

## Zajszűrő eljárások alkalmazása, teljesítményük vizsgálata zajos beszéd automatikus felismerésénél

Sztahó Dávid, Szaszák György, Vicsi Klára

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Távközlési és Médiainformaticai Tanszék, Beszédaakusztikai Laboratórium  
1111, Budapest, Stoczek utca 2.  
sztaho@tmit.bme.hu, szaszak@tmit.bme.hu, vicsi@tmit.bme.hu

**Kivonat:** A jelen cikk célja több zajszűrő eljárás teljesítményének összehasonlítása autók belső terében történő automatikus beszédfelismeréskor. A kutatást autóban felvett hanganyagon végeztük el német nyelvre. A zajszűrő eljárások teljesítményének összehasonlítását egy csatornán végeztük. Négyféle zajszűrő eljárást vizsgáltunk: Spectral Subtraction, Wiener-filter, Minimum Mean-Square Error Log-Spectral Amplitude Estimator, valamint modulációs spektrum szűrésén alapuló zajcsökkentés. Minden eljárásnál használtunk egy felüláteresztő szűrőt is, amellyel az autó mélyfrekvenciás zaját tudtuk kiküszöbölni. A beszédfelismerési tesztekre Rejtett Markov-modell alapú felismerőt használtunk. A tesztsorozatokot két részre bontottuk. Az első tesztsorozat során megvizsgáltuk az egyes zajszűrő eljárások alkalmazhatóságát a beszédfelismerésben úgy, hogy a TELEAUTO személygépkocsi belterében rögzített hanganyagot használtuk tanításra és tesztelésre is, az adott zajszűrő eljárás alkalmazása után. A második tesztsorozat során pedig megvizsgáltuk, hogy a SpeechDat adatbázissal betanított HMM modellekkel a szűrt felvételek milyen eredményeket adnak a szűretlen személygépkocsikban rögzített felvételekkel történő felismeréshez képest. A kapott eredmények azt mutatják, hogy a zajszűrő eljárások közül az MMSE adja a legjobb felismerési százalékot az általunk vizsgált módszerek közül. Továbbá a teszteredményekből az is egyértelmű, hogy a felismerés szempontjából az a leghatékonyabb eljárás, ha a zajos beszédfelismerésnél a hasonló zajban felvett beszédadatbázissal történik a betanítás.

### 1 Bevezetés

Napjainkban a beszédjelek kiemelése a zajos környezetből, vagyis a zajjal terhelt beszédjelek javítása kiemelt kutatási téma. Ennek oka a számos felhasználási lehetőség, amellyel egy hatékony beszéd kiemelő rendszer rendelkezik. Ma már számos technológiai eljárás létezik ennek megvalósítására. Általánosságban elmondható, hogy a mai kifejlesztett beszédfelismerők megcélzott felhasználási környezete alacsony zajszintű. Éppen ezért egy ilyen felismerő alkalmazása zajos környezetben akkor lehetséges, ha a felismerő bemenetére már zajszűrt beszédjel kerül, vagy a felismerő akusztikai előfeldolgozó eljárását zajtűrő eljárásra cseréljük. A zajszűrés külön problémát jelent különösen olyan esetekben, amikor változó, hol állandó, hol

impulzusszerű, valamint változó hangszintű és -színezetű zaj váltakozva van jelen, mint például gépkocsik belső terében.

A jelen cikk célja több zajszűrő eljárás teljesítményének összehasonlítása autók belső terében történő, Rejtett Markov-modelleken alapuló automatikus beszédfelismeréskor.

Az egycsatornás zajszűrő eljárások összehasonlításánál az alábbi eljárásokat vizsgáltuk: Spectral Subtraction [2], Wiener-filter [3], Minimum Mean-Square Error Log-Spectral Amplitude Estimator [6][7], valamint modulációs spektrum szűrésén alapuló zajcsökkentés [1]. Miután a személygépkocsikban felvett zajos beszédet a különböző zajszűrő eljárásokkal megszürtük, az anyagon beszédfelismerési tesztekét végeztünk. A felismerési feladatok között német nyelvű információlekérés, rövidebb megerősítés jellegű mondatok, valamint hosszabb általános mondatok felismerése szerepelt. Mivel a rendelkezésre álló TELEAUTO-német adatbázis eredeti, zajos, német nyelvű felvételeit magyar anyanyelvű személyek mondták be, ezért az így betanított felismerő német anyanyelvű személyek bemondásainak felismerésére csak korlátozottan lesz alkalmas. Ezért a tesztek során két fázist különítettünk el. A tesztsorozatok első fázisában a TELEAUTO-német adatbázist használva megvizsgáltuk, hogy az egyes zajszűrő eljárások alkalmazása a felismerések során hogyan teljesít az eredeti zajos felvételekkel történő betanításhoz és felismeréshez képest. Ehhez mind a tanítás, mind pedig a tesztelés felvételeit zajszűrtük, majd beszéd-felismerési kísérleteket folytattunk.

A második fázisban az anyanyelvi német mobil telefon beszédet tartalmazó SpeechDat(II) adatbázissal [4] végeztük a HMM modellek betanítását, a tesztek pedig most is az autóban készült felvételek zajszűrt változataival történtek. A kapott eredmények megmutatták, hogy a szűrt felvételek milyen felismerési teljesítményt adnak a szűretlen felvételekkel történő felismeréshez képest.

A cikk részeiben először bemutatjuk a használt vizsgálati módszert, az alkalmazott adatbázisokat. Majd rövid áttekintést adunk az általunk alkalmazott zajszűrő eljárásokról. Utána pedig a két tesztsorozat módszerét, valamint azok eredményeit ismertetjük.

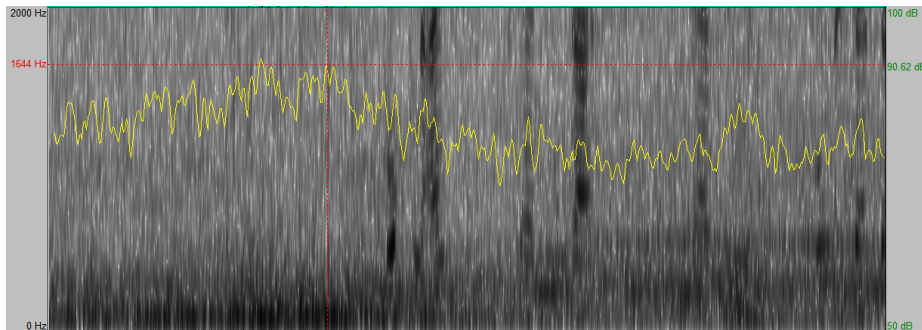
## 2 Vizsgálati módszer leírása

A kísérletek során két adatbázist használtunk fel:

A TELEUTO projekt kapcsán elkészült német nyelvű, autók belső terében készült hanganyag (TELEAUTO-német) négy mikrofonnal lett (2 szimmetrikusan elhelyezkedő vezetőülés melletti, 1 középső tükörnél lévő, és 1 headset mikrofon) rögzítve. Navigációs rendszereknél általában használt 400 darab rövid, megerősítés jellegű mondatot, 1386 rövid kérést kifejező mondatot, valamint 1432 hosszabb mondatot tartalmaz. A hanganyagot magyar anyanyelvű személyek mondták be német nyelven. A felvételek az autó haladása közben készültek, változó zajos környezetben, 48kHz-es sávszélességgel, 16 bites számábrázolással.

A jelen vizsgálatban csak egymikrofonos, a középső tükörnél lévő mikrofonnal készült felvételeket használtuk. A felvételek során a jel-zaj viszony -10 és 10 dB között változott. Gyakran előfordult, hogy a zaj átlagos intenzitás szintje nagyobb volt, mint

a beszédé. Az 1. ábrán egy zajos felvétel látható az adatbázisból. A piros vonalak egy tisztán zaj részt mutatnak. Jól látható, hogy ennek intenzitása (sárga görbe) nagyobb, mint ami a későbbi beszédnél tapasztalható.



**1. ábra.** Egy felvétel spektrogramja a TELEAUTO-német adatbázisból.

A másik adatbázis a német nyelvű SpeechDat(II)-német adatbázis [4] volt, amely közel 5000 vezetékes telefonfelvételt és 1400 mobilhálózaton keresztüli beszélgetés felvételét tartalmazza, amelyek a telefon sávszélességének megfelelően 8000 Hz-es sávszélességűek és 16 bites számábrázolásúak. Ez a hanganyag mobilhálózaton keresztül már anyanyelvi német beszélőkkel készült, de nem személygépkocsikban került rögzítésre és ez igen lényeges akusztikai különbség a két adatbázis felvételei között. A tesztelésnél, a következtetések levonásánál ezt mindenképp figyelembe kell venni. Sajnos csak ilyen adatbázisok álltak a rendelkezésünkre a vizsgálathoz.

A felismerést Rejtett Markov Modell alapú felismerővel valósítottuk meg, amelyre a Sphinx szoftvert [5] használtuk. A modellek felépítéséhez az előfeldolgozás során a 8 kHz-es felvételeket 130Hz és 3700Hz közötti tartományra szűrtük, majd 25 ms-os Hamming ablakolást követően 512 pontos FFT-t számítottunk. A spektrumot kritikus sávok szerint szűrve MFC együtthatókká transzformáltuk, tehát a „klasszikus” 39 elemű jellemzővektorok lettek létrehozva (13 MFCC együttható, valamint ezek első és második deriváltja) 10 milliszekundumos kereteltolással. Ezekből 16 Gauss-keveréssel 3 állapotú trifón – tehát környezetfüggetlen – beszédhangmodellek készültek.

A tesztelésnél a 3. pontban részletezett zajszűrő eljárások alkalmazása után végzett felismerési kísérleteket hasonlítottuk össze egymással és a szűretlen zajos beszédfelvételekkel kapott felismerési eredményekkel.

### 3 Zajszűrő eljárások

Az autó belső terében készült zajos felvételek minőségének javításához, a beszédjel kiemeléséhez, a beszéd érthetőbbé tételéhez a következőkben szereplő zajszűrő eljárásokat alkalmaztuk. Minden eljárásnál használtunk egy felüláteresztő szűrőt is, amellyel az autó mélyfrekvenciás zaját tudtuk kiküszöbölni.

### 3.1 Wiener-szűrő [3]

A Wiener-szűrők központi szerepet játszanak számos alkalmazásban, például lineáris predikció, jelkódolás, visszhang kioltás, jelvisszaállítás és csatornaki egyenlítés megoldásaiban. A Wiener szűrőt úgy számoljuk, hogy a szűrő kimenete és a kívánt jel átlagos négyzetes távolsága minimális legyen. Általános esetben a szűrő azt feltételezi, hogy a jelek stacionárius folyamatok, ám a szűrő együtthatóinak időnkénti újraszámolásával adaptívvá tehető a jel karakterisztikájához. A Wiener-szűrő véges impulzus válaszu (FIR) szűrőként való megvalósítása adott számú lineáris egyenletet ad, amelyeknek létezik zárt alakú megoldása. A 2. ábra a Wiener-szűrőt ábrázolja, a  $\mathbf{w}$  együttható vektorral, az  $\mathbf{y}(m)$  bemenő jellel, és az  $\hat{x}(m)$  kimenő jellel, amely  $\hat{x}(m)$  a kívánt cél  $x(m)$  legkisebb átlagos négyzetes becslése. A szűrő bemeneti-kimeneti összefüggése:

$$\hat{x}(m) = \sum_{k=0}^{P-1} w_k y(m-k) = \mathbf{w}^T \mathbf{y} \quad (1)$$

A becslt és a cél jel közötti különbségből adódó hiba:

$$\mathbf{e}(m) = x(m) - \hat{x}(m) = x(m) - \mathbf{w}^T \mathbf{y} \quad (2)$$

A legkisebb négyzetes hibájú Wiener-szűrőt a következő egyenlet alapján kapjuk meg:

$$\mathbf{R}_{yy} \mathbf{w} = \mathbf{r}_{yx} \quad (3)$$

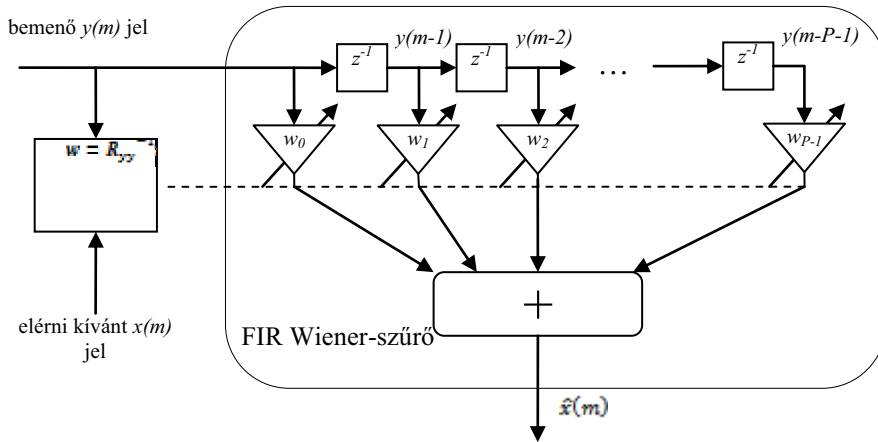
ahol szükségünk van a bemenő jel autókorrelációs mátrixára, valamint a bemenő és a kívánt jel keresztkorrelációs vektorára. Mivel az additív zajjal terhelt jel felírható  $\mathbf{y}(m) = \mathbf{x}(m) + \mathbf{n}(m)$  formában, a jel és a zaj korrelálatlanságából adódóan az autókorrelációs mátrix és a keresztkorrelációs vektor felírható

$$\mathbf{R}_{yy} = \mathbf{R}_{xx} + \mathbf{R}_{nn}, \text{ és} \quad (4)$$

$$\mathbf{r}_{xy} = \mathbf{r}_{xx} \quad (5)$$

formában, ahol az  $\mathbf{R}_{yy}$ ,  $\mathbf{R}_{xx}$ ,  $\mathbf{R}_{nn}$  a zajos jel, a zajmentes jel és a zaj autókorrelációs mátrixai,  $\mathbf{r}_{xy}$  pedig a zajos jel és a zajmentes jel keresztkorrelációs vektora. A (3), (4) és (5) egyenletekből a következő összefüggés adódik a Wiener-szűrő meghatározására:

$$\mathbf{w} = (\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-1} \mathbf{r}_{xx} \quad (6)$$



2. ábra. A Wiener-szűrő felépítésének illusztrációja.

### 3.2 A „Spectral Subtraction” eljárás [2]

A spectral subtraction egy olyan eljárás, amely alkalmas egy additív zajban megfigyelt jel teljesítmény-spektrumának, illetve magnitúdó-spektrumának visszaállítására a zajos jelből. A zaj spektrumát becsülni lehet olyan időszakokból, amikor nincs értékelhető jel, csupán a zaj van jelen. A spectral subtraction számításigénye kicsi, ám a zaj hirtelen változásai negatív teljesítmény-, illetve magnitúdó-spektrumot eredményezhetnek, amelyeket nemnegatív tartományba kell átranszformálni. Ezen nemlinearitás a jelet torzítja.

A zajos jelet leírhatjuk  $y(m) = x(m) + n(m)$ -ként az időtartományban, ahol  $y(m), x(m), n(m)$  a zajos jel, a jel és az additív zaj. Frekvenciatartományban a  $Y(f) = X(f) + N(f)$  összefüggéssel írható le, ahol  $Y(f), X(f), N(f)$  a zajos jel, az eredeti jel és a zaj Fourier-transzformáltjai. Az eljárás blokkdiagramja a 3. ábrán látható, ahol a spectral subtraction-t megvalósító egyenlet:

$$|\widehat{X}(f)|^b = |Y(f)|^b - \alpha |N(f)|^b \quad (7)$$

A negatív teljesítmény- és magnitúdóspektrum kiküszöbölése érdekében egy utófeldolgozási lépést teszünk az inverz Fourier-transzformáció elé:

$$|\widehat{X}(f)| = \begin{cases} |\widehat{X}(f)|, & \text{ha } |\widehat{X}(f)| > \beta |Y(f)| \\ f_n[|Y(f)|], & \text{különben} \end{cases} \quad (8)$$



3. ábra. A spectral subtraction eljárás blokkdiagramja.

### 3.3 A „Minimum Mean-Square Error Spectral Amplitude Estimator”(MMSE) eljárás [6][7]

A spectral subtraction módszeren alapuló eljárások esetén a rövid távú spektrális amplitúdó (STSA) a jel spektrális komponenseinek varianciájának maximum likelihood becslő négyzetgyökeként adódik. A Wiener-szűrőknél az STSA becslő a jel spektrum-komponenseinek optimális legkisebb átlagos négyzetes hibájaként kapható meg. Mivel mindkét STSA becslő eljárás adott körülmények között kínál optimális megoldást, egyik sem általános optimális spektrális amplitúdó becslő. Ezért a következő eljárás, amelyet kipróbáltunk olyan, amely az STSA becslést közvetlenül a zajos megfigyelésből nyeri.

Jelölje  $x(t)$  és  $d(t)$  a beszéd és a zaj folyamatát, valamint  $y(t)$  a megfigyelt jelet.

$$y(t) = x(t) + z(t) \quad (9)$$

Jelölje  $X_k \triangleq A_k \exp(j\alpha_k)$ ,  $D_k$  és  $Y_k \triangleq R_k \exp(j\vartheta_k)$  a beszédjel, a zaj és a megfigyelt jel  $k$ . spektrális összetevőjét.  $Y_k$  megadható a következőképpen is:

$$Y_k = \frac{1}{T} \int_0^T y(t) \exp\left(-j \frac{2\pi}{T} kt\right) dt, \quad k = 0, \pm 1, \pm 2 \dots \quad (10)$$

Az eljárás célja az  $A_k$  tényező becslése:

$$\hat{A}_k = E(A_k | Y_k) = \Gamma(1,5) \frac{\sqrt{v_k}}{\gamma_k} \exp\left(-\frac{v_k}{2}\right) \left[ (1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] R \quad (11)$$

ahol  $\Gamma(\cdot)$  jelöli a gamma függvényt,  $I_0$  és  $I_1$  jelöli a nullad-, és elsőrendű módosított Bessel-függvényeket,  $\lambda_d(k) = E\{|D_k|^2\}$  a zaj varianciája,  $v_k$  definíciója pedig

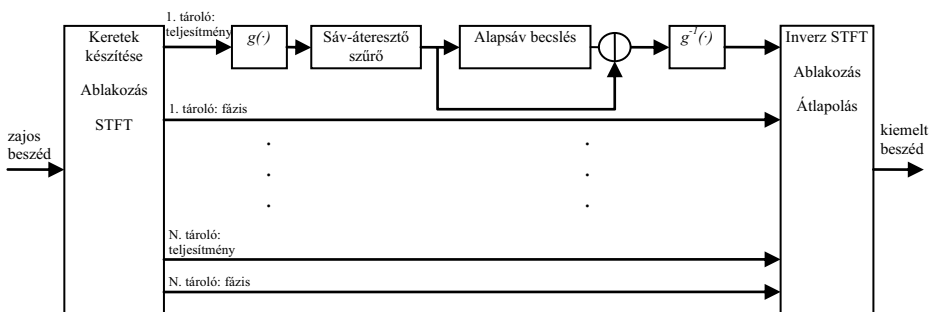
$$v_k = \frac{\xi_k}{1 + \xi_k} \gamma_k, \quad \text{ahol } \xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \text{ és } \gamma_k = \frac{R_k^2}{\lambda_d(k)} \quad (12)$$

### 3.4 Modulációs spektrum alapú eljárás [1]

A modulációs spektrum a teljesítményspektrum (vagy annak egy tömörített változatának) komponenseiből felépített időbeli sorozatnak Fourier-transzformáltja. A kü-

lőnböző modulációs spektrumok beszédérthetőségre gyakorolt hatását számos kutató vizsgálta, és általánosságban megállapították, hogy az 1 Hz és 16 Hz közötti modulációs spektrum játssza a legnagyobb szerepet az érthetőségben.

A 4. ábra illusztrálja a modulációs spektrum szűrésén alapuló zajszűrési eljárást. Az  $x(t)$  bemenő jelet keretenként 20 ms-os Hamming ablakkal és 10 ms-os időléptéssel elemzik. Jelölje  $X_k(f)$  a short-time Fourier-transzformációt az  $f$ . frekvenciatárolóban. Definiáljuk a rövid távú teljesítményspektrumot  $P_k(f) = |X_k(f)|^2$  -ként. Legyen  $N$  az FFT tárolók száma, valamint  $K$  a zajos beszéd megfigyelések keretszáma. Ekkor  $P_k(f), k = 1, \dots, K$  jelöli az  $f$ . frekvenciatároló idősorozatát. A modulációs szűrés során sáváteresztő szűrőt alkalmazunk a  $g(P_k(f)), k = 1 \dots K$  idősorozatra  $f = 1 \dots N$  szerint. A 301 hosszúságú Parks-McLellan módszerrel tervezett FIR szűrőt 1 és 16 Hz közötti tartományon alkalmaztuk. A  $g(\cdot)$  egy tömörítő függvény, amelyet a teljesítményspektrumon alkalmaztunk a dinamika tartomány csökkentése érdekében. A tömörített időbeli burkológörbe alapsávú komponensét becsüljük a sávszűrt komponensből. A tömörítő függvény  $g^{-1}(\cdot)$  inverzét alkalmazzuk a tömörítő hatás visszaállítására. A kapott közelítő burkolót felhasználva áll elő a rövid távú magnitúdóspektrum. Ezt a módosított magnitúdóspektrumot és az eredeti rövid távú fázisspektrumot használják fel a feljavított jel előállításához inverz FFT, ablakozás és átlapolás segítségével.



4. ábra. A modulációs szűrésen alapuló beszédkiemelő eljárás blokkdiagramja.

## 4 Zajszűrési eljárások kiértékelése

### 4.1 Tesztelési eljárások

A teszteket két adatbázissal, a TELEAUTO-német és a SPEECHDAT(II)-német adatbázissal végeztük, és a két adatbázis együttes kiértékeléséből vontuk le a következtéseinket. Az autó zajának kiküszöböléséhez a zajszűrő eljárás mellett még egy felüláteresztő szűrőt is alkalmaztunk, a mélyfrekvenciák eltávolítása érdekében. Így a fentebb bemutatott zajszűrő eljárásokat össze tudtuk hasonlítani aszerint, hogy beszédfelismerésre mennyire alkalmasak.

Az első fázisban a kísérletek során a TELEAUTO-német adatbázisa került felhasználásra. Az egyes zajszűrő eljárások tesztelése során az adatbázis három különböző hosszúságú mondatot tartalmazó részét két részre bontottuk és részenként 70%-ot használtunk a felismerő betanítására, a maradék 30%-ot pedig a felismerési hanganyagra. A tanítás előtt a felvételeken elvégeztük az adott zajszűrést, majd a tanítást már ezekkel a mintákkal kezdtük el.

Minden szűrő eljárást a változó zajhoz adaptálva alkalmaztunk a mondatok előtt vagy után található, csak zajból álló jel felhasználásával.

Az anyanyelvi német autós beszéd megfelelő biztonságú felismerése anyanyelvi német beszédatadabázis használatát követeli meg. Ezért a tesztelések második fázisában a tanításnál az anyanyelvi német SpeechDat(II) adatbázist használtuk fel. A teszteléskor továbbra is az előző fázisban felhasznált, autóban felvett hanganyaggal megegyező mintahalmazt használtunk már a megfelelő zajszűrő eljárás lefuttatása után. Az anyanyelvi különbségekből adódó alapvető kiejtésbeli eltérés miatt ezeknél a felismerési kísérleteknél az abszolút felismerési teljesítmény nem lehet mérvadó. Ezért itt mindig az eredeti zajos felvételekkel történő felismeréshez viszonyítottuk a különböző zajszűrő eljárásokkal kapott felismerési eredményt.

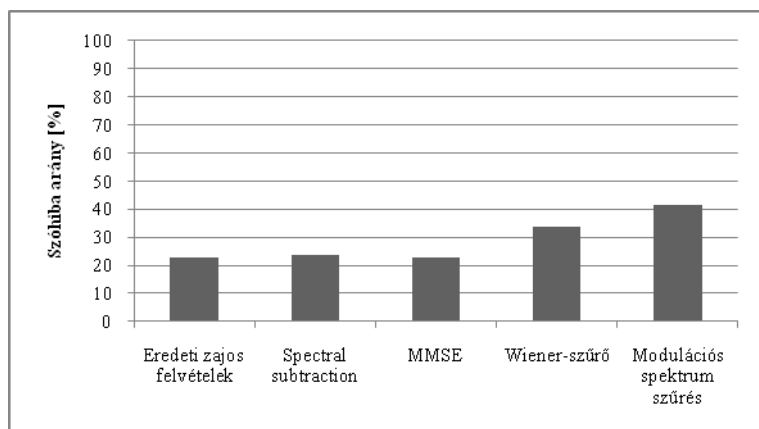
#### 4.2 Beszédfelismerési eredmények

Az 1. táblázatban és 5. ábrán az első teszt sorozat eredményei láthatóak. Az 1. táblázat részletesen tartalmazza a TELEAUTO adatbázis három mondat típusával kapott felismerést. Az elvártaknak megfelelően az egyik legjobb felismerési eredményt akkor kaptuk, ha a betanítás és a tesztelés is az eredeti, zajos felvételekkel történt. Ehhez képest legjobb zajszűrők által elért felismerést az MMSE Spektrális Amplitúdó Becslő, valamint Spectral Subtraction eljárás adta, amelyek csupán legfeljebb 1%-kal tértek el az eredeti, zajos felvételekkel történő tesztek eredményétől, valamint közel 10%-kal jobb felismerést produkáltak, mint a további két zajszűrő módszer. Ezek a módszerek tehát egy felismerési feladat során alkalmazhatónak adódtak.

1. táblázat: Az első fázis teszt sorozataiban kapott felismeréseinek szóhiba arányai százalékban.

|                            | Rövid, megerősítés jellegű mondatok | Kérés jellegű mondatok | Hosszú, általános mondatok | Átlagos felismerés |
|----------------------------|-------------------------------------|------------------------|----------------------------|--------------------|
| Eredeti zajos felvételek   | 18,5                                | 15,3                   | 34,4                       | 22,7               |
| Spectral subtraction       | 22,6                                | 15,1                   | 33,3                       | 23,7               |
| MMSE                       | 17,8                                | 16                     | 34,1                       | 22,6               |
| Wiener-szűrő               | 35,6                                | 19,5                   | 46,7                       | 33,9               |
| Modulációs spektrum szűrés | 37,7                                | 28                     | 59,1                       | 41,6               |



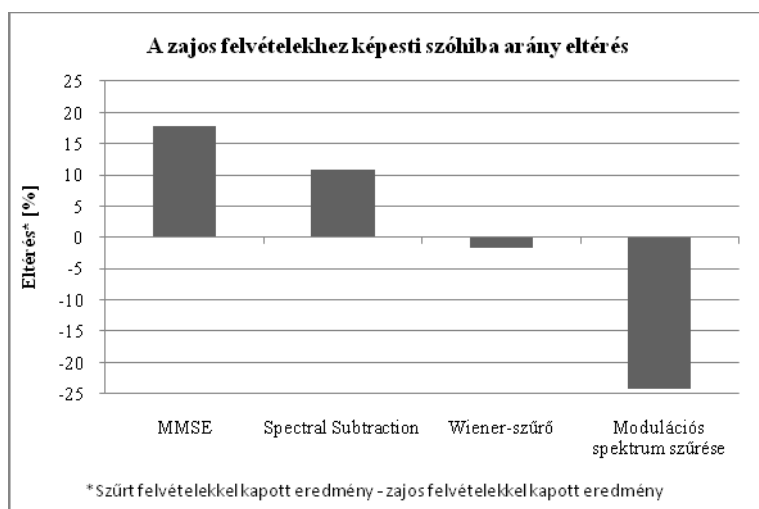


**5. ábra.** A TELEAUTO adatbázissal készített felismerés eredményei: az eredeti zajos felvételeket, valamint az adott zajszűrő eljárást alkalmazva a tanításra és tesztelésre

A 6. ábrán a második tesztsorozat eredményei láthatóak. Az ábra a szűretlen hangmintákkal végzett felismerési teljesítményhez képesti javulást vagy romlást mutatja az egyes zajszűrő eljárások esetében. A 2. táblázatban részletesen láthatóak az egyes mondat típusokkal kapott eredmények. Látható, hogy azok az eljárások, amelyek az első tesztsorozat esetében jól teljesítettek, itt is hasonló tulajdonságot mutatnak, de itt az MMSE kicsit kiemelkedik a többi közül.

2. táblázat: A német nyelvű SpeechDat adatbázissal végzett tanítás és a TELEAUTO adatbázissal végzett tesztelés során az adott zajszűrő eljárás és az eredeti zajos felvételek felismerési szóhiba arányai közötti eltérés százalékban.

|                                    | Rövid,<br>megerősítés<br>jellegű<br>mondatok | Kérés<br>jellegű<br>monda-<br>tok | Hosszú,<br>általános<br>mondatok | Átlag-<br>os<br>eltérés |
|------------------------------------|----------------------------------------------|-----------------------------------|----------------------------------|-------------------------|
| <b>MMSE</b>                        | 32,2                                         | 14,1                              | 7,3                              | <b>17,8</b>             |
| <b>Spectral Subtraction</b>        | 31,5                                         | 2,1                               | -0,9                             | <b>10,9</b>             |
| <b>Wiener-szűrő</b>                | 10,9                                         | -9,1                              | -6,5                             | <b>-1,6</b>             |
| <b>Modulációs spektrum szűrése</b> | -25,4                                        | -39,2                             | -7,9                             | <b>-24,2</b>            |



**6. ábra.** A német nyelvű SpeechDat adatbázissal végzett tanítás és a TELEAUTO adatbázissal végzett tesztelés során az adott zajszűrő eljárás és az eredeti zajos felvételek felismerési teljesítménye közötti eltérés.

A James G. Lyons, Kuldip K. Paliwal-féle modulációs spektrum alapú eljárás a vártnál lényegesen rosszabb felismerési eredményt adott mind a két adatbázissal végzett betanítás esetén. Annak ellenére, hogy az irodalomban lényeges szubjektív érthetőségnövekedésről számolnak be a kutatók e szűrés alkalmazása esetén.

## 5 Összefoglalás

A cikkben olyan zajszűrő eljárásokat hasonlítottunk össze, amelyek alkalmasak additív zaj szűrésére, a hasznos jel (beszéd) kiemelésére. Az összehasonlítás során személygépkocsiban felvett időben változó zajkörnyezetű folyamatos beszéd felismerését vizsgáltuk.

Az eljárások két fázisban kerültek tesztelésre. Az első fázisban ugyanazon személygépkocsi belső terében felvett hanganyaggal történt a betanítás is és a tesztelés is, a TELEAUTO-német adatbázissal. Ezekben az esetekben a szóhibaarány 20-25% közöttinek adódott a -10 és 10 dB közötti jel-zaj viszony határok között. Ez azt mutatja, hogy szűrés nélkül is a legjobb szűréssel kapott eredményhez közeli elfogadható eredményt kapunk abban az esetben, ha az akusztikus modell betanítása hasonló zajos körülmények között történik, mint ami a felismeréskor is előfordul.

Abban az esetben, amikor nem áll rendelkezésre megfelelő zajos adatbázis a betanításhoz, amit a kísérletben a SPEECHDAT(II) német adatbázis használatával modelleztünk, egyes zajszűrő eljárások a 4. fejezetben tárgyalt kísérletek szerint adaptív módon sikerrel alkalmazhatók személygépkocsiban adódó zajkörnyezetben.

A legjobban teljesítő zajszűrő eljárás a Minimum Mean-Square Error Spectral Amplitude Estimator (MMSE) volt, amelynek használatával közel 18%-kal adott jobb

felismerési eredményt, mint az eredeti zajos felvételek esetén. Az egyes módszerek egy beszéddetektor segítségével automatikusan adaptívvá tehetőek.

## **Köszönetnyilvánítás**

Ez a kutatás a Jedlik OM-00102/2007 számú "TELEAUTO" projekt keretén belül készült.

## **Hivatkozások**

1. Lyons J. G., Paliwal K. K.: Effect of Compressing the Dynamic Range of the Power Spectrum in Modulation Filtering Based Speech Enhancement. *Interspeech 2008* (2008) 387–390
2. Berouti, M., Schwartz, R., Makhoul, J.: Enhancement of Speech Corrupted by Acoustic Noise. *IEEE ICASSP* (1979) 208–212
3. Vaseghi, S. V.: *Advanced Signal Processing and Digital Noise Reduction*. Wiley & Teubner Communications (1996)
4. SpeechDat.: <http://www.speechdat.org/>
5. Sphinx.: <http://cmusphinx.sourceforge.net/html/cmusphinx.php>.
6. Yariv, E., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (1984) 1109–1121
7. Yariv, E., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (1985) 443–446