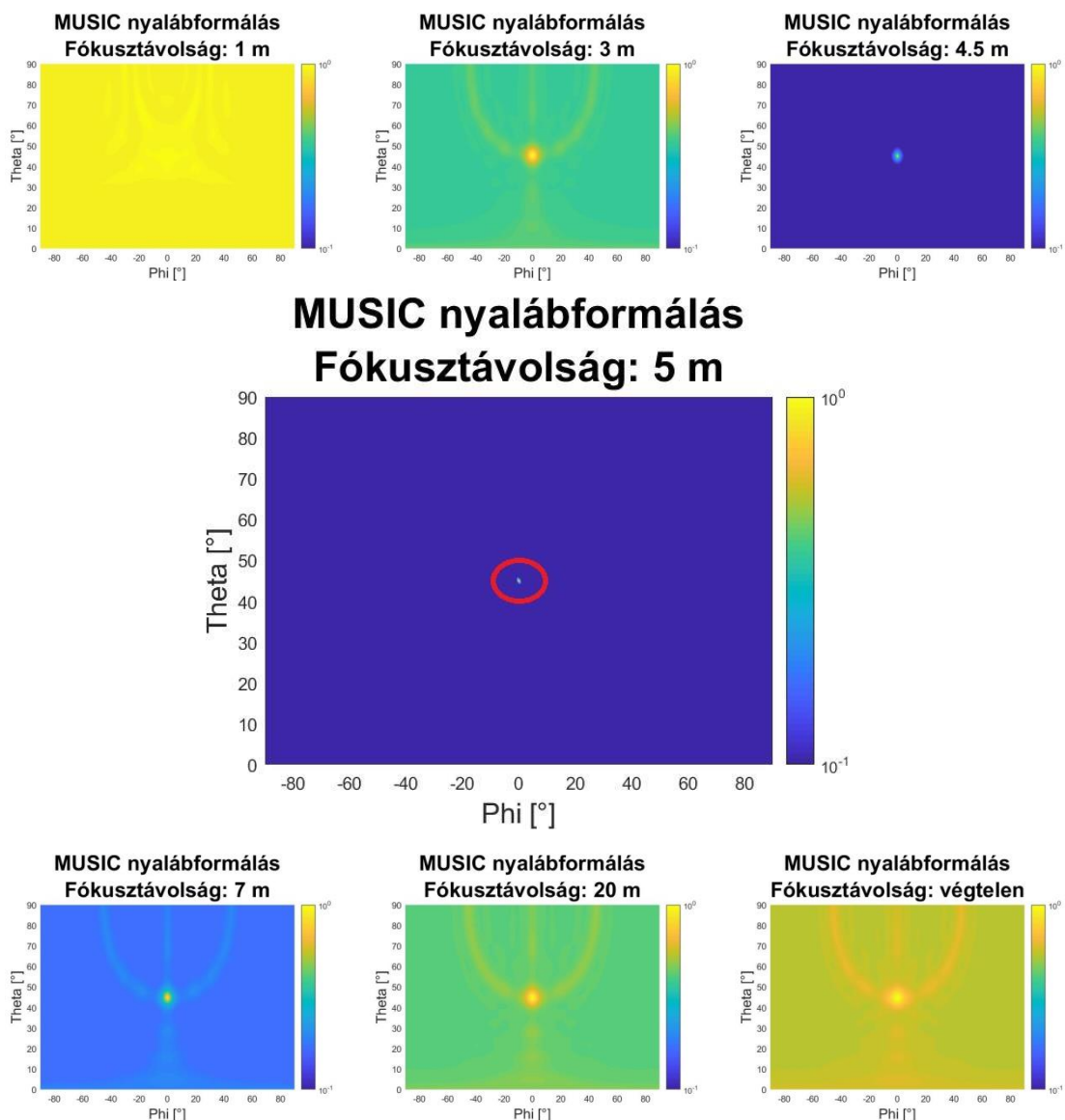
ahol \mathbf{y} a becslés hagyományos nyalábformálással, és az $\mathbf{x} \geq 0$ feltétel fizikailag azt jelenti, hogy egy hangforrásnak a teljesítménye nemnegatív.

2.5 Távolságbecslés az akusztikus vászon kiterjesztésével

Az akusztikai fókuszálás során fontos, hogy a fókusztávolságot megfelelően válasszuk meg. A szem, mint optikai képalkotó működése során is kulcsfontosságú ez a paraméter, mert ha egy adott tárgyra fókuszálunk, akkor azt élesen, az attól különböző távolságra lévő tárgyakat pedig homályosabban látjuk. Hasonló jelenség lép fel akusztikai kameraként használt mikronrendszerek használata során is, ha nem pontosan akkora a fókusztávolság, mint a hangforrás távolsága, akkor elmosódottabb kameraképet kapunk.

A kameraképein látható a képalkotás fókusztávolságtól való függése. A mérési elrendezés hasonló az eddigi háromdimenziós szimulációkéhoz, viszont a vászon pontjainak a mikrofóntomb közepétől vett távolságai minden ábrára más (rendre 1, 3,

4,5, 5, 7, 20 méter és végtelen), és egy fehérzaj forrás helyezkedik el a tömbtől 5 méterre, $\varphi=0^\circ$ és $\theta=45^\circ$ irányban (ezekben a szimulációkban háttérzaj nincs hozzáadva a mikrofonok vett jeleihez). 5 méteres fókusz távolságnál a mikrofonrendszer nagyon nagy pontossággal meg tudja határozni a hangforrás irányát, azonban ahogy távolodunk a helyes fókusz távolságtól, egyre rosszabb kameraképet kapunk. A nyálábformáló algoritmusok becslésének maximumhelye egyre bizonytalanabb, sokkal nagyobb amplitúdójúak lesznek a kereszt mikrofonelrendezésből adódó melléknyalábok, és a becslés alapszintje is egyre magasabb a maximumhoz képest. Ebből kiindulva többféle fókusz távolság használata felhasználható távolságbecslésre.



5. ábra: A mikrofontömbtől 5 méterre elhelyezkedő hangforrásról alkotott kameraképek, különböző fókusz távolságokkal.

A nyalábformáló algoritmusok használata során az akusztikus vászon virtuális forráspozíciói általában egy síkon vagy egy gömbfelület mentén helyezkednek el szimulációk és mérési eredmények feldolgozása során. A virtuális forráspozíciók ilyen elrendezése lehetővé teszi a hangforrások irányának a becslését, viszont a távolságuk ismeretlen marad. Enélkül nem tudhatjuk pontosan a pozíciójukat, és így a sebességükre sem lehet becslést adni, ezért érdemes valamilyen módon kiterjeszteni az akusztikus vásznat úgy, hogy egy adott irány mentén több pontja is legyen különböző távolságokra. Ezáltal lényegében egy háromdimenziós vásznat hozunk létre, ami már használható távolságbecslésre is.

Az első kézenfekvő megoldás az lehet, hogy a teljes vizsgált szögtartományon belül felvesszünk vásznakat különböző távolságokra. A virtuális forráspozíciók koordinátái úgy kaphatjuk meg, hogy az eredetileg vizsgált pontok koordinátáit egy-egy konstanssal megszorozzuk. A becslés elve az, hogy mindegyik vászon mindegyik pontjára kapunk egy amplitúdót a használt nyalábformáló algoritmus segítségével, és a forrás pozícióját annak a pontnak feltételezzük, amelyikre a legnagyobb amplitúdót kapjuk. Ahhoz, hogy az irány mellett a távolságot is pontosan meg tudjuk állapítani, a pontokat sűrűn kell felvenni. Ez viszont azt jelenti a vászon ilyen módú kiterjesztése esetén, hogy összesen nagyon sok virtuális forráspozíciót kell felvenni, az eredetinek a sokszorosát, ami jelentősen megnöveli a számításgépi igényt, ezért ez a módszer nagyon lassú, a gyakorlatban nem érdemes alkalmazni.

A módszer gyorsítható úgy, ha nem az egész vizsgált szögtartományon belül vesszünk fel vásznakat különböző távolságokra, hanem csak a becsült irányok körül egy szűkebb tartományon. Ehhez először fel kell venni egy kiindulási ponthalmazt a mikrofontömbtől közepes távolságra (közepes távolság alatt itt az értendő, hogy a várható legközelebbi és legtávolabbi forrás is jól megkülönböztethető legyen a kameraképen). Miután ennek a segítségével meghatároztuk a forrás (vagy források) irányát, a vászonnak egy olyan részletét kiválasztjuk, aminek a közepén van a becsült irány, majd pedig ezt a részletet sokszorosítjuk az előző módszerhez hasonlóan. Ezzel sikerült lecsökkenteni az összes virtuális forráspozíciók számát, és felgyorsul az algoritmus.

Tovább gyorsíthatjuk az algoritmust, ha nem a vászonnak egy-egy részletét, hanem csak a becsült irányhoz tartozó egyetlen pontot sokszorosítjuk. Az eredeti vásznon elvégezzük a nyalábformálást, és az amplitúdótérképen talált maximumhoz tartozó pont irányában felvesszünk több pontot különböző távolságokban, amikre

fókuszálunk, és a rajtuk elvégzett nyalábformálás maximumhelyén feltételezzük a hangforrás távolságát. Ennek az eljárásnak az előnye, hogy sok különböző távolság felvétele esetén is kicsi a számításigény, így ez nem korlátozza a becslés pontosságát. Hátránya viszont, hogy az iránybecslés pontatlansága tovább növelheti a távolságbecslés pontatlanságát, főleg zajjal terhelt méréseken.

Az algoritmus tovább optimalizálható, a számítási költségek tovább csökkenthetők. Megtehetjük azt is, hogy nem vesszük fel egyszerre az összes lehetséges távolságot egy irányban, hanem csak egyre elvégezzük a nyalábformálást, majd utána lépésről lépésre haladunk egyre közelebb a valódi pozícióhoz. Ez az elv további részletes kidolgozást igényel, ez ennek a dolgozatnak nem tárgya.

3 Mozgó forrás követése Kálmán-szűrővel

A Kálmán-szűrő egy olyan algoritmus, aminek a segítségével mozgó, változó rendszerek állapotára végezhetünk optimális becslést. Ehhez a becsléshez felhasználjuk a mérési eredményeken kívül a rendszer korábbi állapotáról rendelkezésre álló információt is. Ez nagyobb pontosságot eredményez, mintha csak a mért eredményeket vennénk figyelembe, és predikciót is végezhetünk a rendszer egy jövőbeli állapotára. A Kálmán-szűrővel minimalizálhatjuk a becslés hibájának négyzetes középértékét. Esetünkben a változó rendszer egy mozgó hangforrás, aminek a pozíciójára és a sebességére végzünk becslést.

A következő fejezetben az algoritmus két változatát mutatom be. Az első lineáris rendszerek esetén használható, lineáris állapotbecslésen alapuló Kálmán szűrő, a második pedig ennek egy kiterjesztése, az úgynevezett Unscented Kalman Filter (UKF), ami használható nemlineáris rendszerek becslésére is. Végül pedig javaslatot teszek arra, hogyan bővíthető ki ez a módszer úgy, hogy egyszerre több hangforrás követésére is alkalmas legyen.

3.1 Optimális lineáris állapotbecslés

Feltesszük, hogy a rendszerünk lineáris és időinvariáns. Diszkrét időben így az állapotvektorra a következő összefüggést írhatjuk le [23], [33], [34]:

$$\mathbf{x}(n+1) = \mathbf{A}\mathbf{x}(n) + \mathbf{B}\mathbf{u}(n) + \mathbf{w}(n). \quad (20)$$

Itt az $\mathbf{x}(n)$ az állapotvektor az n -edik mintavételi időpontban; $\mathbf{u}(n)$ a bemeneti vektor (gerjesztés); \mathbf{A} és \mathbf{B} a rendszerre jellemző mátrixok, amik leírják az állapotváltozást két mintavételi időpont között; $\mathbf{w}(n)$ pedig a bemenetet terhelő zajvektor (angolul process noise), ami a modell pontatlanságait reprezentálja. Az $\mathbf{x}(n)$ állapotvektor mozgó hangforrás követése esetében pozíció- és sebességinformációt is tartalmaz (három dimenzióban így egy hatelemű vektor), az előbbit például egy nyalábformáló algoritmus által becsült maximumhelyből nyerhetjük ki.

Az $\mathbf{y}(n)$ kimeneti vektorra felírható egyenlet:

$$\mathbf{y}(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{u}(n) + \mathbf{v}(n), \quad (21)$$

ahol \mathbf{C} és \mathbf{D} a rendszerre jellemző mátrixok (\mathbf{D} -t nullának vehetjük, mert a bemenet nem hat közvetlenül a kimenetre); $\mathbf{v}(n)$ pedig a méréseket terhelő zajvektor. A két zajvektor, $\mathbf{w}(n)$ és $\mathbf{v}(n)$, zérus várható értékű, egymással korrelálatlan, normális eloszlású fehérzajok, rendre $\mathbf{Q}(n)$ és $\mathbf{R}(n)$ kovarianciamátrixszal:

$$E[\mathbf{w}(n)\mathbf{w}(m)^T] = \mathbf{Q}(n)\delta_{n-m}, \quad (22)$$

$$E[\mathbf{v}(n)\mathbf{v}(m)^T] = \mathbf{R}(n)\delta_{n-m}, \quad (23)$$

$$E[\mathbf{w}(n)\mathbf{v}(m)^T] = 0. \quad (24)$$

A δ_{n-m} a Kronecker delta függvény, aminek értéke 1, ha $n=m$, és 0, ha $n \neq m$.

Első lépésként a Kálmán-szűrő az állapotegyenlet alapján egy előzetes (a priori) becslést végez el az állapotvektorra és a kimenetre az $n+1$ -edik mintavételi időpontra:

$$\mathbf{x}^- = \mathbf{A}\tilde{\mathbf{x}}(n) + \mathbf{B}\mathbf{u}(n), \quad (25)$$

$$\tilde{\mathbf{y}}(n) = \mathbf{C}\mathbf{x}^-(n). \quad (26)$$

A mérési eredmények ($\mathbf{y}(n)$) és a becslés ($\tilde{\mathbf{y}}(n)$) közötti különbség:

$$\mathbf{d}(n) = \mathbf{y}(n) - \tilde{\mathbf{y}}(n). \quad (27)$$

Ennek a különbségnek a segítségével tudunk az állapotvektorra egy utólagos (a posteriori), a mérési eredményeket figyelembe vevő becslést végezni egy korrekciós lépéssel:

$$\tilde{\mathbf{x}}(n+1) = \mathbf{x}^+ = \mathbf{x}^- + \mathbf{K}_n\mathbf{d}(n), \quad (28)$$

ahol \mathbf{K}_n egy megfelelően választott korrekciós mátrix. Az $\tilde{\mathbf{x}}(n)$ (az n -edik időponti állapotvektorra már korábban végzett becslés eredménye) várható értéke $\mathbf{x}(n)$, a kovarianciamátrixát vegyük $\mathbf{P}(n)$ -nek.

Célunk tehát az optimális becslést lehetővé tevő \mathbf{K}_n korrekciós mátrix meghatározása. Ehhez szükségünk van az állapotvektor kovarianciamátrixára, amit a következőképpen számíthatunk ki:

$$\mathbf{P}_n^- = \mathbf{A}\mathbf{P}_{n-1}\mathbf{A}^T + \mathbf{Q}_n. \quad (29)$$

Ebből az optimális \mathbf{K}_n mátrix:

$$\mathbf{K}_n = \mathbf{P}_n^- \mathbf{C}^T (\mathbf{C}\mathbf{P}_n^- \mathbf{C}^T + \mathbf{R}_n)^{-1} = \mathbf{P}_n^+ \mathbf{C}^T \mathbf{R}_n^{-1}. \quad (30)$$

Ezek után már meghatározhatjuk mérési eredményeket figyelembe vevő állapotvektorbecslés (\mathbf{x}^+) kovarianciamátrixát:

$$\begin{aligned} \mathbf{P}_n^+ &= (\mathbf{I} - \mathbf{K}_n \mathbf{C}) \mathbf{P}_n^{-1} (\mathbf{I} - \mathbf{K}_n \mathbf{C})^T + \mathbf{K}_n \mathbf{R}_n \mathbf{K}_n^T = \\ &= (\mathbf{P}_n^{-1} + \mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} = \\ &= (\mathbf{I} - \mathbf{K}_n \mathbf{C}) \mathbf{P}_n^-. \end{aligned} \quad (31)$$

Érdemes megjegyezni, hogy a korrekciós mátrix csak a zajok kovarianciájától (\mathbf{R} és \mathbf{Q}), illetve a rendszermátrixoktól (\mathbf{A} és \mathbf{C}) függ, az aktuális méréstől és az állapotvektortól nem. Ez azt jelenti, hogy értéke kiszámolható előre, és ezzel gyakorlati alkalmazásokban implementálva a Kálmán-szűrőt a számítási költségeket jelentősen csökkenthetjük.

3.2 Unscented Kalman Filter

A gyakorlatban a vizsgált rendszerünk sokszor nemlineáris, így szükség van a Kálmán szűrő kiterjesztésére (EKF, Extended Kalman Filter). Ezek a kiterjesztett algoritmusok legtöbbször valamilyen linearizáláson alapulnak. Egy másik lehetőség az úgynevezett Unscented Kalman Filter (UKF), amelynek során a becült kovarianciamátrix segítségével szórt állapotvektorokat hozunk létre egy ellipszis vagy ellipszoid mentén, majd ezeken végezzük el a Kálmán-szűrést [23], [34], [35], [36].

Az UKF használata során a következő lépéseket végezzük el:

1. Mindegyik időlépésben választunk $2N$ darab szigma-pontot (ahol N az állapotér dimenziója) az alábbiak szerint:

$$\mathbf{x}_i^\sigma, \mathbf{x}_{N+i}^\sigma = \mathbf{x}_n \pm \boldsymbol{\sigma}_i, \quad i = 1 \dots N, \quad (32)$$

ahol $\boldsymbol{\sigma}_i$ a $\sqrt{N\mathbf{P}_n}$ mátrix i -edik sora. Ha így választjuk meg $\boldsymbol{\sigma}_i$ értékét, akkor a szigma pontok statisztikai éppén megegyeznek az állapotvektorával, azaz a pontok átlaga az $\mathbf{x}(n)$ állapotvektor, a tapasztalati szórásuk pedig kiadja a \mathbf{P}_n kovarianciamátrixot.

2. A szigma pontokra alkalmazzuk az állapotegyenletet, és a kapott pontokat $\mathbf{x}_i^{\sigma*}$ -vel jelöljük. Ezek átlagával becsüljük az új állapotot:

$$\tilde{\mathbf{x}}^- = \frac{1}{2N} \sum_{i=1}^{2N} \mathbf{x}_i^{\sigma*}, \quad (33)$$

a kovarianciájuk pedig:

$$\tilde{\mathbf{P}}^- = \frac{1}{2N} \sum_{i=1}^{2N} (\mathbf{x}_i^{\sigma^*} - \tilde{\mathbf{x}}^-) (\mathbf{x}_i^{\sigma^*} - \tilde{\mathbf{x}}^-)^T. \quad (34)$$

Ehhez pedig hozzáadva a rendszerzajt (process noise, \mathbf{Q}), akkor megkapjuk az állapotvektor kovarianciájára a becslésünket:

$$\mathbf{P}^- = \tilde{\mathbf{P}}^- + \mathbf{Q}. \quad (35)$$

3. Létrehozunk újabb szigma-pontokat $\tilde{\mathbf{x}}^-$ várható értékkel és \mathbf{P}^- kovarianciával. Ezekre alkalmazzuk a kimeneti egyenletet, és az így kapott pontok (\mathbf{y}_i^σ) átlaga lesz a kimenetre végzett becslés ($\tilde{\mathbf{y}}$). Meghatározzuk ennek a vektornak a kovarianciáját (\mathbf{P}_{yy}) és az $\tilde{\mathbf{x}}$ -el vett keresztkorrelációs mátrixát (\mathbf{P}_{xy}):

$$\mathbf{P}_{yy} = \frac{1}{2N} \sum_{i=1}^{2N} (\mathbf{y}_i^\sigma - \tilde{\mathbf{y}}) (\mathbf{y}_i^\sigma - \tilde{\mathbf{y}})^T, \quad (36)$$

$$\mathbf{P}_{xy} = \frac{1}{2N} \sum_{i=1}^{2N} (\mathbf{x}_i^{\sigma^*} - \tilde{\mathbf{x}}^-) (\mathbf{y}_i^\sigma - \tilde{\mathbf{y}})^T. \quad (37)$$

4. Ezek alapján létrehozuk a korrekciós mátrixot:

$$\mathbf{K}_n = \mathbf{P}_{xy} \mathbf{P}_{yy}^{-1}, \quad (38)$$

amivel már ki tudjuk fejezni az állapotvektor és kovarianciájának a mérési eredményeket figyelembe vevő becslését:

$$\mathbf{x}_{n+1} = \tilde{\mathbf{x}}^+ = \mathbf{x}^- + \mathbf{K}_n (\mathbf{y}_n - \tilde{\mathbf{y}}), \quad (39)$$

$$\mathbf{P}_{n+1} = \mathbf{P}^+ = \mathbf{P}^- + \mathbf{K}_n (\mathbf{P}_{yy} + \mathbf{R}) \mathbf{K}_n^T. \quad (40)$$

3.3 Egynél több hangforrás követése

A gyakorlatban szükség lehet arra, hogy egyszerre egynél több hangforrást kövessünk. A Kálmán-szűrő a nyalábformáló algoritmusokkal együttműködve képes egy forrás követésére, de kibővíthető úgy, hogy kettő vagy több is követhető legyen, illetve úgy, hogy a források számát az algoritmus előre nem ismert bemeneti paraméterként kapja meg.

A Kálmán-szűrő ilyen formájú kibővítéséhez először azt kell megvizsgálni, hogy hogyan kapja meg bemenetként a vizsgált rendszer állapotáról készült mérési eredményeket. Az eddigiekhez hasonlóan, a mérési eredmények az adott nyálábformáló algoritmus (pl. MUSIC) által létrehozott kameraképből nyerhetők ki, most azonban egy abszolút maximum helyett több lokális maximumot kell keresni. Ez nem várt nehézségeket és hibalehetőségeket vonhat maga után, például:

- zajos kameraképen, jelentősen különböző erősségű csúcsok esetében, a nagyobb amplitúdójú csúcs közvetlen közelében is lehet találni lokális maximumokat, míg a valódi forrásirányt, ami a kisebb amplitúdójú csúcshoz tartozik, figyelmen kívül hagyjuk;
- ha két forrás iránya az akusztikai vásznon nagyon közel van egymáshoz, de különböző távolságra vannak, akkor az algoritmus nehezen tudja megkülönböztetni őket, és a közös csúcs mellett a kameraképen helytelen irányban talál egy második lokális maximumot;
- figyelmet kell fordítani arra is, hogy a használt maximumkereső algoritmustól függően egy adott mérési időpillanatban a különböző forrásokhoz tartozó mérési eredmények az előzőkhez képest más sorrendben is feljegyzésre kerülhetnek, ami úgy nyilvánul meg, hogy a rendszer becslései szerint a hangforrások egymással helyet cserélnek, akár gyakran is, miközben ez valójában nem történik meg.

Az első probléma ellen lehet védekezni a maximumkereső algoritmus helyes megválasztásával. Én a MATLAB Image Processing Toolbox „imregionalmax” nevű függvényét alkalmaztam, ami képes két-, vagy akár többdimenziós mátrixokban lokális maximumok keresésére. Jó alternatíva lehet a FastPeakFind képfeldolgozó algoritmus [37], ami jobban tud zajos képeket feldolgozni, viszont a lehető legnagyobb pontosság elérése érdekében érdemes bemeneti paraméterként megadni a csúcsok várt nagyságát és szórását.

A második probléma ellen nehéz a védekezés korlátozott felbontású amplitúdótérképen, viszont a harmadik probléma orvosolható. Miután minden egyes időpillanatban a kamerakép alapján becslést végeztünk a források pozíciójára, leellenőrizhetjük, hogy más sorrendben lettek-e feljegyezve ez előző időpillanathoz képest (például az aktuális és az előző időpillanathoz tartozó pozícióvektorok

különbsége alapján). Ha igen, akkor a helyvektor egy-egy forráshoz tartozó koordinátacsoportjainak a sorrendjét megfelelően fel kell cserélni. Ez a módszer két hangforrás esetén viszonylag egyszerűen implementálható, de három vagy több forrásnál a lehetséges sorrendek száma faktoriális szerint növekszik, így a gyakorlatban egyre nehezebben valósítható meg.

A másik fontos kérdés a Kálmán-szűrő kibővítése során, hogy milyen módon kell megváltoztatni az algoritmus működése során felhasznált vektorokat és mátrixokat. Az állapotvektor eddig egy hatelemű vektor volt, n darab követendő forrás esetén $n*6$ eleme van, az egy forráshoz tartozó pozíció- és sebességinformáció egymással szomszédos mezőkön található meg. A pozíció továbbra is a mérési eredményekből nyerhető ki, a sebességre pedig a szűrő ad becslést. Az algoritmust megvalósító osztályból létrehozott objektumnak bemeneti paraméterként meg kell adni a P_0 , Q és R_0 mátrixokat (lásd 5.1.1 szakasz), ahol a „0” alsó index a mátrixok kezdeti értékét jelenti. Ezeket a mátrixokat a következőképpen kell kibővíteni:

$$P'_0 = \begin{bmatrix} P_0 & \mathbf{0}_{6 \times 6} & \cdots \\ \mathbf{0}_{6 \times 6} & P_0 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (41)$$

$$Q' = \begin{bmatrix} Q & \mathbf{0}_{6 \times 6} & \cdots \\ \mathbf{0}_{6 \times 6} & Q & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (42)$$

$$R'_0 = \begin{bmatrix} R_0 & \mathbf{0}_{3 \times 3} & \cdots \\ \mathbf{0}_{3 \times 3} & R_0 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (43)$$

így n darab hangforrás követésekor a P'_0 és Q' mátrixok $6n*6n$, az R'_0 mátrix pedig $3n*3n$ méretű lesz.

Az osztály implementálásában módosítani kell az állapotegyenlet A mátrixát is. Ez egy forrás esetén a következő formát veszi fel:

$$A = \begin{bmatrix} 1 & 0 & 0 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 & T_s & 0 \\ 0 & 0 & 1 & 0 & 0 & T_s \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (44)$$

ahol T_s az időlépés hossza diszkrét időben. Több forrás esetében a mátrix felépítése:

$$A' = \begin{bmatrix} A & \mathbf{0}_{6 \times 6} & \dots \\ \mathbf{0}_{6 \times 6} & A & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (45)$$

n db forrás követéséhez egy $6n*6n$ -es mátrixra van szükség.

A mérési eredmények és a Kálmán-szűrő ilyen formájú kibővítése után a rendszer már képes lesz egyszerre egynél több hangforrást is követni.

4 Szimulációk

A fent részletezett nyálábformálási módszerek összehasonlíthatók szimulációk elvégzésével. MATLAB környezetben elvégezhetők olyan számítások, amik akár kétdimenziós, akár háromdimenziós térbeli mérési elrendezéseket szimulálnak.

4.1 A szimulációs keretrendszer

A munkám során felhasznált MATLAB környezetben megírt szimulációs keretrendszer objektum-orientált, különböző osztályokból létrehozott elemek segítségével lehet a szükséges számításokat elvégezni. A keretrendszer néhány fontosabb, általam is felhasznált osztálya [23]:

- Environment: az akusztikus környezetet valósítja meg (amely általában levegő), annak olyan tulajdonságait tárolja, mint például a légnyomást, a hőmérsékletet és a relatív páratartalmat, illetve ki tudja számítani a hangsebességet az adott közegben;
- Source: a hangforrás osztálya, amelyet egy trajektória és egy jel ír le;
- SoundField: egy Environment és egy Source elemből áll, képes kiszámolni egy adott pozícióban az időfüggő hangnyomást;
- MicrophoneArray: a mikroföntömböt megvalósító osztály, a mikrofonok koordinátáit tárolja;
- Canvas: a vászon pontjainak koordinátáit tároló osztály;
- Propagation: az akusztikus környezet, a vászon és a mikroföntömb alapján képes kiszámítani az érzékelési mátrixot;
- Mixdown: a vett jel alapsávba való lekeverését végzi el;
- CsmEstimator: a spektrális keresztkorrelációs mátrixot számítja ki;
- ConventionalBeamformer: a hagyományos nyálábformálást elvégző osztály, egy Propagation és egy CsmEstimator elemet kell megadni neki;

- Music: a MUSIC iránybecslést elvégző osztály, egy-egy Propagation és CsmEstimator elemeken kívül még a források előre megbecsült számát kell megadni;
- CompressedBeamformer: a Compressed Beamforming algoritmust elvégző osztály, egy Propagation és egy MixDown elemet kell megadni neki;
- Kalman: a Kálmán-szűrőt megvalósító osztály; a konstruktorában az időablak hosszát, az állapotvektor kovarianciamátrixának kezdeti értékét, a rendszerzaj kovarianciamátrixát, a kimenetet terhelő zaj kovarianciamátrixát és a blokkszámot kell megadni. Ennek kiterjesztései a Kalman2, illetve a Kalman_n osztályok, amikkel több forrás követését is el lehet végezni.

A nyálábformáló algoritmusokban a mikrofonok által vett jelnek egy pillanatban egy adott frekvenciájú komponensét használjuk fel (2.1.3 alfejezet). A spektrumot egy adott pillanatban úgy közelíthetjük, hogy a jelnek egy rövid szakaszán végzünk Fourier-transzformációt, ennek a módszernek azonban korlátozott a pontossága, ugyanis az időablak hossza és a kapott spektrum felbontása fordítottan arányos:

$$\Delta f = \frac{1}{T}, \quad (46)$$

itt a T az időablak hossza (ez a mostani szimulációk során 0,2 másodperc), Δf pedig a maximális frekvenciafelbontás. Ez azt jelenti, hogy minél rövidebb időtartamra akarunk becslést végezni, minél inkább pillanatszerűen akarjuk megkapni a spektrumot, annál alacsonyabb lesz a felbontása. Lehetőségünk van arra, hogy meghatározzuk pontosabban a jel egy adott frekvenciájú komponensét, anélkül, hogy megnöveljük az időablak méretét. Ezt a feladatot a MixDown osztály hajtja végre a következő lépéseken keresztül [38]:

1. A kiválasztott f_0 frekvenciát felhasználva az eredeti jelünket egyoldali frekvenciamodulációval eltoljuk balra:

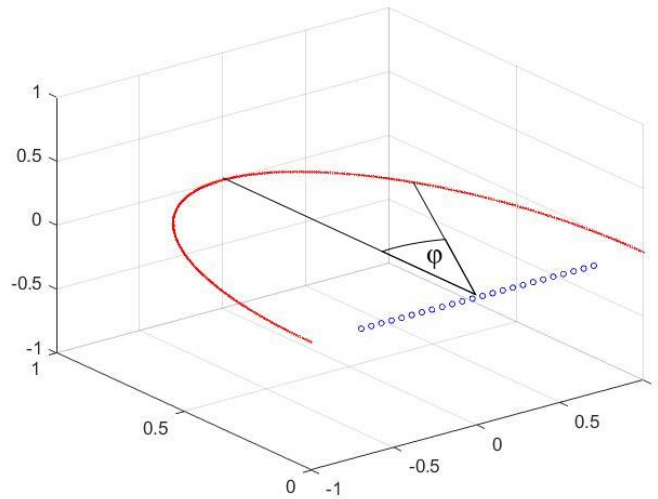
$$x_{mod}(t) = x(t)e^{-j2\pi f_0 t}. \quad (47)$$

2. A modulált jelet egy aluláteresztő szűrővel szűrjük, aminek a vágási frekvenciája jóval kisebb, mint a választott frekvencia.

3. A modulált és szűrt jelet időtartományban dolgozzuk fel. A jel amplitúdója a szűrő beállása után közel állandó, és ezt az értéket vesszük az adott frekvenciájú komponens amplitúdójának.

4.2 Szimulációk két dimenzióban

Kétdimenziós szimulációk esetén 24 darab mikrofont használunk vonal elrendezésben, két szomszédos mikrofon távolsága 6,1 centiméter. A hangsebességet 343,4 m/s-nak vesszük. Ennek megfelelően a felső határfrekvencia az (1) formula alapján 2821,16 Hz. Érdekes a nyalábformálás eredményét ennél lényegesen kisebb, lényegesen nagyobb, illetve nagyjából ekkora frekvencián is megvizsgálni és ez alapján is értékelni a különböző algoritmusok használhatóságát különböző frekvenciákon. Vizsgálati frekvenciáknak 500, 2500 és 5000 Hz-et választottam. A vászon 361 pontból áll, ezek egyenletesen helyezkednek el egy 100 méter sugarú félkörön, aminek a középpontjában van a mikrofontömb. A vászon által behatárolt szögtartomány $[-90^\circ, +90^\circ]$. A térben négy pontszerű, stacionárius forrás helyezkedik el 150 méter távolságban és -45° , -30° , 0° és $+30^\circ$ irányokban (a negyedik forrás 5000 Hz-es vizsgálati frekvencia esetében szükség szerint át lett helyezve $+60^\circ$ -ra, hogy bizonyos jelenségek jobban vizsgálhatók legyenek). A források fehérzajt bocsátanak ki (a mintavételi frekvencia 48 kHz), szórásuk rendre 3, 1, 2 és 1 egységnyi (azaz teljesítményük 9, 1, 4 és 1 egységnyi). Mindkét nyalábformáló algoritmus működését vizsgálom zajmentes és zajos esetben is, utóbbinál 20 dB-es jel-zaj viszonyal. Az eredményüket egy véletlenszerűen kiválasztott pillanatban (0,2 másodperces időtartamra vonatkoztatva) ábrázolom. Az ábrákon a valódi forrásirányokat kézzel, az algoritmusok becslését pirossal jelölöm, ezek a maximumukhoz viszonyított relatív értékeként értelmezendők.



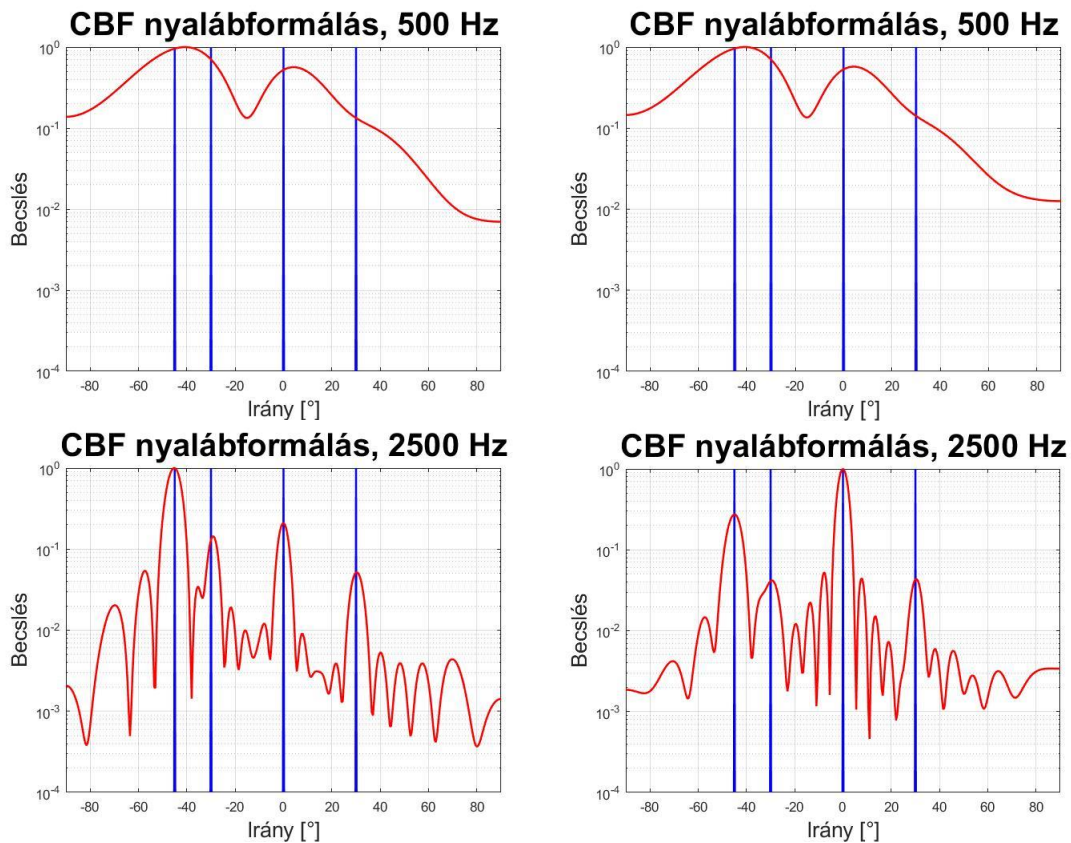
a

6. ábra: Az irányok értelmezése két dimenzióban.

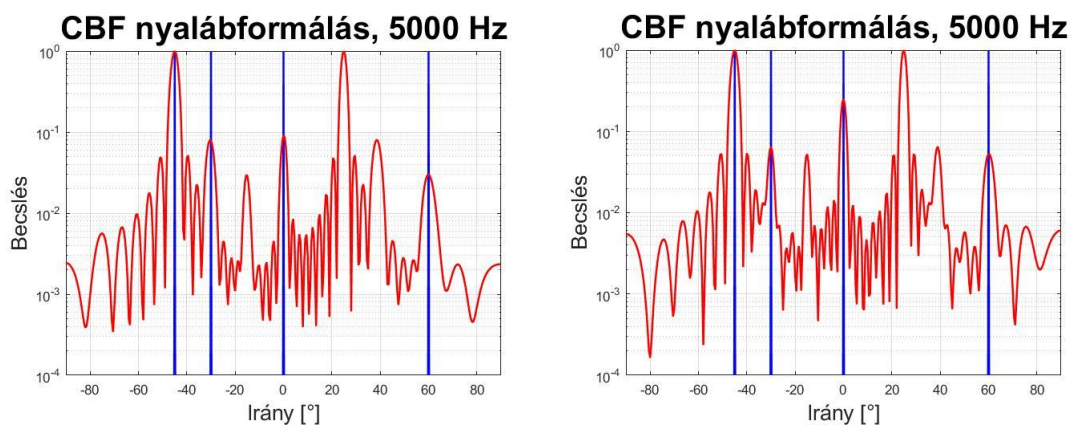
4.2.1 Conventional Beamforming

A 7. ábra mutatja a hagyományos nyálábformálás eredményét, késsel a valóságos forrásirányok, pirossal pedig a becslések eredménye van ábrázolva. 500 Hz-es vizsgálati frekvencián a módszer nem alkalmas a források irányának becslésére. 2500 Hz-en zajtalan esetben az első három forráshoz tartozó főnyalábok még kiemelkednek a melléknyalábok közül, de a negyedik már nagyjából a legnagyobb melléknyalábok szintjén van, ráadásul a második és a negyedik forrás irányának a becslése kicsit pontatlan. Zajos esetben már csak két főnyaláb emelkedik ki a melléknyalábok közül.

5000 Hz-en már egyáltalán nem használható ez a módszer jó jel-zaj viszonyban sem, ugyanis a térbeli átlapolódás miatt megjelenik 20-30 fok környékén egy szellemforrás, ami a legnagyobb főnyalábbal van egy szinten, emiatt nem lehet helyesen meghatározni a források irányát (8. ábra). Ez a jelenség feltűnik a többi módszernél két és három dimenzióban is, ezért azoknak az eredményét nem ábrázolom külön 5000 Hz-es vizsgálati frekvencia mellett.



7. ábra: CBF nyalábformálás zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.

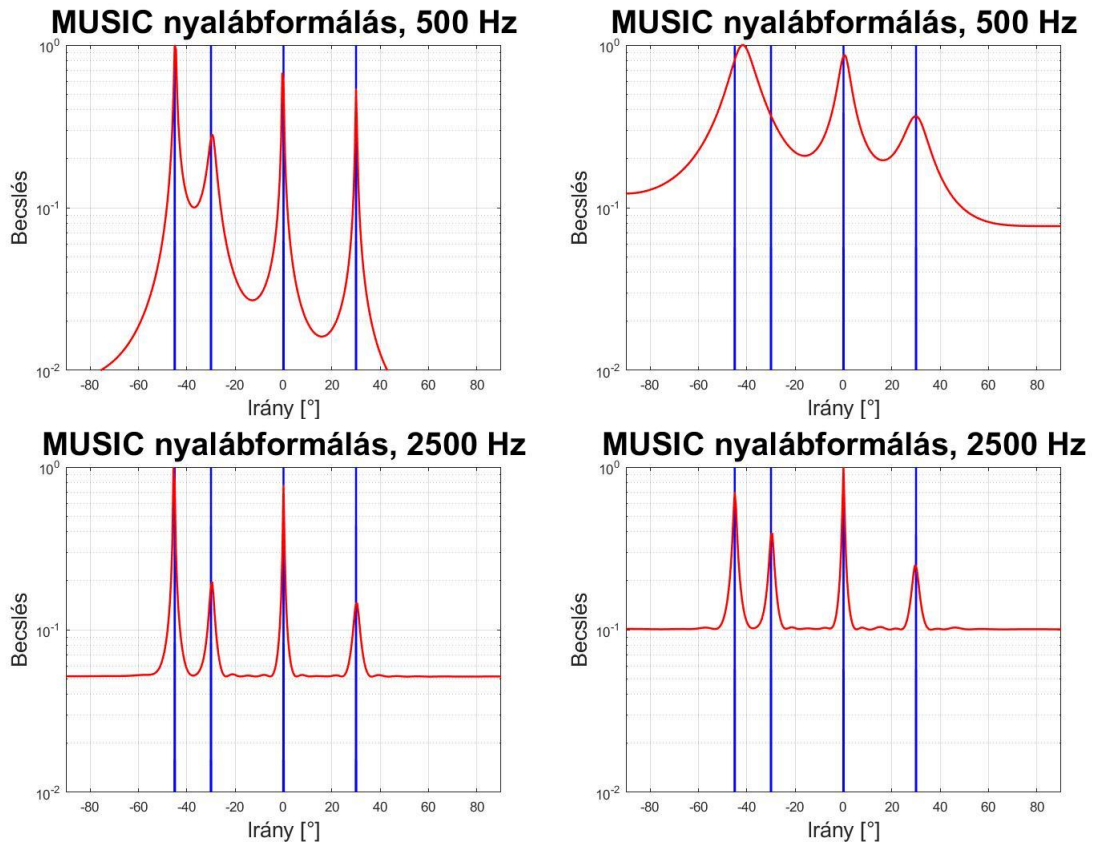


8. ábra: CBF nyalábformálás zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 5000 Hz vizsgálati frekvencián.

4.2.2 Multiple Signal Classification

A 9. ábra mutatja a MUSIC algoritmus alapú iránymeghatározás eredményét. A CBF-fel ellentétben 500 Hz-en kicsit pontatlan, de használható becslést kapunk, csak a zaj bevezetésekor nem sikerül mind a négy forrást elkülöníteni egymástól. 2500 Hz-es

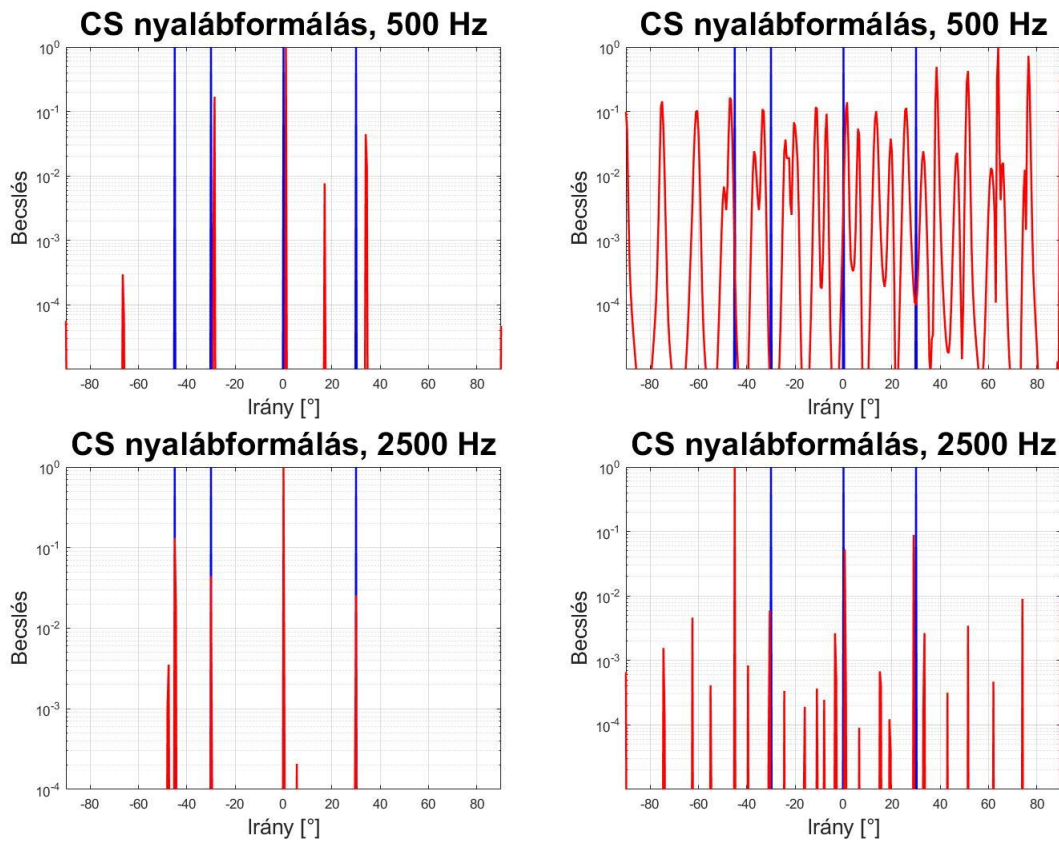
vizsgálati frekvencián még zajjal együtt is viszonylag nagy pontosságú megbecsülhetjük eredményt kapunk.



9. ábra: MUSIC nyálábformálás zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.

4.2.3 Compressive Beamforming

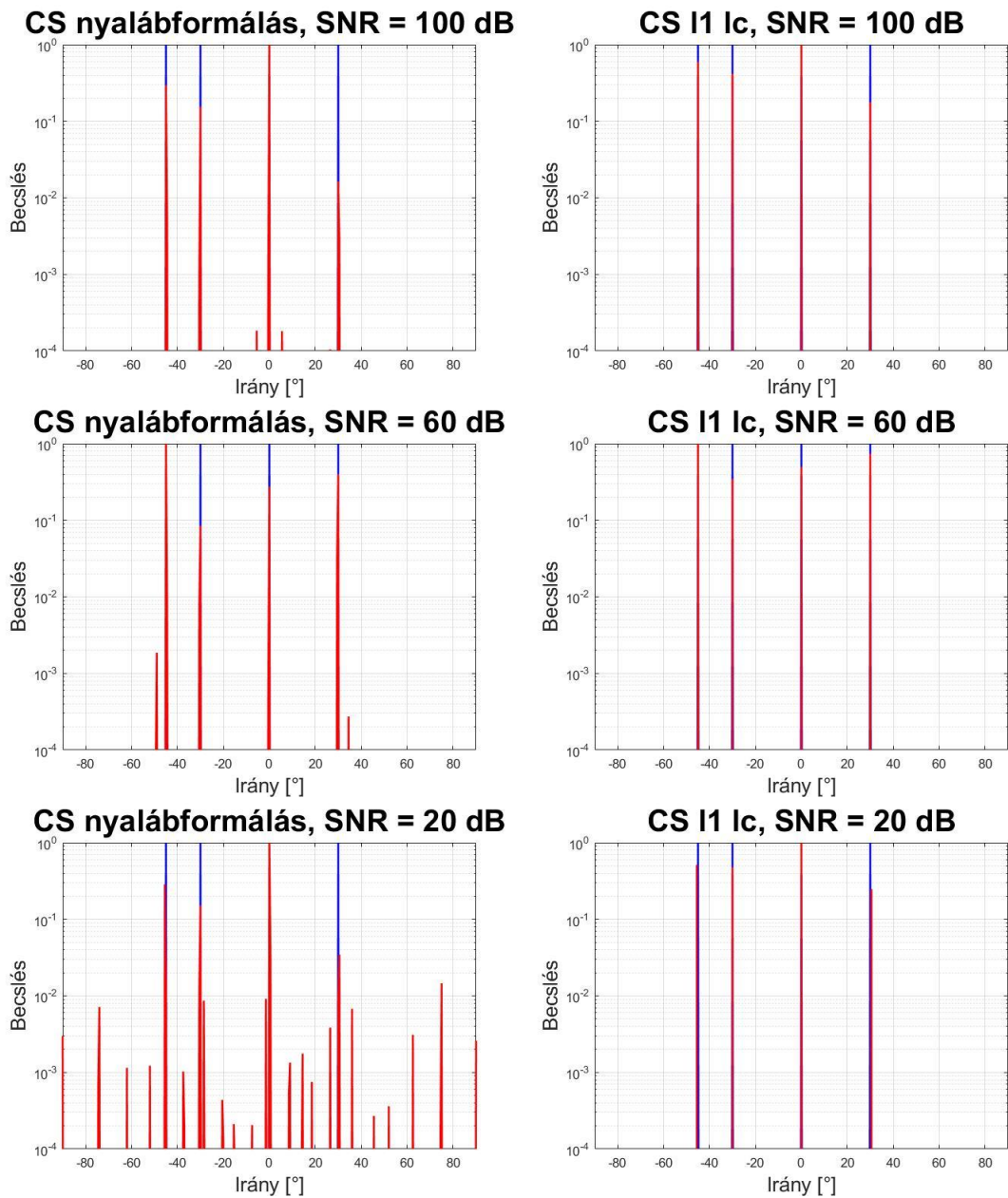
A Compressive Beamforming módszer eredménye alább látható (10. ábra). 500 Hz-en az algoritmus nem tud helyes becslést adni, csak 2500 Hz-en. Megfelelő vizsgálati frekvencia esetén is a zajtűrése jóval kisebb, mint a MUSIC-nak. Ha viszont elég jó a jel-zaj viszony, a Compressive Beamforming nagy pontosságú és ritkaságú megoldást képes adni.



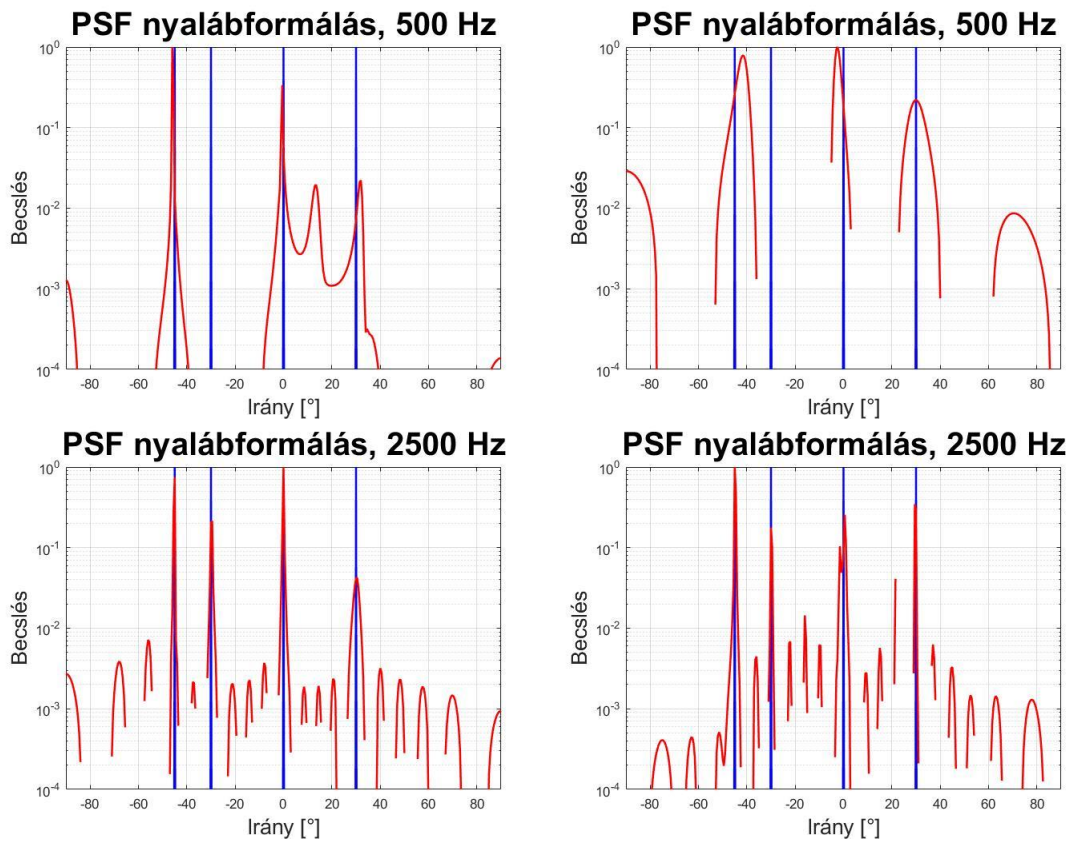
10. ábra: CS iránybecslés zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.

A Compressive Beamforming-ot kiegészítve l_1 minimalizálással és szintkorrekcióval sokkal jobb zajtűrést és tisztább képet kapunk (11. ábra). Az algoritmus még 20 dB jel-zaj viszony mellett is viszonylag pontosan megtalálja a forrásokat az esetek többségében.

A pontszórás-függvény alapú Compressed Beamforming módszert használva az eredetihez hasonló eredményhez jutunk (12. ábra). Lényeges különbség azonban, hogy 2500 Hz-en ennek a módszernek jóval nagyobb a zajtűrése, mint a hagyományos Compressed Beamforming-nak.



11. ábra: CS iránybecslés (bal) és kiegészítésekkel (jobb), 100 dB (fent), 60 dB (középen) és 20 dB (lent) jel-zaj viszony mellett, 2500 Hz vizsgálati frekvencián.



12. ábra: CS iránybecslés zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.

4.2.4 Mozgó hangforrás követése Kálmán szűrővel

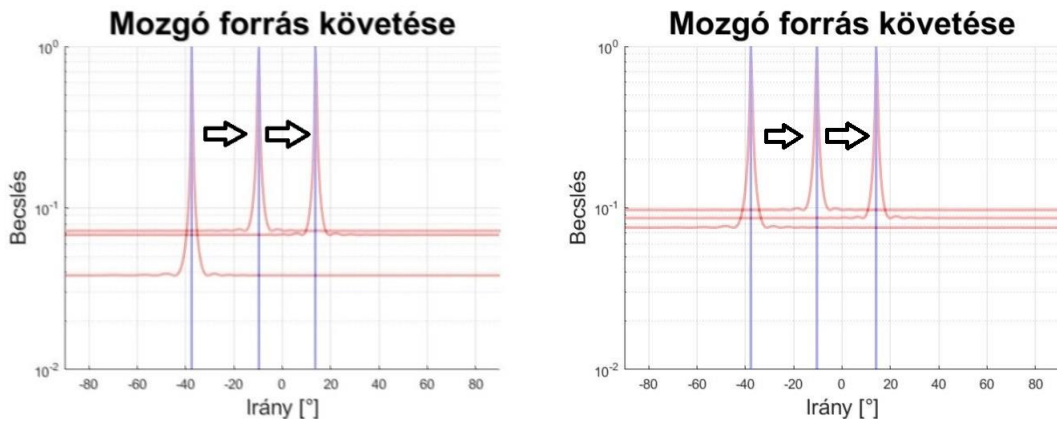
Az eddigi szimulációk során stacionárius fehérzaj forrásokról készült kameraképeket vizsgáltunk. A hangforráshoz hozzárendelve egy trajektóriát, ami mentén mozog, tesztelhetjük a Kálmán-szűrő működését néhány egyszerű szituáción keresztül.

A 13. ábra mutatja egy olyan szimuláció eredményét, amelynek során egy egyenletesen mozgó hangforrás halad el a mikroföntömb előtt. A hangforrás 3 egységnyi szórású fehérzajt bocsát ki, a mikrofonok vonalával párhuzamos egyenes mentén mozog egyenletesen: az $[x, y, z] = [-150, 150, 0]$ koordinátájú pontból indul és az x tengellyel párhuzamosan 50 m/s-os sebességgel mozog. A Kálmán-szűrő mérési eredményként a MUSIC algoritmus maximumhelyeit használja fel. Bal oldalon a zajmentes, jobb oldalon a zajjal terhelt (SNR = 20 dB) szimuláció alatt létrejött kameraképeket láthatjuk, három különböző időpontban. A Kálmán-szűrő paraméterei az 5.1.1 szakaszban részletezett kiindulási paraméterek, a blokkhossz 0,02 s. Az 1. Táblázat látható, hogy a Kálmán-szűrő az egyenletesen mozgó hangforrást jól tudja

követni, az általa becsült beesési szög az esetek nagy részében megegyezik a MUSIC-ével, háttérzaj mellett is.

Időpont (s)	Zajmentes		Zajos	
	MUSIC	Kálmán	MUSIC	Kálmán
1,32	-37,5°	-37,5°	-37,5°	-37,5°
2,96	-10°	-10°	-10,5°	-10,5°
4,28	14°	14°	14°	14°

1. Táblázat: A MUSIC algoritmus és a Kálmán-szűrő által becsült irányok.

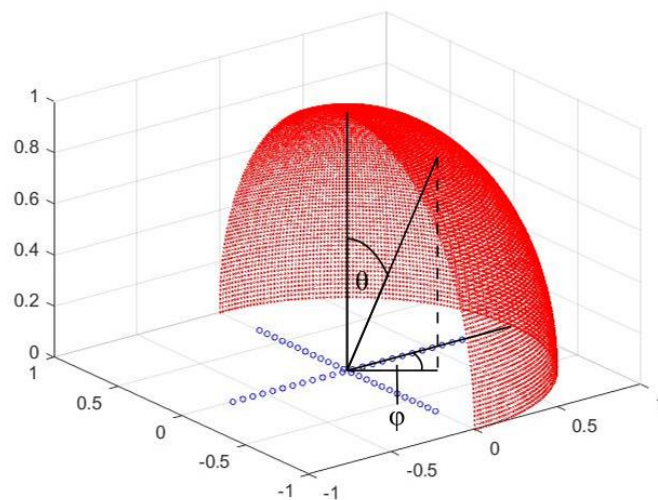


13. ábra: Mozgó fehérzajforrás követése két dimenzióban, bal oldalon zajmentes, jobb oldalon zajos esetben (piros: MUSIC algoritmus eredménye; kék: a Kálmán-szűrő által becsült irány).

4.3 Szimulációk három dimenzióban

Háromdimenziós szimulációk esetén összesen 48 darab mikrofont használunk kereszt elrendezésben (azaz két egymásra merőleges vonalban, mik egyenként 24 mikrofonból állnak), két szomszédos mikrofon távolsága ebben az elrendezésben 6 centiméter. A felső határfrekvencia ebben az esetben 2862,07 Hz. Vizsgálati frekvenciáknak itt is 500, 2500 és 5000 Hz-et választottam. A vászon 32851 pontból áll, ezek egy gömb felületén helyezkednek el, aminek a középpontjában van a mikrofontömb és a sugara 100 méter. A vászon pontjainak irányai két szöggel írhatók le, gömbi koordinátarendszerben, ahol θ a z tengellyel bezárt szög, φ pedig az irány x-y síkra vett vetületének az x tengellyel bezárt szöge. A vászon által behatárolt szögtartomány: $\varphi = [-90^\circ, +90^\circ]$ és $\theta = [0^\circ, +90^\circ]$. A térben négy pontszerű forrás helyezkedik el 150 méter távolságban és $(\varphi, \theta) = (-30^\circ, 30^\circ), (30^\circ, 30^\circ), (30^\circ, 60^\circ), (-30^\circ,$

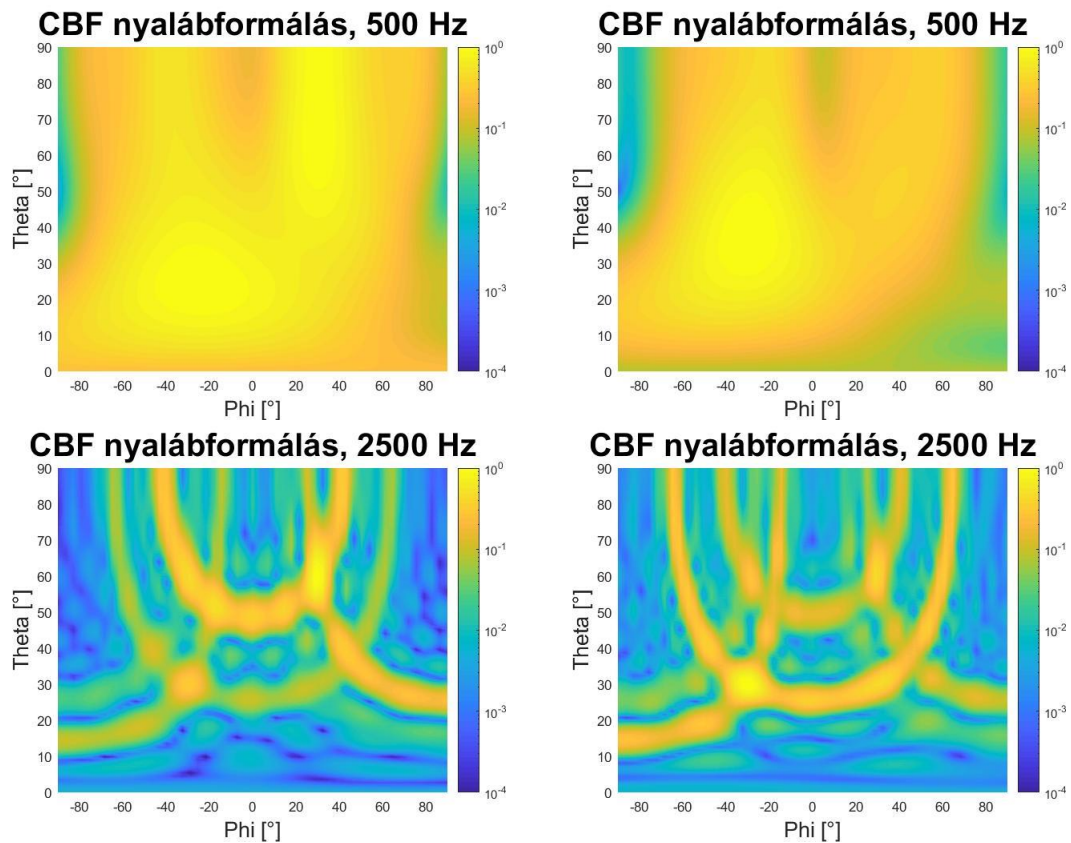
60°) irányokban. A források fehérzajt bocsátanak ki (a mintavételi frekvencia 48 kHz), szórásuk rendre 3, 1, 2 és 1 egységnyi. Mindkét nyalábformáló algoritmus működését vizsgálom zajmentes és zajos esetben is, utóbbinál 20 dB-es jel-zaj viszonyal. Az eredményüket egy véletlenszerűen kiválasztott pillanatban (0,2 másodperces időtartamra vonatkoztatva) ábrázolom.



14. ábra: Az irányok értelmezése három dimenzióban.

4.3.1 Conventional Beamforming

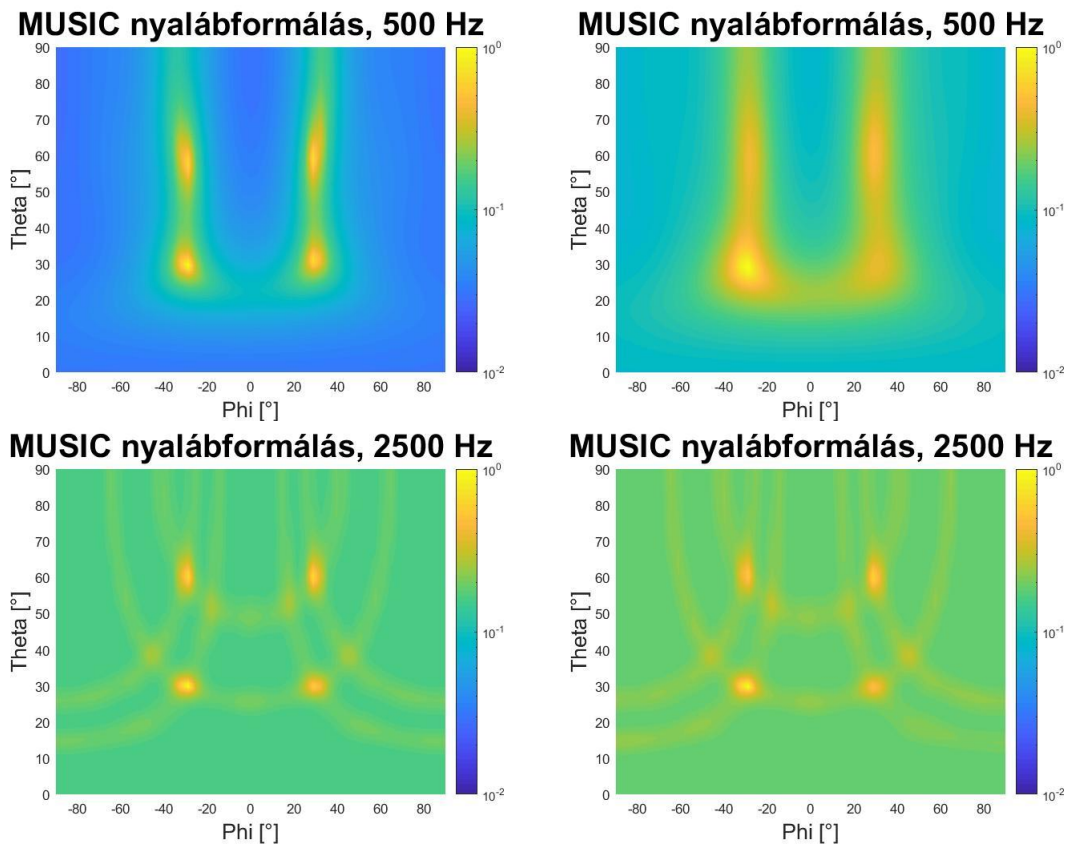
A 15. ábra mutatja a hagyományos nyalábformálás eredményét. A kétdimenziós szimulációhoz hasonlóan, 500 és 5000 Hz-en nem használható jól a módszer, ugyanis előbbinél az elmosódott kép, utóbbinál a térbeli átlapolódás révén létrejött szellemforrások miatt nem tudunk pontos iránybecslést végezni. 2500 Hz-en, zaj nélkül és zajjal is csak pontatlanul tudjuk megbecsülni az irányokat.



15. ábra: CBF nyalábformálás zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.

4.3.2 Multiple Signal Classification

A MUSIC algoritmus is hasonló eredményeket ad két és három dimenzióban (16. ábra). 500 Hz-es vizsgálati frekvencián, kellően jó jel-zaj viszony mellett jól használható a módszer. 2500 Hz-en, a határfrekvencia közelében viszonylag jó a zajtűrése, 20 dB-es jel-zaj viszonyal együtt is képes meglehetősen pontos iránymeghatározásra. Amikor a MUSIC és a CBF is használható, a MUSIC pontosabb és nagyobb bizonyosságú becslést tud adni.



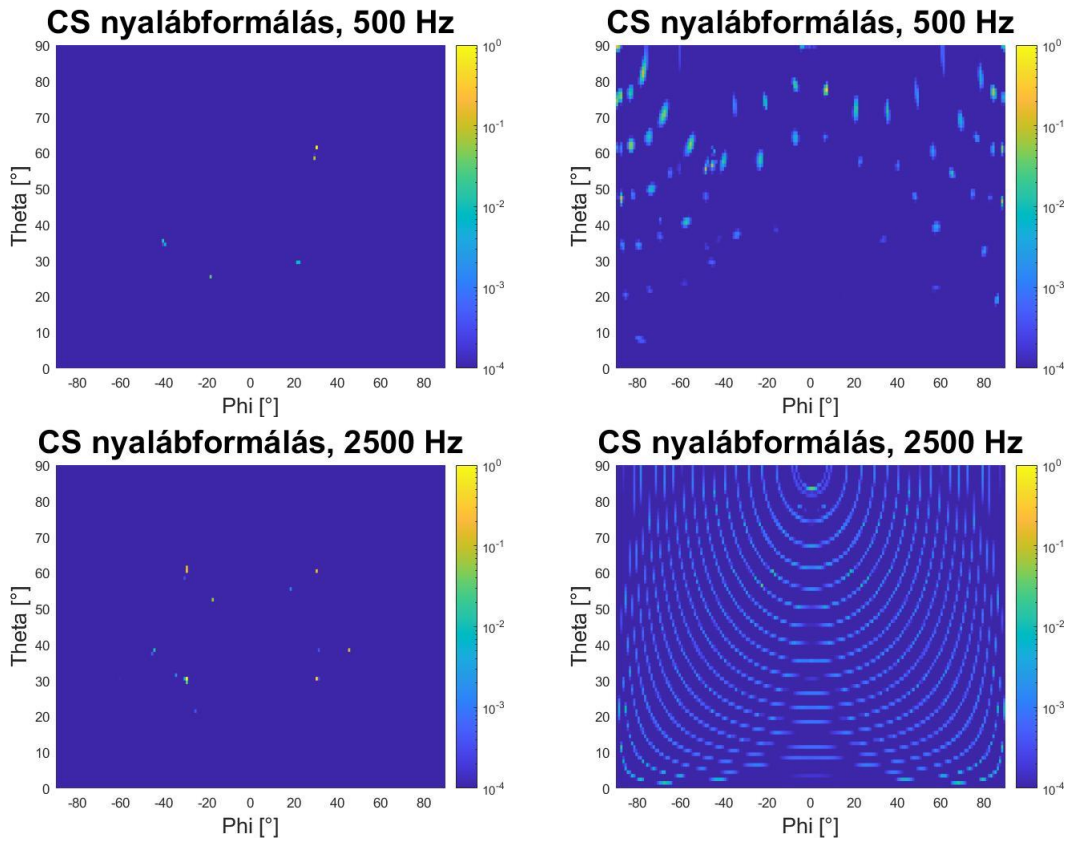
16. ábra: MUSIC nyálábformálás zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.

4.3.3 Compressive Beamforming

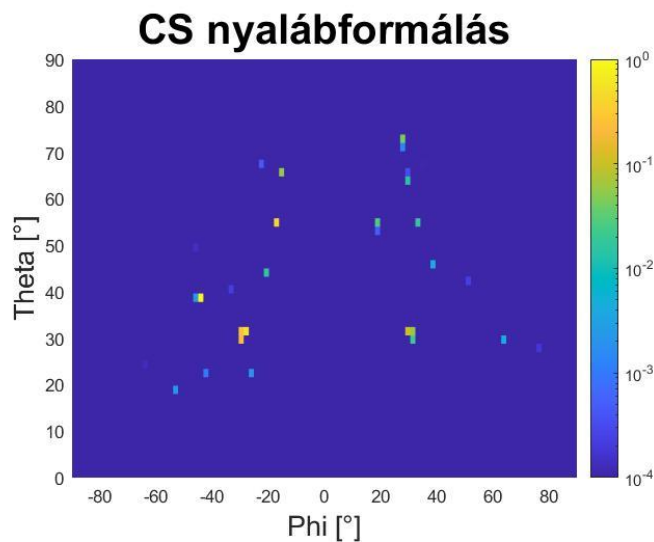
Bár a szakirodalomban még nincsenek háromdimenziós szimulációs eredmények a Compressive Beamforming-ről, sok a hasonlóság a kétdimenziós esettel. Három dimenzióban is csak 2500 Hz-en, kedvező jel-zaj viszony mellett vagy zaj nélkül tud jól használható kameraképet eredményezni (17. ábra). Figyelni kell továbbá arra is (kettő és három dimenzióban egyaránt), hogy a vászon pontjai/irányai kellően sűrűn helyezkedjenek el. Ha a források irányai nem illeszkednek jól a vászonra (ezt a problémát angolul „basis mismatch”-nek nevezik [39]), akkor előfordulhat, hogy a kapott eredmény helytelen (18. ábra). Ez a jelenség csak a Compressive Beamforming-ot érinti, a CBF-et és a MUSIC-ot nem.

L_1 minimalizálás és szintkorrekció bevezetésével (19. ábra) ismét egy jobb zajtűréssel rendelkező eljárást kapunk. A kétdimenziós szimulációkkal szemben azonban 20 dB jel-zaj viszony esetén a MATLAB cvx toolbox-a nem tudta megoldani a megadott konvex optimalizálási problémát. A ritkasági kényszerfeltétel ugyan teljesül,

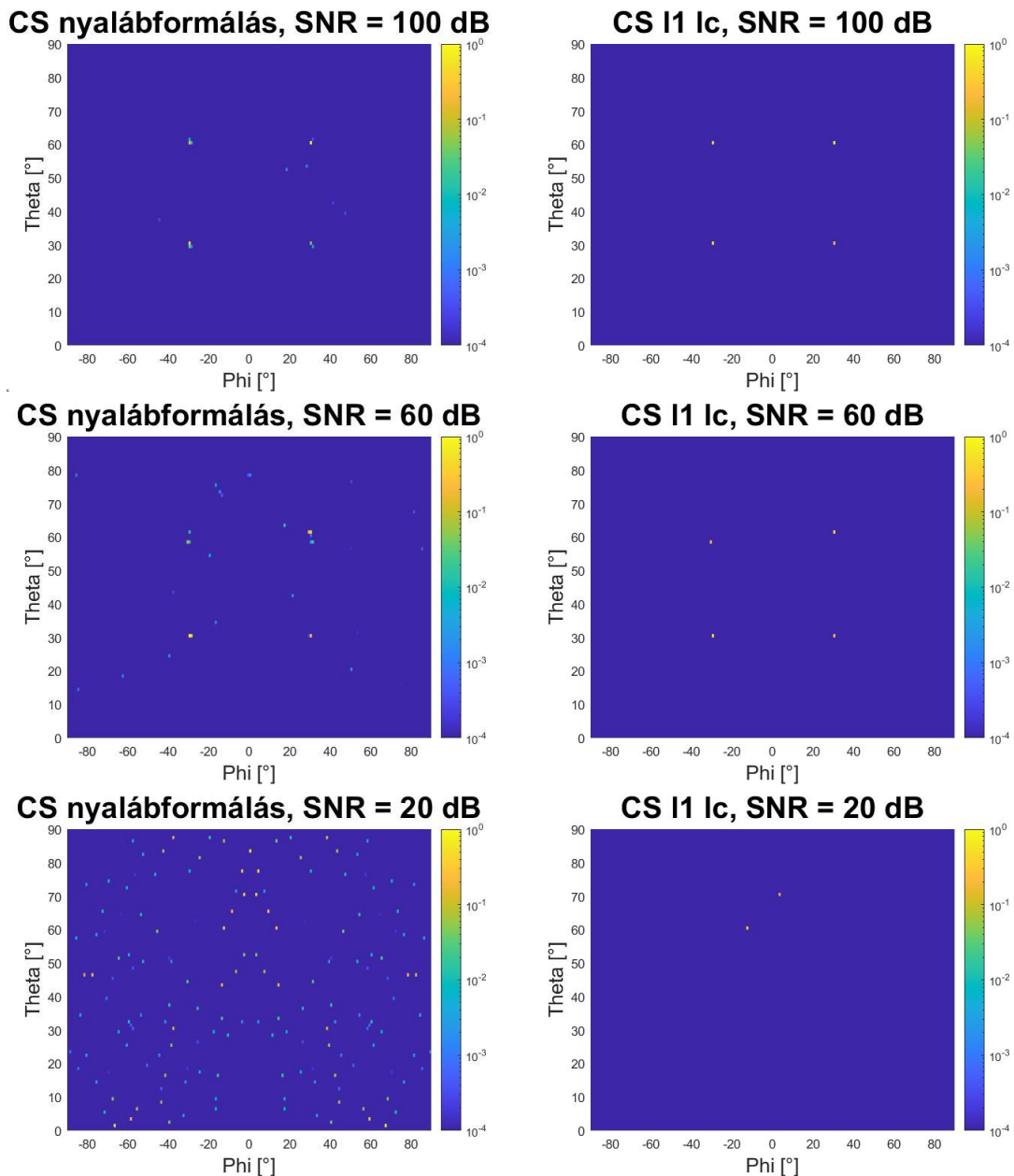
de nagyon sok zérustól különböző eleme van a megoldásnak, ezekből pedig nem tudjuk eldönteni, hogy hol van valójában a forrás.



17. ábra: CS iránybecslés zajmentes (bal) és 20 dB jel-zaj viszonyú (jobb) esetben, 500 Hz (fent) és 2500 Hz (lent) vizsgálati frekvencián.



18. ábra: A „basis mismatch” miatti hibás kamerakép, 2500 Hz vizsgálati frekvencián, zajmentes esetben, 101-szer 51-es méretű vászonnal.



19. ábra: CS iránybecslés (bal) és kiegészítésekkel (jobb), 100 dB (fent), 60 dB (középen) és 20 dB (lent) jel-zaj viszony mellett, 2500 Hz vizsgálati frekvencián.

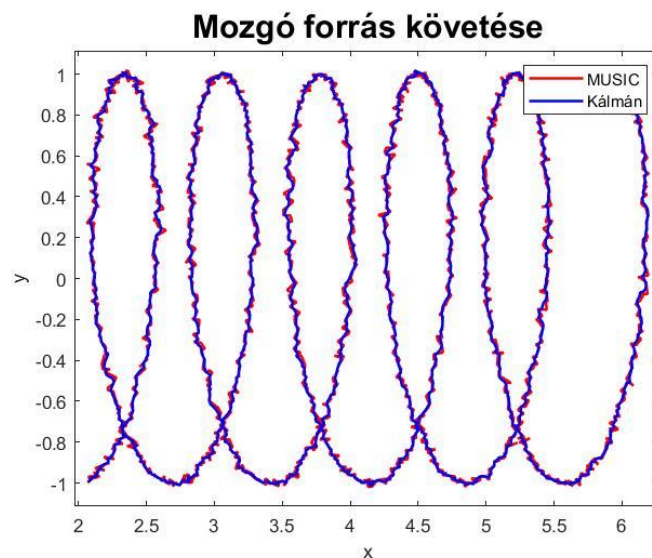
A pontszórás-függvény alapú Compressive Beamforming három dimenzióban a PSF-mátrix nagy mérete miatt igencsak számításigényes (egy 0,2 másodperces időablakra végezve a becslést akár több óráig is eltarthat az iteráció), a többi módszernél sokkal lassabb, valós időben nem lehet alkalmazni.

4.3.4 Mozgó hangforrás követése Kálmán-szűrővel

A Kálmán-szűrő működését három dimenzióban az eddigiekhez képest más jellegű szimulációval mutatom meg. A MUSIC algoritmus által alkotott teljes

kameraképek nincsenek generálva, csak az azokon talált maximumhelyek, mégpedig úgy, hogy az x és az y koordináták egy-egy adott érték körül normális eloszlás segítségével generálódnak (a z koordináta konstans, ebben a szimulációban 5 egység). Ezzel voltaképpen egy olyan hangforrást szimulálunk, amely zajos környezetben helyezkedik el, és ezért a MUSIC algoritmus becslései a várható érték (a forrás valódi pozíciója/iránya) körül ingadoznak. A Kálmán-szűrő paraméterei az 5.1.1 szakaszban részletezett kiindulási paraméterek, a blokkhossz 0,01 s; mérési eredményként a MUSIC maximumhelyeit használja fel.

A 20. ábra mutatja a MUSIC algoritmus maximumhelyei (piros) és a Kálmán-szűrő becslései (kék) által bejárt pályát egy 20 másodperc alatt bejárt szakaszon, mindkét pálya a $z=5$ síkban helyezkedik el. A forrás által kibocsátott hangjelre zaj rakódik, ezért a mérési eredmények a valódi, spirális jellegű pálya körül ingadoznak. A MATLAB-ban elvégzett számítások igazolják azt, ami a kép alapján sejthető: az utóbbi kisebb utat jár be. Ebből arra következtethetünk, hogy spirális (vagy más, gyakorlatban előforduló) trajektórián haladó, zajos környezetben lévő hangforrás vizsgálatokor a MUSIC algoritmus becsléseit a Kálmán-szűrő „kisimítja”, stabilizálja, ilyen mérési szituációkban mindenképpen érdemes használni.

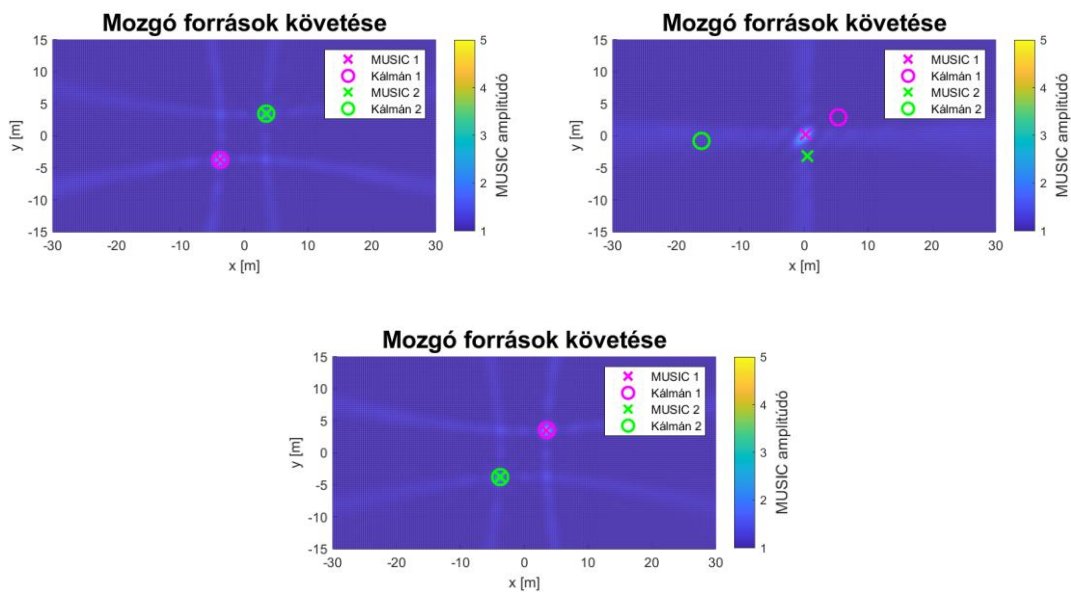


20. ábra: Mozgó hangforrás követése Kálmán-szűrővel, három dimenzióban (piros: MUSIC algoritmus eredménye; kék: a Kálmán-szűrő által becsült irány).

4.3.5 Két hangforrás irány- és távolságbecslése

A következő szimulációban kettő hangforrás egyidejű követését végeztem el, az irányuk mellett a távolságukat is becsülve. A források kezdőpontjainak a koordinátái

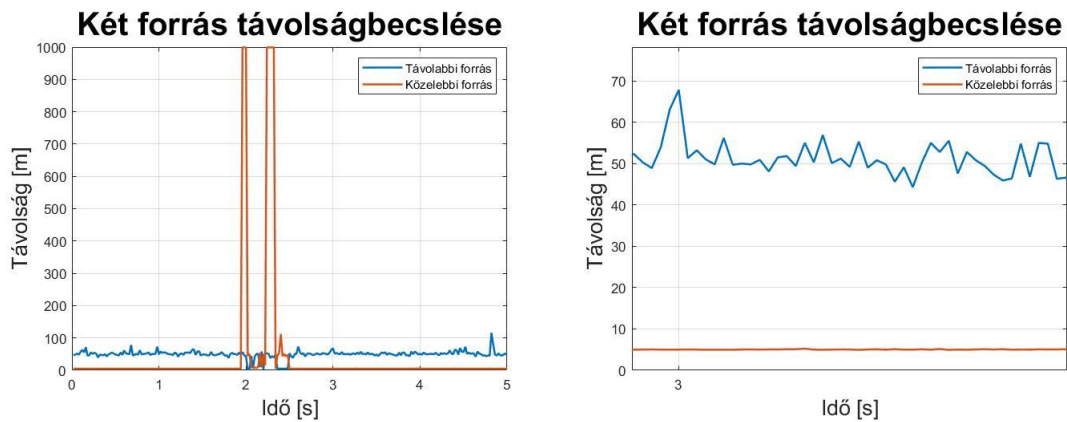
$(x;y;z) = (-20;-20;50)$ és $(2;2;5)$, sebességeik (amik állandók) rendre $(v_x,v_y,v_z) = (10;10;0)$ és $(-1;-1;0)$. Ez azt jelenti, hogy a mikrofontömb síkjától 50, illetve 5 méter távolságra helyezkednek el, majd azzal párhuzamos egyenes pályán mozognak; az irányuk egyre közelednek egymáshoz, egy pillanatig egymás mögött vannak a mikrofontömb középpontjából nézve, majd pedig eltávolodnak. Az általuk kibocsátott fehérzaj jelek szórása rendre 10 és 1 egységnyi. Az akusztikus vászon itt 200-szor 100 pontból áll, egy 60-szor 30 méteres téglalapon egyenletes elhelyezkedéssel, a mikrofontömb síkjától 15 méter távolságra, a közepe az $(x;y;z) = (0;0;15)$ pontban van. A jel-zaj viszony 20 dB. A kamerakép létrehozása a szimulációk alapján összességében legkedvezőbbnek bizonyult MUSIC algoritmussal történik. A távolságot a 2.5 szakaszban részletezett harmadik módszer szerint becsüljük, viszont itt érdemes a MUSIC algoritmusban bemeneti paraméterként megadott előre becsült forrásszámot a valódinál (kettő) nagyobbra állítani (például a kétszeresére) a pontosság növelése érdekében.



21. ábra: Kettő mozgó hangforrás követése MUSIC algoritmussal és Kálmán-szűrővel, három dimenzióban, távolságbecsléssel együtt.

A 21. ábrán látható három pillanatkép a szimulációról időrendben. Az első és a harmadik képen a MUSIC algoritmus és a Kálmán szűrő is megfelelően tudja követni a források irányát. A második kép azt a pillanatot ábrázolja, amikor a mikrofontömb középpontjához viszonyított irányuk nagyon közel vannak egymáshoz (a távolabbi forrás a közelebbi mögött van, ami őt „eltakarja”). A kamerakép alapján nem lehet

megkülönböztetni egymástól a két hangforrást, így a szűrőnek betáplált mérési eredmények is legfeljebb részben lehetnek helyesek, a forráskövetés ezen a szakaszon nem lehetséges ezzel az eljárással.



22. ábra: Kettő mozgó hangforrás követése és távolságuk becslése MUSIC algoritmus segítségével.

A 22. ábrán látható a két forrás becsült távolsága az idő függvényében, bal oldalt a teljes szimulációs időtartamra, jobb oldalt pedig egy rövid szakaszon kinagyítva, egy olyan szakaszon, ahol az irányok közelsége már nem okoz problémát. A közelebbi forrás távolságát az algoritmus nagyon pontosan meg tudja határozni (5 méter), míg a távolabbinál a helyes érték (50 méter) körül nagy ingadozás van. Ez megfelel az elvárásoknak, hiszen nagy távolságokból érkező hanghullámok hullámfrontja egyre jobban közelít a síkhullámhoz (ami a végtelen távolsághoz tartozik), így különböző távolságok között kisebb a különbség.

A problémás szakasz általában 30-35 blokkhosszig tart (ez nem minden futtatáskor ugyanannyi, a források által kibocsátott jel randomgenerált jellege miatt). Ez a használt 0,02 másodperces blokkhossz alapján 0,6-0,7 másodpercrek felel meg. Ennyi idő alatt a távolabbi hangforrás 8,5-10 métert tesz meg, ami 50 méter távolságra 9,7-11,4 fokos szöghkülönbséget jelent. Ezen a szögtartományon belül a következő fajta hibák lépnek fel leggyakrabban a kamerakép alapján becsült irányokra és távolságokra:

- mindkét forrás iránya rossz, a távolságuk felcserélődött;
- a távolabbi forrás iránya és távolsága helyes, a közelebbi forrás iránya rossz és távolsága végtelen;
- a távolabbi forrás iránya és távolsága helyes, a közelebbi forrás iránya rossz és távolsága a másiké;

- a távoli forrás iránya és távolsága rossz, a közeli forrás iránya rossz, távolsága végtelen;
- a távoli forrás iránya és távolság rossz, a közeli forrás iránya rossz, távolsága a másiké;
- a távoli forrás iránya rossz, távolsága a másiké, a közeli forrás iránya és távolsága rossz;
- a távoli forrás iránya és távolsága rossz, a közeli forrás iránya és távolsága rossz;
- a távoli forrás iránya és távolsága rossz, a közeli forrás iránya és távolsága helyes;
- időnként előfordulhat, hogy mindkét forrás iránya és távolsága is helyes, egymás után akár több időpillanatban is, mielőtt újra hibás lenne.

Összességében tehát elmondható, hogy a távolságbecslés olyankor hibás, amikor az iránybecslés is, viszont amikor az irány helyesen lett meghatározva, akkor a távolsággal sincsenek problémák.

5 Mérések, eredmények

Ebben a fejezetben a nyálábformáló algoritmusok közül a legelőnyösebb MUSIC algoritmus által feldolgozott mérési eredményeket mutatok be. A MUSIC algoritmus jobb minőségű kameraképet produkál, a hangforrás pozícióját/irányát képes kisebb bizonytalansággal meghatározni, mint a CBF, a CS-el szemben pedig a jóval kisebb számításigény az előnye. A MUSIC algoritmus eredményei Kálmán-szűrővel további feldolgozásra kerülnek, így újdonságként a két algoritmust mozgó hangforrás lokalizálására és követésére alkalmazom valós mérési eredményeken, szabadtéri környezetben.

5.1 Szabadtéri mérések

Munkám során lehetőségem adódott arra, hogy részt vegyek egy olyan mérésben, ahol pilóta nélküli légi járművek (másképpen drónok) röptetése során azok lokalizációja volt a cél, többféle módszerrel (pl. radar, optikai, akusztikai stb.). A mérést a korábban már részletezett 48 csatornás mikrontömbbel végeztük el. A vászon ebben az esetben egy téglalap mentén helyezkedik el, aminek az oldalai 20 és 10 méter hosszúak, és 200-szor 100 (összesen 20000) pontból áll. A vászon és a mikroföntömb síkjai párhuzamosak, távolságuk 5 méter, a közepeiket összekötő szakasz merőleges ezekre a síkokra.



23. ábra: A mérések ideje alatt reptetett drónok, balról jobbra: DJI Phantom4, Quad480, Eurotor4, Tarot680, Secopx8 és Flying Ranger.



24. ábra: A mérésekhez felhasznált 48 csatornás mikrofontömb, az állvány tetején elhelyezett webkamerával.

5.1.1 A Kálmán-szűrő paraméterfüggő működése

A szimulációs környezetben implementált „Kalman” osztályú objektum létrehozásakor öt bemeneti paramétert kell megadni: az időablak hosszát (T_s), az állapotvektor kovarianciamátrixának kezdeti értékét (P_0), a rendszerzaj kovarianciamátrixát (Q), a kimenetet terhelő zaj kovarianciamátrixát (R_0) és a teljes mérés blokkszámát kell megadni. Kiindulási alapként ezek a paraméterek a következő értékeket veszik fel [23]:

$$P_0 = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 \end{bmatrix}, \quad (48)$$

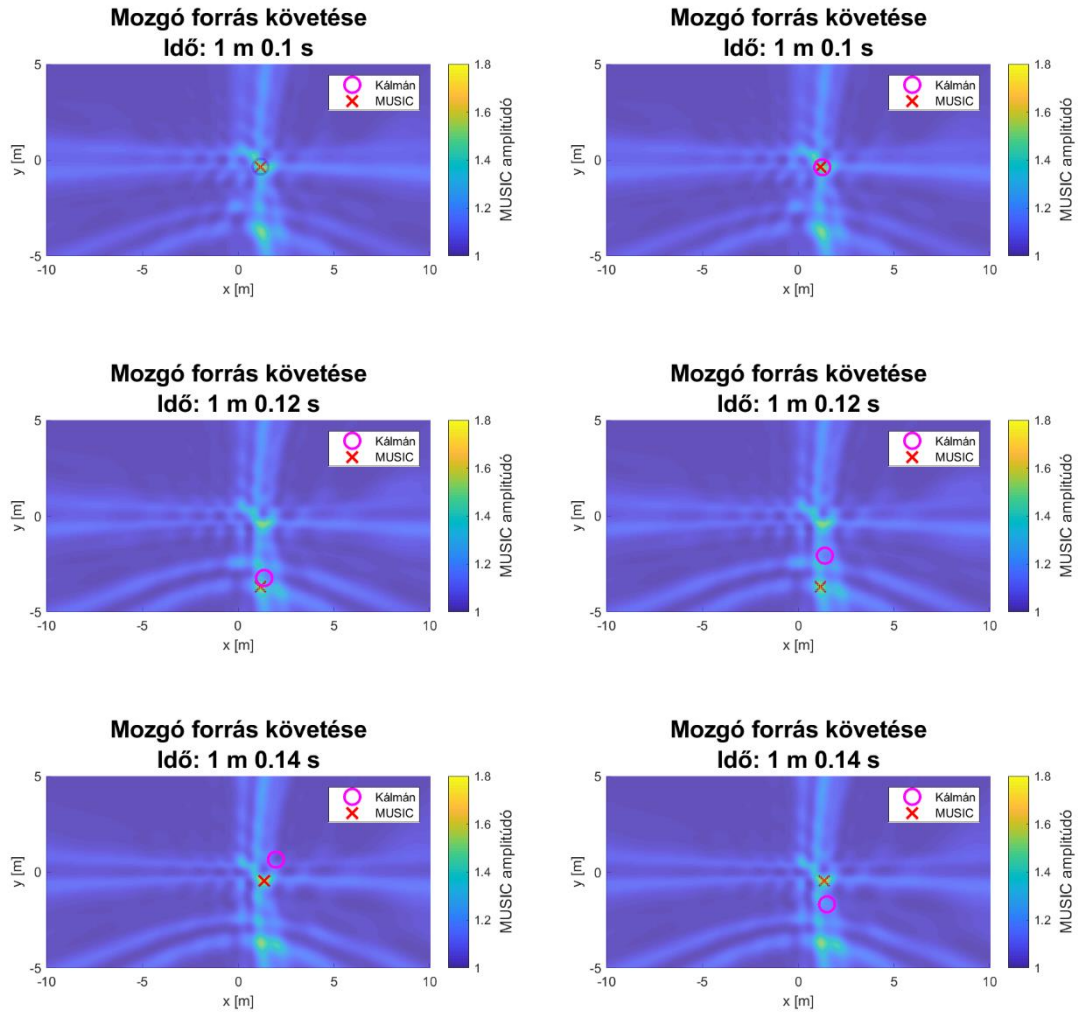
$$Q = T_s \begin{bmatrix} 10^{-3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 10^{-3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 10^{-3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 10^3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10^3 \end{bmatrix}, \quad (49)$$

$$R_0 = \frac{1}{T_s} \begin{bmatrix} 10^{-5} & 0 & 0 \\ 0 & 10^{-5} & 0 \\ 0 & 0 & 10^{-3} \end{bmatrix}. \quad (50)$$

A P_0 mátrix diagonális, mert feltételezzük, hogy kezdetben a becslések függetlenek egymástól, illetve nagyra választottuk a kezdeti pontatlanság miatt. A Q mátrixban az első három és az utolsó három főátló menti mezők értékei közötti különbséget az indokolja, hogy a sebességet jóval pontatlanabban tudjuk becsülni, mint a pozíciót. Az időablak hossza megegyezik a MUSIC algoritmus blokkhosszával, és a teljes mérés időtartamának és ennek az időnek a hányadosa a blokkok száma. Az öt paraméter közül az első négyet változtatva befolyásolhatjuk a Kálmán-szűrő viselkedését, ezért MATLAB környezetben, egy adott drónra (ez esetben a Flying Ranger névre hallgat) vonatkozó mérési eredményekre alkalmazva megvizsgáltam az algoritmus paraméterfüggő működését.

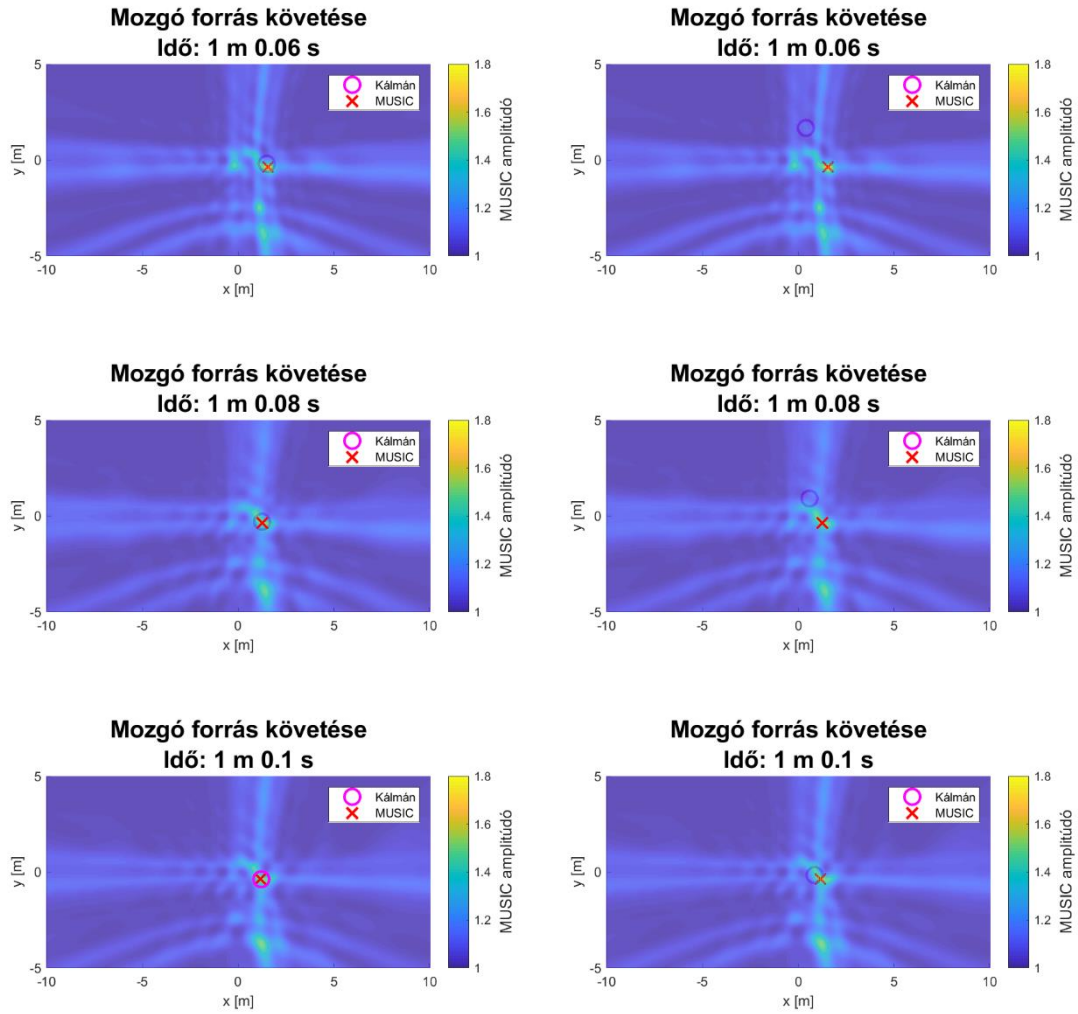
A rendszerzaj (Q) változtatásával befolyásolhatjuk azt, hogy a Kálmán-szűrő által felállított modellre vagy a MUSIC algoritmustól kapott mérési eredményekre helyezünk-e nagyobb hangsúlyt. A rendszerzaj megnövelésével lényegében azt mondjuk, hogy „nem bízunk annyira” a modellben, ezért a mérésekre „hallgatunk inkább”. Ez esetben a szűrő becslései sokkal szorosabban követik a mérési eredményeket, ami pontatlan modell és jó mérési körülmények esetén előnyös lehet, viszont pontos modell és rossz jel-zaj viszonyú mérések esetében a szűrőnek a 4.3.4 szakaszban tárgyalt stabilizáló hatása kevésbé érvényesül. A rendszerzaj csökkentésével ezzel ellentétes hatást érhetünk el, azaz nagyobb hangsúlyt fektetünk a modellre és kisebbet a mérésekre, mert a kisebb rendszerzaj pontosabb modellt feltételez. Ekkor a szűrő becslései sokkal kevésbé szorosan követik a mérési eredményeket, gyorsan változó mozgásnál „lemarad”, viszont a stabilizáló hatása jobban érvényesül.

A kimeneti zaj (R_0) a mérési eredménybe vetett bizalmunkat reprezentálja, ezért a változtatásával ugyanazt befolyásolhatjuk a Kálmán-szűrő működésében, mint a rendszerzajjal, csak azzal ellentétesen. A kimeneti zaj megnövelésével a modellen lesz nagyobb hangsúly, csökkentésével pedig a mérési eredményeken (25. ábra).



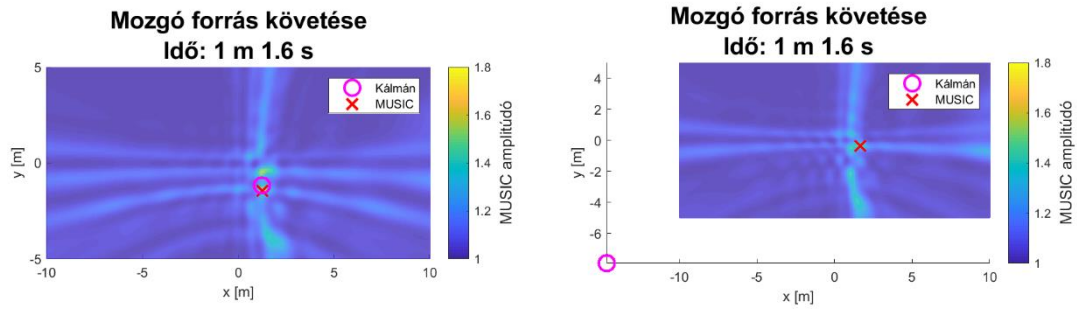
25. ábra: Mozgó hangforrás követése Kálmán-szűrővel, a mérési eredményekre (bal), majd a modellre (jobb) helyezve a hangsúlyt.

A becslés kovarianciamátrixának kezdeti értéke (P_0) lényegében a kezdeti bizonytalanságot adja meg. Ha a mátrixnak nagyobb értéket adunk, akkor ezt a bizonytalanságot nagyobbak feltételezzük. Egy adott szimuláció vagy mérés legelején néhány blokkhossznyi ideig a Kálmán-szűrő nem ad hasznos becslést, időre van szüksége, amíg beáll a MUSIC általi mérési eredményhez közeli értékekre. Ha nagyobb kezdeti bizonytalanságot feltételezünk, akkor több időre van szüksége ahhoz, hogy ez megtörténjen (26. ábra).



26. ábra: Mozgó hangforrás követése Kálmán-szűrővel, normál (bal), majd tízszeres (jobb) kezdeti bizonytalansággal.

A blokkhossz (T_s) a rendszerzaj és a kimeneti zaj kovarianciamátrixait is befolyásolja, előbbivel egyenesen, utóbbival fordítottan arányos. Ezért (bizonyos korlátok között) ha az időablakot (egyben a MUSIC algoritmus blokkhosszát) megnöveljük, a rendszerzaj nagyobb, a kimeneti zaj kisebb lesz, ezzel nagyobb lesz a hangsúly a méréseken. Az időablak csökkentésével pedig a modellen lesz nagyobb hangsúly. Az időablak egy bizonyos ponton túli növelésével (egyelőre ismeretlen okokból) a Kálmán-szűrő becslései teljesen használhatatlanok lesznek, az általa becsült pozíció látszólag szabálytalanul és nagymértékben változik egy-egy időlépésben, és nincs összefüggésben a mérési eredményekkel (27. ábra).

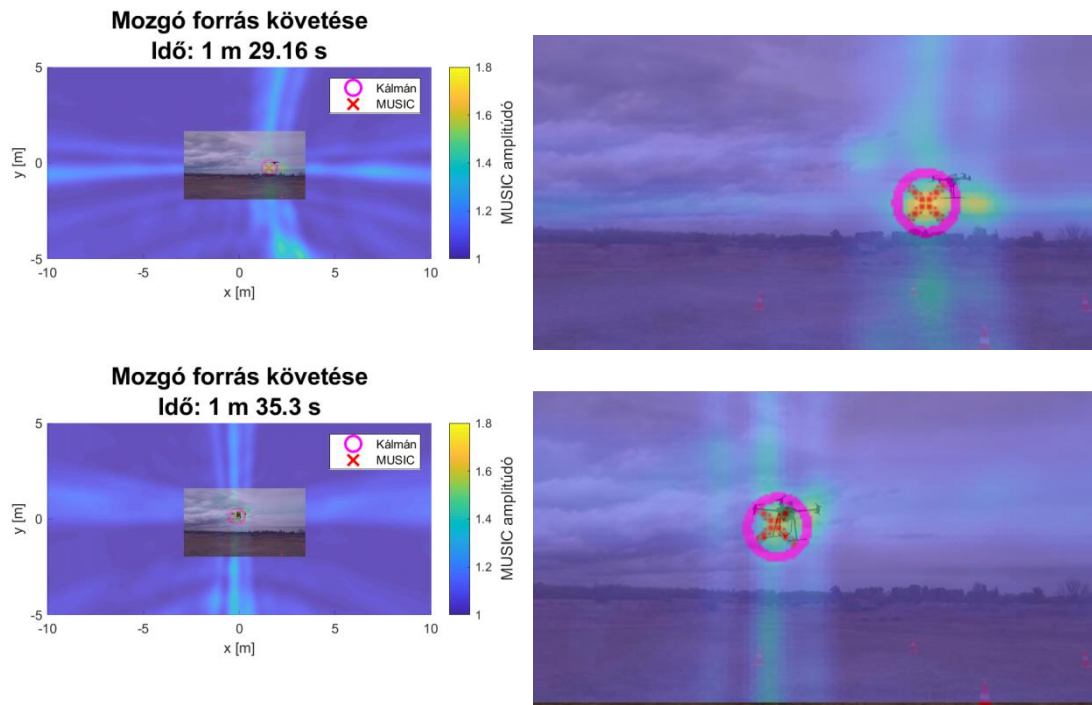


27. ábra: Mozgó hangforrás követése Kálmán-szűrővel, 0,02 másodperces (bal), majd 0,1 másodperces (jobb) blokkhosszal.

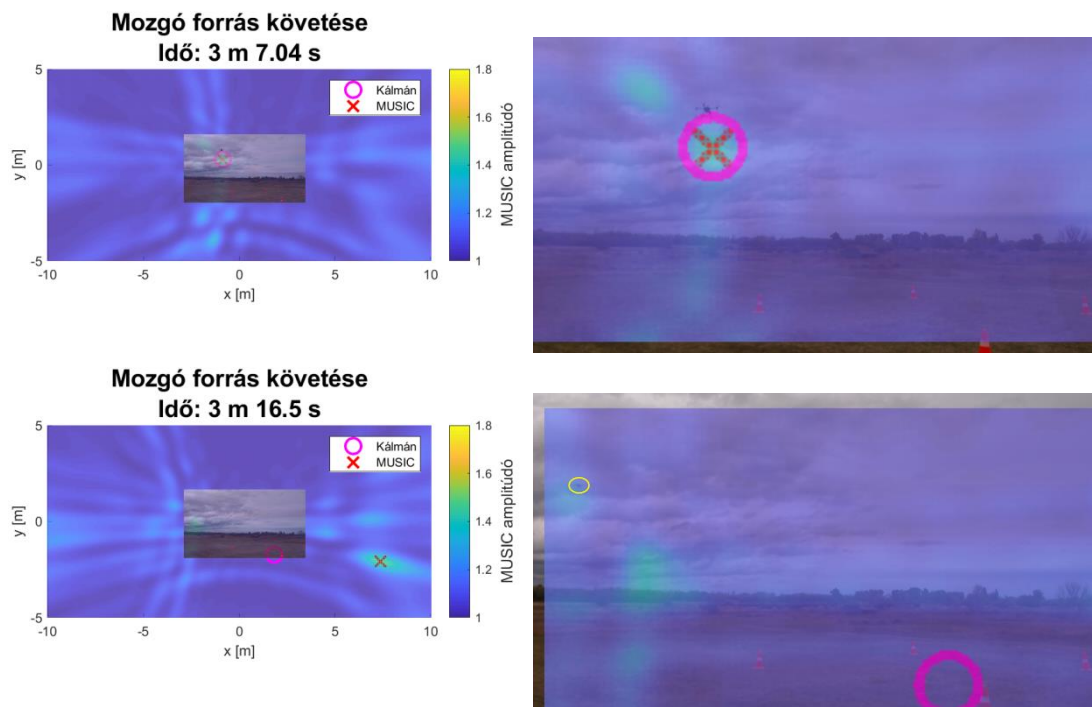
5.1.2 Secopx8

Ebben a szakaszban a Secopx8 nevű drónról készített mérésekből ábrázolok néhány rövid szakaszt. A MUSIC algoritmus által létrehozott kameraképen a maximumkeresés és a Kálmán-szűrő becslését is ábrázolom. Mindemellett ábrázolok az adott amplitúdótérképek időpontjának megfelelő pillanatképeket az állvány tetején lévő webkamera által készített videofelvételből is. Habár a kamera látószöge jóval kisebb, mint az akusztikai vászon által felölelt szögtartomány, az optikai és az akusztikai felvételek közötti összefüggés jól látható, a drón mozgása követhető.

A 28. ábra látható a drón mozgása a vászon által lefedett területen belül. Az algoritmusok jól tudják követni a mozgás irányát, de a távolságbecslés jelenlegi hiányában pontos pozíció- és sebességbecslést még nem lehet végezni. A 29. ábra a drón távolodik a mikrofontömbtől, így a mikrofonok által felvett hangjelek egyre gyengébbek, a hangforrás egyre jobban beolvad a háttérzajba, illetve a fókusztávolság is túl kicsi lesz a valódi távolsághoz képest, és az algoritmusok egyre nehezebben tudják követni.



28. ábra: Secopx8 mozgása a vizsált szögtartományon belül.

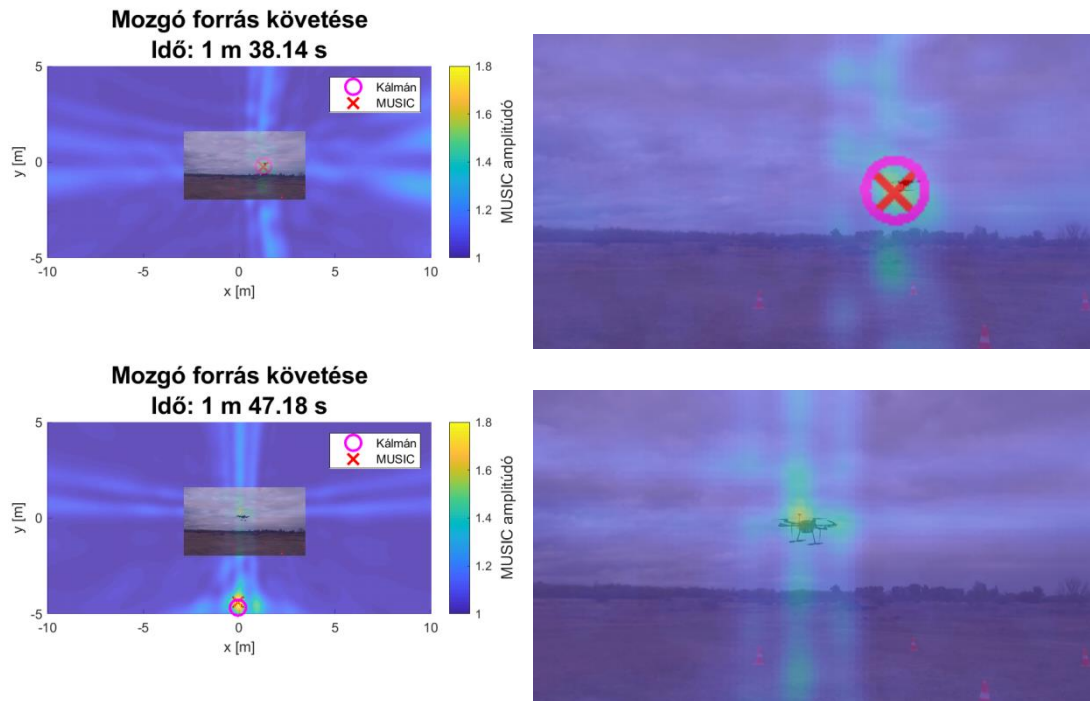


29. ábra: Secopx8 távolodása a mikrofontömbtől.

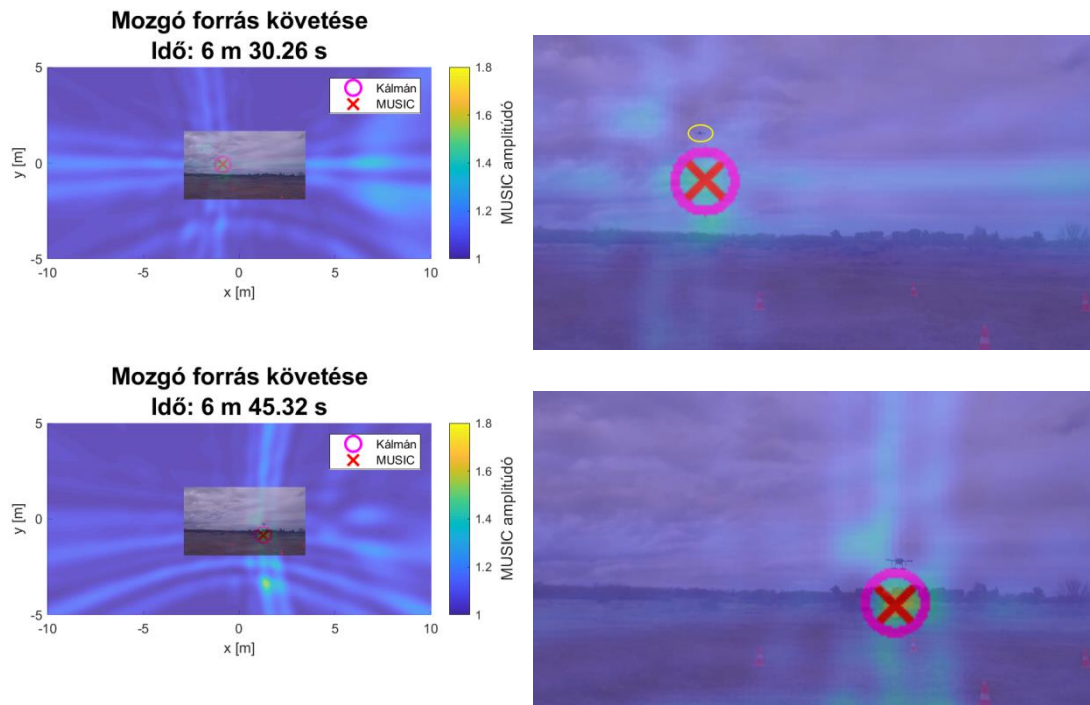
5.1.3 Tarot680

Ebben a szakaszban egy másik drónról (Tarot680) készített mérésekről ábrázolok rövid szakaszokat, ugyanabban a formában. A 30. ábra a drón az előzőhöz

hasznos módon mozog a vizsgált szögtartományon belül, viszont ezen a szakaszon a vászon alsó szélén megjelenik egy zavaró hatás, például reflexió, ami megnehezíti a helyes iránybecslést. A 31. ábra a gép a mérés végén a leszállópályához közeledik. Minél közelebb van a mikrofontömbhöz és az akusztikus vászon síkjához, annál jobb minőségű amplitúdótérképet lehet készíteni, és annál pontosabb becslést lehet végezni.



30. ábra: Tarot680 mozgása a vizsgált szögtartományon belül.



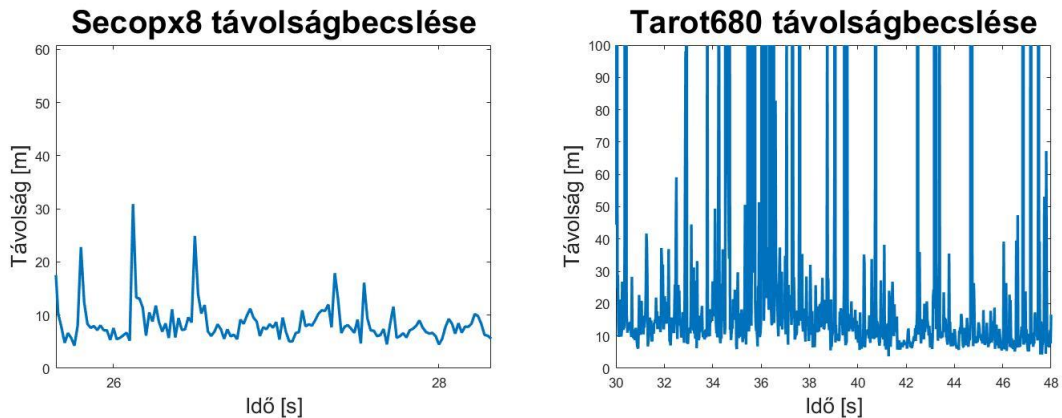
31. ábra: Tarot680 közeledése a leszállóterület felé.

5.1.4 Drónok irány- és távolságbecslése

Az előzőekben a két drónnak csak az irányával foglalkoztunk, aminek a mérését bizonyos korlátok között megfelelően el lehetett végezni. A szimulációk alapján a távolságbecslés MUSIC algoritmussal jól működik akkor, amikor a hangforrás irányát is meg tudjuk határozni, viszont a mérések alatti kevésbé ideális körülmények megnehezíthetik a folyamatot. A kiindulásként iránybecslésre használt vászon az eddigi méréseknél használttól különbözik, a 4.3.5 szakaszban leírttal (60-szor 30 méteres téglalap 15 méter távolságra) egyezik meg.

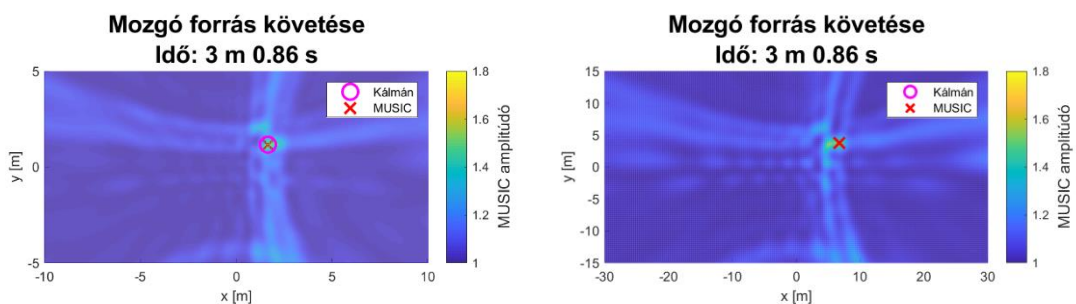
A 32. ábra mutatja a korábbiakban is vizsgált két drón becsült távolságát. A Secopx8 esetében az ábrázolt szakasz a mérés második percében van, miközben a jármű a mikrofontömb előtt, attól időben nagyjából állandó távolságra halad el. A becslés bizonytalansága a szimulációkhoz képest sokkal nagyobb, az értékek sokat ingadoznak, és bár lehet feltételezni egy nagyjából 10 méteres távolságot, ez az eredmény az ábra alapján nem meggyőző. Különösen igaz ez a Tarot680-ra, ahol az ábrázolt szakasz a mérés hetedik percében van, miközben a drón a leszállóterülethez (és a mikrofontömbhöz is) közeledik. A becslések itt teljesen használhatatlanok, az ingadozás az előzőnél is sokkal nagyobb mértékű. Szimulációknál a módszer működött, mérési

eredmények feldolgozásához azonban egyelőre még nem alkalmas, ezért továbbfejlesztésre szorul.



32. ábra: Secopx8 és Tarot680 távolságának a becslése, előbbinél időben közel állandó, utóbbinál egyre csökken a valódi távolság.

A távolságbecslésnek behozatalának egy másik negatív következménye, hogy a Kálmán-szűrő a MUSIC algoritmus alapján biztosított mérési eredményeket nehezebben tudja követni (33. ábra). A hangforrás (ugyanahhoz az időpillanathoz tartozó) becsült irányai közel azonosak, különbség abból adódhat, hogy ha a kiindulásként használt akusztikus vászonok távolsága a mikrofontömbtől különböző (itt például bal oldalon ez 5 méter, míg jobb oldalon 15 méter). A becsült pozíciók viszont már nem csak egy síkon helyezkednek el, a szomszédos időpillanatban felvett pozíciók között sokkal nagyobb távolság lehet, ráadásul a mozgás iránya is hirtelen és nagy gyakorisággal változhat; az ilyen nagymértékben szabálytalan mozgásoknál pedig a Kálmán-szűrő stabilizáló hatása nem tud érvényesülni.



33. ábra: Secopx8 követése Kálmán-szűrővel, bal oldalon távolságbecslés nélkül, jobb oldalon távolságbecsléssel együtt.

5.2 Továbbfejlesztési lehetőségek

Jelen munka többféleképpen is kibővíthető, továbbfejleszhető. Érdeemes megfigyelni, hogy az algoritmusok miként viselkednek különböző típusú hangforrásokra a fehérzajon kívül (pl. azonos frekvenciájú szinuszjeleket sugárzó források lokalizációjára nem alkalmas a MUSIC, mert azok nem korrelálatlanok). A kör és a spirál mikrofon elrendezések is a gyakorlatban megvalósítható elrendezések, ezért érdemes ezekkel is szimulációkat és méréseket elvégezni.

Az l_1 -minimalizálással és szintkorrekcióval kiegészített Compressive Beamforming megfelelő vizsgálati frekvencián pontos iránybecslésre képes, de hasznos lehet megvizsgálni, hogy alacsonyabb frekvenciákon mennyire működésképes a módszer. Az rosszul definiált akusztikai vászon esetén fellépő Basis Mismatch probléma (ezen belül is a virtuális forráspozíciók rossz iránya és távolsága), és ennek kiküszöbölése a CS folytonossá tételével szintén érdekes továbbfejlesztési irány lehet. Bevezethetjük a CS-nek egy olyan változatát, ahol a vásznat leszűkítjük azoknak a pontoknak/irányoknak a halmazára, ahol a CBF vagy a MUSIC által forrást találtunk (például olyan virtuális forráspozíciókra, amikre a legnagyobb becsült forrásamplitúdóhoz képest az elem egy mínusz 10 dB-es küszöbön felül van).

A fókusz-távolság változtatásával lehetőség van az iránybecslésen túl távolságbecslésre is, azonban ez bizonyos mértékben megnöveli a számítási igényeket. A jelenleg kidolgozott módszer alapján MATLAB környezetben implementált algoritmus jelenleg még nem használható valós idejű alkalmazásokban, ezért érdemes lehet az ideális fókusz-távolság keresését optimalizálni valamilyen módszerrel. Mérési eredmények feldolgozása során a módszer nem működött megfelelően, ezért a számítási költségek csökkentése mellett fontos a pontosság és megbízhatóság növelése is. Miután az iránnyal és a távolsággal együtt képesek vagyunk konkrét pozíciót becsülni, megnyílik a lehetőség a teljesítmény, illetve a Kálmán-szűrő által a sebesség becsülésére.

6 Összefoglalás

Dolgozatomban céloim nyalábformáló algoritmusok, illetve a Kálmán-szűrő bemutatása, illetve azok tesztelése volt szimulációkkal és mérési eredmények feldolgozásával. Az algoritmusokat akusztikai kameraként használt mikrofonok vett jelen alkalmazva a mikrofontömb alkalmassá válik arra, hogy hangforrásokat lokalizáljon.

Három főbb nyalábformálási módszert mutattam be, ezek a hagyományos nyalábformálás (CBF), a MUSIC, és a Compressive Sensing alapú Compressive Beamforming. Ezek közül a CBF a legegyszerűbb, ugyanakkor ez az algoritmus eredményezi a legkisebb felbontású kameraképet, és kisméretű hangforrások helybecslésére nem alkalmas. A MUSIC egy lineáris algebrai módszer, aminek előnye a CBF-fel szemben, hogy jobb a zajtűrése, kisméretű forráslokálizációra, és nagyobb a felbontása, hátránya viszont, hogy csak inkohérens források esetén működik. A Compressive Beamforming a nyalábformálást egy konvex optimalizálási problémára vezeti vissza, számításigénye jóval nagyobb, mint a CBF-é és a MUSIC-é, és a zajtűrése a MUSIC-nál kisebb. Az itt bevezetett ritkasági kényszerfeltételnek köszönhetően azonban nagyon nagy felbontású kameraképet kaphatunk vele. Az módszert kiegészíthetjük újraszúlyozott l_1 -minimalizálással és szintkorrekcióval, illetve bevezethetjük egy olyan változatát, amely pontszórásfüggvényeken alapul. Mindkettő nagyobb zajtűrésű eljárást eredményez. Javaslatot tettem az akusztikus vászon egy olyan módú kibővítésére, amellyel a nyalábformáló algoritmusokkal a hangforrások iránya mellett a távolságukat is lehet becsülni.

A Kálmán-szűrő egy olyan algoritmus, amelyet használhatunk mozgó hangforrások követésére. Az éppen aktuális mérési eredmények mellett (amit a nyalábformáló algoritmusok biztosítanak) felhasználja a vizsgált rendszer korábbi állapotát is, ezzel pedig elérhetjük, hogy zajjal terhelt mérés során pontosabb becslést végezzünk, és predikciót is végezhetünk. A Kálmán-szűrőnek tárgyaltam a hagyományos változatát és egy nemlineáris rendszerekre alkalmazható kiterjesztését, és az egynél több hangforrás követésére alkalmas kibővítését is.

Az algoritmusokkal két- és háromdimenziós szimulációkat végeztem, ahol a forrásokat fehérzajnak feltételeztem. Két dimenzióban vonal, három dimenzióban pedig

kereszt alakú mikrofonelrendezést használtam. Végül a szakmában újdonságként szabadtéri mérési eredményekkel is teszteltem a nyalábformáló módszereket. A Kálmán-szűrő a szimulációk alapján az összességében legkedvezőbb MUSIC algoritmussal együttműködve képes hangforrások lokalizálására és követésére, azonban bizonyos zavaró hatások (pl. rossz jel-zaj viszony, reflexiók) korlátozzák a helyes működését. A felhasznált keretrendszer továbbfejleszhető többféleképpen, például kidolgozhatóak a Compressive Beamforming-nak kedvezőbb változatai, és a távolságbecslés tovább gyorsítható és pontosítható.

Köszönetnyilvánítás

Szeretném megköszönni konzulensemnek, Dr. Fiala Péternek a segítségét, tanácsait és iránymutatását. Köszönöm Dr. Rucz Péternek is a munkám haladását segítő ötleteit, észrevételeit. Végül megköszönném családomnak támogatásukat, építő kritikáikat és érdeklődésüket.

Irodalomjegyzék

- [1] J. Benesty, J. Chen, Y. Huang: „Microphone Array Signal Processing”. Springer-Verlag, Berlin, Heidelberg (2008).
- [2] H. E. Camargo, R. A. Burdisso, P. A. Ravetta, A. K. Smith: „A comparison of beamforming processing techniques for low frequency noise source identification in mining equipment”. American Society of Mechanical Engineers, 2009, pp. 1-7.
- [3] Z. D. Zaharis, I. P. Gravas, P. I. Lazaridis, T. V. Yioultsis, C. S. Antonopoulos, T. D. Xenos: „An Effective Modification of Conventional Beamforming Methods Suitable for Realistic Linear Antenna Arrays”. IEEE Transactions on Antennas and Propagation, Vol. 68 (7), July 2020, pp. 5269-5279, doi: 10.1109/TAP.2020.2977822.
- [4] M. Mohanna, M. L. Rabeh, E. M. Zieur, S. Hekala: „Optimization of MUSIC algorithm for angle of arrival estimation in wireless communications”. NRIAG Journal of Astronomy and Geophysics, Vol. 2 (1), June 2013, pp. 116-124.
- [5] Q. Zhao, W. Liang: „A Modified MUSIC Algorithm Based on Eigen Space”. In: Jin D., Lin S. (eds) Advances in Computer Science, Intelligent System and Environment. Advances in Intelligent and Soft Computing, Vol 104. Springer, Berlin, Heidelberg, 2011, doi: 10.1007/978-3-642-23777-5_45.
- [6] S. Li, X. Chen, R. He: „Robust Cyclic MUSIC Algorithm for Finding Directions in Impulsive Noise Environment”. International Journal of Antennas and Propagation, Vol. 2017, Article ID 9038341, 9 pages, 2017, doi: 10.1155/2017/9038341.
- [7] A. Xenaki, P. Gerstoft, K. Mosegaard: „Compressive beamforming”. The Journal of the Acoustical Society of America, Vol. 136 (1), 2014, pp. 260-271.
- [8] M. Coldrey, M. Viberg: „Generalization and analysis of the conventional beamformer for localization of spatially distributed sources”. 2006 14th European Signal Processing Conference, Florence, 2006, pp. 1-5.
- [9] T. C. Yang: „Deconvolved Conventional Beamforming for a Horizontal Line Array”. IEEE Journal of Oceanic Engineering, Vol. 43 (1), Jan. 2018, pp. 160-172, doi: 10.1109/JOE.2017.2680818.
- [10] P. Gupta, S. P. Kar: „MUSIC and improved MUSIC algorithm to estimate direction of arrival”. 2015 International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur, 2015, pp. 0757-0761, doi: 10.1109/ICCSP.2015.7322593.
- [11] L. Yaning, F. Juntao, R. Xinghao, M. Le: „An improved MUSIC algorithm for DOA estimation of non-coherent signals with planar array”. J. Phys.: Conf. Ser. 1060 012026, 2018.

- [12] A. Xenaki, P. Gerstoft: „Grid-free compressive beamforming”. *The Journal of the Acoustical Society of America*, Vol. 137 (4), 2015, 1923-1935.
- [13] Y. Park, Y. Choo, W. Seong: „Multiple snapshot grid free compressive beamforming”. *The Journal of the Acoustical Society of America*, Vol. 143, 2018, 3849-3859.
- [14] M. Imran, A. Hussain, N. M. Qazi, M. Sadiq: „A methodology for sound source localization and tracking: Development of 3D microphone array for near-field and far-field applications”. *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Islamabad, 2016, pp. 586-591, doi: 10.1109/IBCAST.2016.7429936.
- [15] F. Keyrouz, K. Diepold, S. Keyrouz: „High performance 3D sound localization for surveillance applications”. *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, London, 2007, pp. 563-566, doi: 10.1109/AVSS.2007.4425372.
- [16] M. Lustig, D. Donoho, J. M. Pauly: „Sparse MRI: The application of compressed sensing for rapid MR imaging”, *Magn. Reson. Med.* 58, 2007, 1182–1195.
- [17] N. Wagner, Y. C. Eldar, Z. Friedman: „Compressed beamforming in ultrasound imaging”. *IEEE Transactions on Signal Processing* Vol. 60, 2012, 4643–4657.
- [18] E. J. Candés, T. Tao: „Decoding by linear programming”. *IEEE Trans. Inf. Theory* 51, 4203–4215 (2005).
- [19] H. Yao, P. M. Shearer, P. Gerstoft: „Compressive sensing of frequency-dependent seismic radiation from subduction zone megathrust ruptures”. *Proc. Natl. Acad. Sci. U.S.A.* Vol. 110, 2013, 4512–4517.
- [20] J. Romberg: „Imaging via compressive sampling”. *IEEE Signal Processing Magazine* Vol. 25, 2008, 14–20.
- [21] Csóka Bence: „Hangforrások lokalizációja mikrofonrendszerekkel”, *BME TDK dolgozat* (2019, 2020)
- [22] J. Novoa, R. Mahu, A. Díaz, J. Wuth, R. Stern, N. B. Yoma: „Weighted delay-and-sum beamforming guided by visual tracking for humanrobot interaction”. 2019, arXiv:1906.07298.
- [23] Tapolczai Péter: „Mozgó zajforrások észlelése és követése akusztikus nyalábformálással”, *BME Diplomatervezés* (2018).
- [24] R. Schmidt: „Multiple emitter location and signal parameter estimation”. *IEEE Transactions on Antennas and Propagation* Vol. 34, 1986, pp. 276–280.
- [25] E. J. Candés, M. B. Wakin: „An introduction to compressive sampling”. *IEEE Signal Processing Magazine* Vol. 25, 2008, 21–30.
- [26] R. G. Baraniuk: „Compressive sensing”. *IEEE Signal Processing Magazine* Vol. 24, 2007, 118–121.

- [27] E. J. Candés, J. Romberg, T. Tao: „Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. IEEE Transactions on Information Theory Vol. 52, 2006, 489–509.
- [28] M. Elad: „Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing”. Springer, New York, 2010, pp. 1–359.
- [29] S. Foucart, H. Rauhut: „A Mathematical Introduction to Compressive Sensing”. Springer, New York, 2013, pp. 1–589.
- [30] J. A. Tropp: „Just relax: Convex programming methods for identifying sparse signals in noise”. IEEE Transactions on Information Theory Vol. 52, 2006, 1030–1051.
- [31] E. J. Candés, M. B. Wakin, S. Boyd: „Enhancing sparsity by reweighted l_1 minimization,”. J. Fourier Anal. Appl. Vol. 14, 2008, 877–905.
- [32] C. F. Mecklenbräuker, P. Gerstoft, A. Panahi, and M. Viberg: „Sequential Bayesian sparse signal reconstruction using array data”. IEEE Transactions on Signal Processing Vol. 61, 2013, 6344–6354.
- [33] D. Simon: „Optimal State Estimation – Kalman, H_∞ , and Nonlinear Approaches”, John Wiley & Sons, Inc., Hoboken, New Jersey (2006).
- [34] Belső Zoltán, Gáti Balázs, Koller István, Rucz Péter, Turóczi Antal: "Design of a nonlinear state estimator for navigation of autonomous aerial vehicles" Repüléstudományi közlemények (Aviation scientific publications) XXVII/3 pp. 255–276 (2015).
- [35] E. A. Wan, R. Van Der Merwe: „The unscented Kalman filter for nonlinear estimation”. Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), Lake Louise, Alberta, Canada, 2000, pp. 153-158, doi: 10.1109/ASSPCC.2000.882463.
- [36] G. Valverde, V. Terzija: „Unscented Kalman filter for power system dynamic state estimation”. IET Generation, Transmission & Distribution, Vol. 5 (1), 2011, pp. 29-37, doi: 10.1049/iet-gtd.2010.0210.
- [37] Natan (2021). Fast 2D peak finder (<https://www.mathworks.com/matlabcentral/fileexchange/37388-fast-2d-peak-finder>), MATLAB Central File Exchange. Retrieved April 26, 2021.
- [38] Fiala Péter, Rucz Péter: „Hangjelek digitális feldolgozása az akusztikai gyakorlatban”. Mérési leírás.
- [39] Y. Chi, L. L. Scharf, A. Pezeshki, A. R. Calderbank: “Sensitivity to basis mismatch in compressed sensing”. IEEE Transactions on Signal Processing Vol. 59, 2011, 2182–2195.

HALLGATÓI NYILATKOZAT

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

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

Kelt: Budapest, 2021. 06. 15.

Csóka Bence

.....
Csóka Bence