

# Beszéd kiemelése zajból a rekonstruált fázistérben

PINTÉR ISTVÁN

Kecskeméti Főiskola GAMF Kar, Kalmár Sándor Informatikai Intézet,  
Automatizálási és Alkalmazott Informatikai Szakcsoport  
pinter.istvan@gamf.kefo.hu

Lektorált

**Kulcsszavak:** beszéd kiemelése zajból, jel-altér, rekonstruált fázistér, dimenzió-beágyazás

A dolgozat beszédjel zajból való kiemelésére szolgáló módszert ismerteti, amely a rekonstruált fázistér és a dimenzió-beágyazás fogalmaira épül. Az algoritmus a beszédet a zajtól a transzformált térben elvégzett nemlineáris művelettel választja szét. A dolgozat a beszéddel nem korrelált additív zaj esetében elért jelenlegi eredményeinket mutatja be.

## 1. Bevezetés

A gépi beszédfeldolgozásban régóta meglévő feladat a beszédjel kiemelése a zajból [1]. Az elmúlt mintegy három évtizedben több módszert is kidolgoztak a zaj csökkentésére. A legtöbb eljárásban közös az a feltételezés, hogy a lineáris beszédjel-modellnek megfelelő zajjellemzők időben lassan változnak. Példaként a hallásmodell-alapú szűrőssal végzett feldolgozás említhető, ahol a rész-sávokban Wiener-szűrést alkalmaznak [2]. A tapasztalatok szerint mintegy 6...9 dB-nél nagyobb jel/zaj viszony (SNR) esetén érhető el e módszerekkel jó eredmény [3]. Ennél kisebb SNR illetve nemstacionárius zaj esetén a zajcsökkentő módszerek alapja többnyire nemlineáris modell. Nemlineáris rendszermodellre példa az emberi hallórendszer egy modellje, amit zajcsökkentő eljárásban is alkalmaznak [4], míg nemlineáris jelmodell a jelen dolgozatban is vizsgált beszédmodell, nevezetesen a rekonstruált fázistérben történő beszédábrázolás [5].

A cikk felépítése a következő. Először ismertetjük a zajmentes beszédjel ábrázolását, amit a továbbiakban felhasználunk. Ezt követi a beszéd-altér fogalmára alapozott zajcsökkentési eljárás leírása a szakirodalom alapján, ahol bemutatjuk a rekonstruált fázistérben működő változatot is. A negyedik szakaszban ismertetjük az eljárás megvalósításával elért zajcsökkentési eredményeinket, majd a cikket a következtetések zárják.

## 2. A zajmentes beszédjel ábrázolása a transzformált térben és a rekonstruált fázistérben

A cikkben ismertetett zajcsökkentő eljárás két feltételezésre épül: az egyik, hogy létezik a beszédminta-sorozat optimális ábrázolása, a másik pedig, hogy a beszédfeldolgozási feladatokhoz is használható a rekonstruált fázistér fogalma.

Ami az első feltételezést illeti, ebben az esetben a feldolgozás alatt álló szegmens  $N$  számú  $\alpha_n$  beszédmintájából vektort alkotunk, így a szegmens az  $N$  dimen-

ziós tér egy pontjának felel meg. Ez az  $\underline{s}$  vektor az úgynevezett  $\{\underline{t}_n\}$  természetes, ortonormált bázisban a bázisvektorok lineáris kombinációjaként írható fel, ahol az együtthatók a beszédminták:  $\alpha_n = (s, \underline{t}_n)$ , és az  $N$  dimenziós  $\underline{t}_n$  oszlopvektor  $n$ -edik komponense 1, a többi 0.

A gépi beszédfeldolgozás gyakorlati problémáinak megoldása során szerzett tapasztalatok szerint a beszédjel esetén létezik olyan bázis, amelybeli reprezentációban  $N$ -nél kevesebb számú összetevővel is leírható az  $\underline{s}$  vektor [6]. Emiatt fel lehet tenni azt a kérdést, hogy melyik az az ortonormált bázis, amelyben  $\underline{s}$  kevesebb összetevővel adható meg

$$\hat{\underline{s}} = \sum_{n=0}^{L-1} a_n \cdot \underline{v}_n \quad (1)$$

alakban, ahol  $\{\underline{v}_n\}$  a keresett ortonormált bázis és  $L < N$ , továbbá ez az előállítás optimális abban az értelemben, hogy a

$$J(\underline{e}) = E \left\{ \|\underline{e}\|^2 \right\} = E \left\{ \|\underline{s} - \hat{\underline{s}}\|^2 \right\} = \quad (2)$$

$$= \sum_{n=L}^{N-1} \underline{v}_n^T \cdot E \left\{ \underline{s} \cdot \underline{s}^T \right\} \cdot \underline{v}_n = \sum_{n=L}^{N-1} \underline{v}_n^T \cdot \underline{R} \cdot \underline{v}_n$$

kritériumfüggvény értéke, vagyis az eltérésnégyzet várható értéke minimális (ideális esetben  $L < N$  és  $\|\underline{e}\|=0$ ). A továbbiakban feltesszük, hogy  $E\{\underline{s}\}=0$ , így  $\underline{R} = \underline{K}$ , ami a kovarianciamátrix. Ismeretes, hogy a  $\{\underline{v}_n\}$  optimális ortonormált bázist a kovarianciamátrix sajátvektor-rendszerre adja, az eltérésnégyzet várható értéke pedig

$$J(\underline{e}) = \sum_{n=L}^{N-1} \lambda_{n_n}$$

ahol  $\lambda_n$  a kovarianciamátrix  $n$ -edik sajátértéke. Az  $\underline{s}$  beszédminta-vektor új (optimális) bázisbeli reprezentációját a

$$\underline{T} = (\underline{v}_0^T, \underline{v}_1^T, \dots, \underline{v}_L^T, \dots, \underline{v}_{N-1}^T)^T \quad (3)$$

mátrix segítségével lehet kiszámítani, ami praktikusan rendre a nagyság szerint csökkenő sorrendben felírt sajátértékeknek megfelelő sajátvektorokból, mint sorvektorokból áll.

A második feltételezés a beszédjelnek a rekonstruált fázistérben történő ábrázolására vonatkozik. A rekonstruált fázistér fogalma a diszkrét dinamikai rendszer  $\underline{x}_{n+1} = \underline{F}(\underline{x}_n)$  alakú mozgásegyenletéhez kapcsolható, ahol  $\underline{x}_n$  és  $\underline{x}_{n+1}$  a D dimenziós fázistérbeli pontok,  $\underline{F}$  megfelelő leképezés. A fázistérbeli pontok  $\{\underline{x}_n\}$  halmaza a trajektória, vagy pálya. Ez a pálya közvetlenül nem figyelhető meg, csak az  $\underline{x}_n \rightarrow \underline{g}(\underline{x}_n)$  nemlineáris leképezésén át. Így adódik az  $\alpha_n = \underline{g}(\underline{x}_n)$  megfigyelhető (mérhető) valós szám, vagy beszédminta. Ezeket rendre egymás után  $T_{MV}$  időnként véve adódik az  $\{\alpha_n\}$  beszédminta-sorozat.

Igazolható, hogy ha  $M > 2 \cdot D + 1$ , akkor az  $\alpha_n$  számsorozatból az eredeti  $\{\underline{x}_n\}$  vektorsorozattal ekvivalens  $\{\underline{y}_n\}$  vektorsorozat állítható elő az úgynevezett dimenzióbeágyazás módszerével. A dimenzióbeágyazás az

$$\underline{y}_n(M, \tau) = (\alpha_n, \alpha_{n+\tau}, \dots, \alpha_{n+(M-1)\tau}) \quad (4)$$

vektort eredményezi, ahol  $\tau > 0$  az időeltolás (itt mintaszámmal adott),  $M > 0$  a beágyazási dimenzió. A fentebb említett ekvivalencia azt jelenti, hogy létezik olyan egyértelműen invertálható, sima  $\underline{h}: \underline{y}_n(M, \tau) \rightarrow \underline{x}_n$  leképezés, amivel a két vektorsorozat egymásba átvihető [7]. A dimenzióbeágyazás műveletéhez szükséges M dimenzió és a  $\tau$  időeltolás értékét numerikus kísérletekkel lehet meghatározni az adott beszédtechnológiai alkalmazási feladathoz. A szakirodalmi adatok szerint az  $M \cdot \tau \cdot T_{MV}$  beágyazási időablak 1...5 ms [8].

### 3. Zajcsökkentés a rekonstruált fázistérben altér módszerrel

A rekonstruált fázistér fogalmával leírható zajcsökkentő algoritmus lényegében általánosítása egy, a szakirodalomban régebben közölt eljárásnak [6], ezért először ezt ismertetjük.

A módszer alapja az előző pontban ismertetett tulajdonság, vagyis az, hogy az optimális beszédábrázolás az N számú bázisvektor helyett L számú bázisvektorral is megoldható, ideális esetben zérus eltérésnégyzettel. Az N mintából álló beszédvektor így az N dimenziós tér L dimenziós altérben található, emiatt ezt beszéd-altérnek is nevezik. A zajcsökkentő algoritmus ebben a beszéd-altérben állít elő optimális becslt beszédvektort a kiindulásképpen rendelkezésre álló zajos beszédmintából. A feladat az, hogy a beszédjellel nem korrelált, additív zajjal terhelt

$$\underline{u} = \underline{s} + \underline{w} \quad (5)$$

beszédmintasorozat ismeretében adjuk meg a tiszta beszédvektor  $\underline{s}$  becslését úgy, hogy az  $\underline{s} - \underline{\tilde{s}}$  eltérésvektor hosszának várható értéke a legkisebb legyen, azaz

$$E \{ \|\underline{s} - \underline{\tilde{s}}\|^2 \} \rightarrow \min \quad (6)$$

teljesüljön. Hasonlóan az előző szakaszban foglaltakhoz, itt is meg kell találni az optimális ortonormált bázist, ám most csak  $\underline{u}$  ismert. Feltesszük, hogy  $E\{\underline{w}\} = 0$ , és mivel előző feltevésünk miatt  $E\{\underline{s}\} = 0$ , adódik  $E\{\underline{u}\} = 0$ .

További feltevés, hogy a 0 várható értékű zaj típusa fehér zaj, vagyis kovarianciamátrixa  $\underline{K}^{ZAJ} = \sigma^2 \cdot \underline{I}$ , ahol  $\sigma > 0$  és  $\underline{I}$  NxN méretű egységmátrix. Belátható, hogy ekkor a korrelátlanság miatt a zajos beszéd kovarianciamátrixa a beszéd és a zaj kovarianciamátrixok összege:

$$\underline{K}^{ZAJOS} = E\{\underline{u} \cdot \underline{u}^T\} = \underline{K}^{BESZÉD} + \underline{K}^{ZAJ}, \quad (7)$$

továbbá igazolható az is, hogy a zajos beszéd kovarianciamátrixának és a zajmentes beszéd kovarianciamátrixának sajátvektorai azonosak. Ez utóbbi tulajdonság teszi lehetővé, hogy a zajcsökkentés a meglévő zajos beszédvektorból kiindulva ugyanabban az ortonormált bázisban végezhető el, mint amiben a beszéd reprezentációja optimális és az optimális beszéd-reprezentációt adó  $\{\underline{v}_n\}$  ortonormált bázis kiszámítható a rendelkezésre álló zajos beszéd kovarianciamátrixából is, a zajmentes beszédvektor kovarianciamátrixának ismerete nélkül.

Továbbá, mivel a kovarianciamátrixok összegezhethők, belátható, hogy a zajos beszéd kovarianciamátrixa a transzformált térben a következő diagonálmátrix lesz:

$$\begin{aligned} \underline{K}^{ZAJOS} \Big|_{\{\underline{v}_n\}} &= \\ &= \text{diag}(\lambda_0 + \sigma^2 \dots \lambda_{L-1} + \sigma^2 \quad \sigma^2 \dots \sigma^2). \end{aligned} \quad (8)$$

Az első szakaszban ismertetett feltevésünknek megfelelően a  $\underline{v}_0, \underline{v}_1, \dots, \underline{v}_{L-1}$  vektorok által kifeszített altérben a zajmentes beszéd ideálisan reprezentálható. Szemléletesen szólva a zajos beszéd esetében itt „beszéd és zaj is található”, míg ezen altér ortogonális kiegészítésében, a  $\underline{v}_L, \underline{v}_{L+1}, \dots, \underline{v}_{N-1}$  vektorok által kifeszített altérben „csak zaj található”.

Ezek után a zajcsökkentő eljárást a  $\underline{H}$  lineáris transzformáció alakjában keressük, vagyis

$$\underline{\tilde{s}} = \underline{H} \cdot \underline{u}. \quad (9)$$

A becslés hibája az  $\underline{r} = \underline{s} - \underline{\tilde{s}}$  maradékjel. A [6] szerzői megmutatták, hogy az

$$\underline{r} = \underline{r}^{BESZÉD} - \underline{r}^{ZAJ} \quad (10)$$

maradékjel két összetevőből áll, egy a beszéddel, egy a zajjal korrelált. Emiatt nemcsak a beszéddel korrelált hibaösszetevő minimalizálása a feladat, hanem ezzel egyidejűleg a zajjal korrelált összetevő előírt szint alatt tartása is cél az optimális lineáris transzformáció keresésekor. A feladatot [6]-ban mind az időtartományban, mind a spektrális tartományban előírt feltételek esetében megoldották. Saját, időtartományra vonatkozó eredményeinket [9]-ben tettük közzé. A második esetben is a beszéddel korrelált hibaösszetevő minimalizálása a cél, de most minden egyes spektrális komponensre külön-külön írunk elő zajszintcsökkentési feltételt, azaz

$$J(\underline{r}^{BESZÉD}) \Big|_{\underline{H}} \rightarrow \min \quad (11)$$

feltéve, hogy:

$$\begin{aligned} E \left\{ \left| \underline{v}_n^T \cdot \underline{r}^{ZAJ} \right|^2 \leq \beta_n \cdot \sigma^2 \right\} \quad n = 0, 1, \dots, L-1, \\ E \left\{ \left| \underline{v}_n^T \cdot \underline{r}^{ZAJ} \right|^2 = 0 \right\} \quad n = L, \dots, N-1. \end{aligned} \quad (12)$$

Szemléletesen szólva az első feltétel a beszéd-altérben megmaradt zajra vonatkozó komponensenkénti előírás rendre  $\beta_n > 0$ -val megadott feltételekkel, a második a zaj-altérbeli komponensek nullázását írja elő.

Az optimális transzformáció mátrixa a Karush-Kuhn-Tucker-féle feltételes szélsőérték-keresési módszer alapján [6] szerint a következő:

$$\begin{aligned} \underline{\underline{H}}^{OPT} &= \underline{\underline{V}} \cdot \underline{\underline{G}} \cdot \underline{\underline{V}}^T, \\ \underline{\underline{G}} &= \text{diag}(g_{0,0}, \dots, g_{N,N}), \\ g_{n,n} &= \begin{cases} \sqrt{\gamma_n} & n = 0, 1, \dots, L-1 \\ 0 & n = L, \dots, N-1 \end{cases} \end{aligned} \quad (13)$$

ahol  $\underline{\underline{V}}$  a sajátvektorokból, mint oszlopvektorokból álló mátrix. Az említett mű  $\gamma_n$  megválasztására két javaslatot is ad, ezek közül a jelen munkában a

$$\gamma_n = \exp\left(-\frac{\kappa \cdot \sigma^2}{\lambda_{n, \text{BESZÉD}}}\right) \quad (14)$$

összefüggéssel dolgoztunk. A zajcsökkentés mértékét a  $\kappa \geq 1$  tapasztalati konstanssal lehet beállítani, egyben a tisztított beszéd torzulását is befolyásolva ezzel.

A fent összefoglalt módszer általánosítható a rekonstruált fázistér esetére is. Ugyanis ez utóbbi, mint modellháttér lehetővé teszi, hogy egyetlen megfigyelt  $\underline{\underline{u}} = \underline{\underline{s}} + \underline{\underline{w}}$  N dimenziós zajos beszédvektorból állítsunk elő M dimenziós vektorokból álló adatrendszert a dimenzió beágyazás módszerével. Az így keletkező  $\underline{\underline{U}}_{M \times N}$  trajektóriamátrixra a konstrukciója miatt igaz, hogy

$$\underline{\underline{U}}_{M \times N} = \underline{\underline{S}}_{M \times N} + \underline{\underline{W}}_{M \times N} \quad (15)$$

Mivel  $u_n = s_n + w_n$  minden összetartozó mintára fennáll, a trajektóriamátrixban lévő vektorokból, mint adatrendszerből számolható kovarianciamátrixra nézve teljessül, hogy

$$\underline{\underline{K}}_{\underline{\underline{U}}_{M \times N}} = \underline{\underline{K}}_{\underline{\underline{S}}_{M \times N}} + \underline{\underline{K}}_{\underline{\underline{W}}_{M \times N}}, \quad (16)$$

ahol  $\underline{\underline{K}}_{\underline{\underline{W}}_{M \times N}} = \sigma^2 \cdot \underline{\underline{I}}_{M \times M}$

Emiatt az előzőekben ismertetett gondolatmenet most is alkalmazható, amivel előállítható a becsült  $\underline{\underline{S}}$  trajektória-mátrix. Ebből kell a becsült  $\underline{\underline{s}}$  beszédminta sorozatot visszaállítani, ami  $\underline{\underline{U}}$  előállítása alapján tehető meg. Az eredeti altér-módszertől ez az eljárás abban különbözik, hogy itt más adatrendszer konstruálható a kovariancia-mátrix becslésére és a becsült minta végle-

ges értéke a becsült trajektória-mátrix egynél több eleméből számítható ki.

Az általunk használt trajektória-mátrix a beszéd-szegmens periodikus kiterjesztésén alapul, ezáltal minden minta pontosan M-szer szerepel, akárcsak a becsült trajektóriamátrixban is, így egyetlen becsült beszédminta előállítása átlagolással történhet és nem szükséges súlyozómátrix, ami más konstrukciónál megjelenik [10].

A trajektóriamátrix  $u_{i,j}$  eleme eszerint

$$u_{i,j} = u_{(j+i \cdot \tau) \bmod N}, \quad (17)$$

itt tehát N a beszéd-szegmens mintáinak száma, M a beágyazási dimenzió,  $\tau$  az időeltolás.

A zajcsökkentési eljárás alapja a fenti trajektóriamátrix, mint M dimenziós vektorokból álló adatrendszer alapján becsülhető kovarianciamátrix. Megjegyezzük, hogy ez a kovarianciamátrix különbözik mind a [6]-ban használt empirikus Toeplitz-kovarianciamátrixtól, mind pedig az [5]-ben, illetve [10]-ben ismertetett változatoktól.

#### 4. Zajcsökkentési eljárás megvalósítása, numerikus kísérleti eredmények

A jelen dolgozatban vizsgált kiinduló beszédállomány úgy jött létre, hogy a leírt mondatot egy magyar anyanyelvű, férfi bemondó valósította meg, a beszédmintákat 8 kHz mintavételi frekvenciával és 16 bites lineáris kvantálással állították elő. Az aktív beszédszakaszon számolt globális jel/zaj viszony 45,8 dB volt. A zajos beszédállományokat ebből mesterséges zajosítással készítettük, a jelen munka során használt zajok az RSG-10 zaj adatbázisból származnak [11], 8 kHz mintavételi frekvenciájúra átalakítva az eredetileg 19980 Hz-cel mintavett jeleket.

A vizsgálatban használt zajtípusok: fehér zaj, rózsaszín zaj, hírközlő csatorna zaja. A zajszint beállításának alapjául a tiszta beszéd aktív beszédszakaszain számolt energia szolgált. A zajcsökkentés hatékonyságát az

$$SRR = 10 \cdot \lg\left(\frac{E^{\text{BESZÉD}}}{E^{\text{MARADÉK}}}\right), \quad (18)$$

(Signal to Residual Ratio) számmal jellemeztük, ahol a számlálóban az aktív beszédszakaszok indexhalmaza számolt beszédenergia, a nevezőben ugyanezen indexhalmazon számított, zajcsökkentés utáni maradékjel energiája szerepel.

Jel/zaj viszony (dB)	SRR (dB)		
	Fehér zaj	Nagyfrekvenciás hírközlő csatorna zaja	Rózsaszín zaj
15	9,3	9,3	9,0
12	9,1	9,0	8,3
9	8,7	8,6	7,3
6	8,1	7,9	5,7
3	7,2	6,9	3,7
0	6,0	5,7	1,8
-3	4,5	4,2	0,0

1. ábra  
A javulás értékei különböző jel/zaj viszony értékek és zajtípusok esetén (szegmenshossz: 800 minta, beágyazási dimenzió: 20, időeltolás: 1 minta, beszéd-altér dimenziója: 7, a tapasztalati konstans:  $\kappa=5$ )

A zajcsökkentést időben átlapolt beszédsegmentek sorozatán haladva, rendre szegmensről szegmensre végeztük egyetlen szegmens mintái alapján. Az adott szegmens mintáit Hanning-ablakkal súlyoztuk, 50%-os átlapolást alkalmaztunk és a tisztított beszédminta sorozatot az átlapolás és hozzáadás módszerével számítottuk ki. A szegmens hosszát, a beágyazási dimenzió és az időeltolás értékét, a beszéd-altér dimenzióját és a  $\kappa$  tapasztalati konstans a tisztított beszéd meghallgatása alapján határoztuk meg. Az előző szakaszban ismertetett spektrális tartománybeli módszerben szereplő  $\gamma_n$  értékekhez a  $\sigma^2$  értékét az első zaj-altérbeli sajátértékkel, a  $\lambda_n^{\text{BESZÉD}}$  értékeket a beszéd-altérbeli sajátértéknek a  $\sigma^2$  becslésétől mért eltéréseivel becsültük. A sajátérték-sajátvektor számítását a Jacobi-módszerrel végeztük, a zajcsökkentő eljárást C nyelvű programmal valósítottuk meg.

Az 1. ábrán (lásd az előző oldalon) összefoglaltuk a kapott számszerű adatokat, amik megfelelnek a szakirodalomban közölt eredményeknek [5,10]. A táblázatból kiolvasható, hogy az SRR-ben is megjelenő javulás 6 dB-nél kisebb SNR esetén mutatható ki. Ennek oka véleményünk szerint az, hogy a vizsgált eljárásnak nemcsak zajelnyomó, hanem beszédtorzító hatása is van, megfelelően a beszéddel korrelált maradékjel-összetevő létezéséről a 3. szakaszban leírtaknak. A módszer működőképességét szemlélteti az alábbi, időtartománybeli 2. ábra, mely -3 dB SNR és fehér zaj esetében készült. Amellett, hogy a módszer zajcsökkentő képessége szembeötlő, az ábrán a beszédtorzító hatás is jól követhető.

## 5. Következtetések

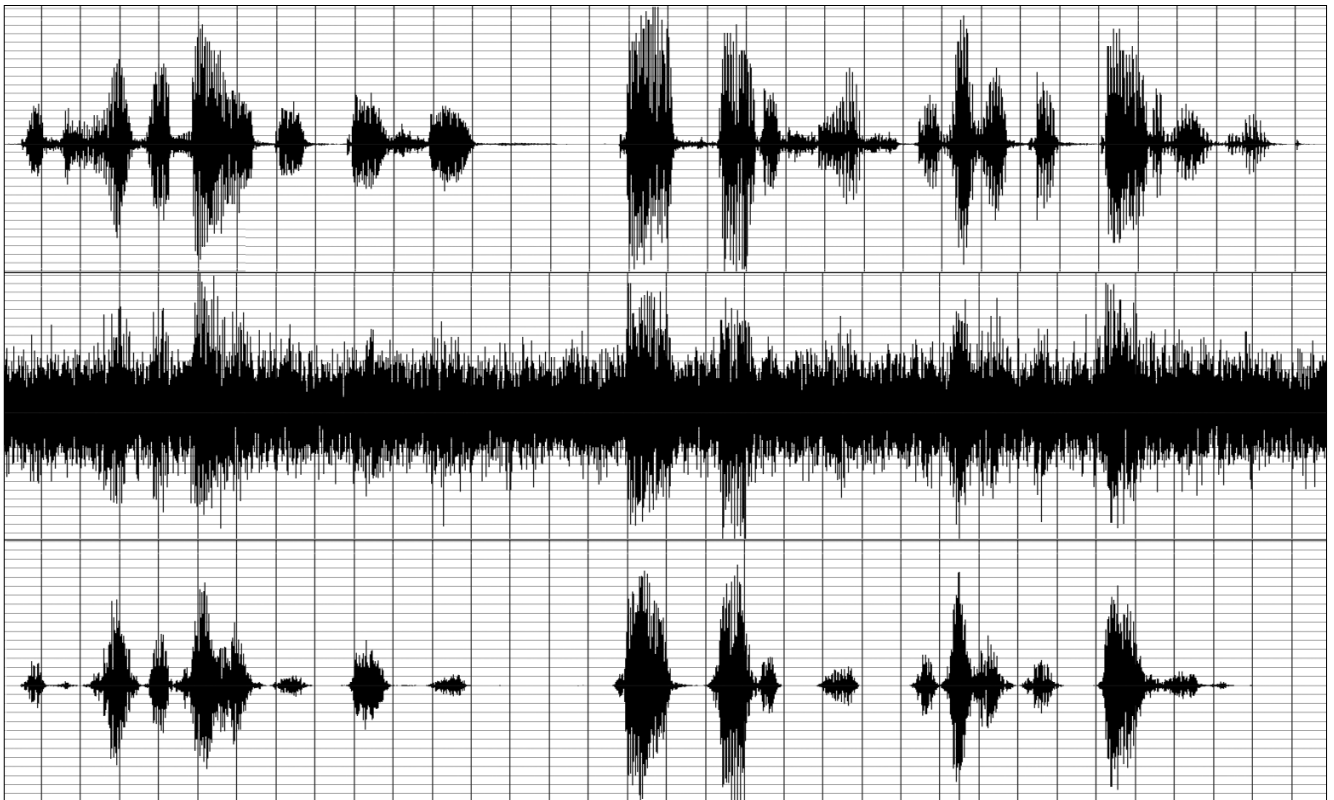
Dolgozatunkban rekonstruált fázistérben működő zajcsökkentő eljárást ismertettünk.

Az eljárás a dimenzióbeágyazásra épül és feltételezi, hogy a beszéd és a zaj altér elfogadhatóan válaszható szét a dimenzióbeágyazás után kapott adatrendszerből számítható optimális ortonormált bázis által kifeszített euklideszi térben. Az említett ortonormált bázist az adatrendszer kovarianciamátrixának sajátvektorai alkotják, amit a Jacobi-eljárással számítottunk ki. Az adatrendszert a tisztítandó beszédsegmentek periodikus kiterjesztésével alkottuk meg, eltérően a szakirodalomban található megoldásoktól. A mi módszerünk nem igényel tapasztalati súlyozómátrixot a beszédminta becslésekor.

A programot egy magyar mondat zajosításával kapott zajos beszéddel teszteltük háromféle zaj és hétféle zajszint esetén. A javulást számszerűen is jellemeztük, a paramétereket a tisztított beszéd meghallgatásával állítottuk be. Megállapítható, hogy a legjobb eredményt körülbelül 100 ms hosszúságú szegmens, 50%-os szegmensléptetés, Hanning-ablak, átlapolás és hozzáadás típusú szegmentált feldolgozás, 20 dimenziós beágyazott tér, 1 minta beágyazási időlépés és 7 dimenziós beszéd-altér esetén értük el.

Ezek a vonatkozó szakirodalomban jelenleg megtalálható adatoknak jól megfelelnek, numerikus kísérleteink alapján azt is mondhatjuk, hogy nemcsak fehér zaj, hanem a nagyfrekvenciás csatorna zaja és rózsaszín zaj esetén is. Ugyanakkor a közölt módszer fehér zajra

Zajcsökkentés -3 dB SNR és fehér zaj esetében a táblázatban leírt paraméterek mellett  
(felül: eredeti bemondás, középen: zajosított beszédminta-sorozat, alul: a zajcsökkentés utáni minta-sorozat)



kidolgozott, a színes zaj elnyomása nem optimális, ahhoz fehérítő transzformáció beépítése is szükséges.

További feladat a beágyazási dimenzió, az időeltolás és a beszéd-altér dimenzió értékeinek automatikus meghatározása, valamint ezek birtokában a zajcsökkentő eljárás módszeres tesztelése nagy beszéd-adatbázison.

### Köszönetnyilvánítás

A szerző ez alkalommal is megköszöni Gordos Gézának, Németh Gézának és Tatai Péternek a segítséget és biztatást, amit beszédfeldolgozási célú algoritmusfejlesztési munkái során kapott.

### A szerzőről

**Pintér István** 1983-ban kapta meg okleveles villamosmérnöki diplomáját a BME Villamosmérnöki Karán, majd ugyanitt PhD fokozatot szerzett a műszaki tudomány informatika tudomány szakán 1997-ben. Az egyetem elvégzése után 1 évig a MIKI-ben dolgozott, majd 1984-től kezdve a GAMF-on, illetve 2000-től ennek jogutódján a KF GAMF Karán, jelenleg főiskolai tanárként. Fő érdeklődési területe a gépi beszédfeldolgozás, a digitális jelfeldolgozás és a jelfelismerés. Az első területen új beszédreprezentációk kidolgozásával és a beszédjel zajból való kiemelésével foglalkozik, a második témakörön belül elsősorban diszkrét ortogonális transzformációk jelfeldolgozási alkalmazásaival, a sort a mesterséges neurális hálózatok jelfelismerési célú alkalmazásai zárják. Az említett szakterületeken cikkei jelentek meg hazai és külföldi folyóiratokban, konferencia-kiadványokban. 2007-ben elnyerte a HTE Pollák-Virág díját.

### Irodalom

- [1] J.S. Lim, A.V. Oppenheim:  
Enhancement and bandwidth compression of noisy speech.  
Proc. IEEE 67 (12) 1979., pp.1586–1604.
- [2] Yang Gui, Kwan, H.K.:  
Adaptive subband Wiener filtering for speech enhancement using critical-band gammatone filterbank.  
Proc. 48th Midwest Symposium on Circuits and Systems, 2005, Vol. 1., pp.732–735.
- [3] Haci Tasmaz, Ergun Ercelebi:  
Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE-STSA estimation in various noise environments.  
Digital Signal Processing,  
(p.16, in press, available online 12 October 2007)
- [4] T.F. Quatieri, R.B. Dunn:  
Speech enhancement based on auditory spectral change.  
Proc. International Conference on Acoustics, Speech and Signal Processing in Orlando, IEEE, 13-17 May 2002, pp.257–260.
- [5] J. Sun, N. Zheng, X. Wang:  
Enhancement of Chinese speech based on nonlinear dynamics.  
Signal Processing 87 (2007), pp.2431–2445.
- [6] Y. Ephraim, H.L. Van Trees:  
A signal subspace approach for speech enhancement.  
IEEE Trans. on Speech and Audio Processing, July 1995, Vol. 3., No.4., pp.251–266.

- [7] H. Kantz, T. Schreiber:  
Nonlinear Time Series Analysis.  
Cambridge University Press, 1997.
- [8] G. Kubin, C. Lainscsek, E. Rank:  
Identification of Nonlinear Oscillator Models for Speech Analysis and Synthesis.  
In: Chollet et al. (Eds.): Nonlinear Speech Modeling. LN AI 3445, Springer Verlag 2005., pp.74–113.
- [9] I. Pintér:  
Noise suppression using non-linear speech model.  
Pollack Periodica, Vol. 2. Suppl., 2007.  
Akadémiai Kiadó, pp.121–133.
- [10] M.T. Johnson, R.T. Pavinelli:  
Generalized phase space projection for nonlinear noise reduction.  
Physica D 201 (2005), pp.306–317.
- [11] [http://spib.rice.edu/spib/select\\_noise.html](http://spib.rice.edu/spib/select_noise.html)