Beszédjelek pillanatnyi jellemzőinek becslése a Teager-operátorral és a Hilbert-Huang-transzformációval

PINTÉR ISTVÁN

Kecskeméti Főiskola GAMF Kar, Automatizálási és Alkalmazott Informatikai Tanszék
pinter.istvan@gamf.kefo.hu

Lektorált

Kulcsszavak: Teager-operátor, HHT, pillanatnyi amplitúdó és frekvencia, visszaállítás pillanatnyi jellemzőkből

A beszédjelek finomszerkezetének vizsgálatához a nemlineáris és nemstacionárius jellemzők meghatározására szolgáló módszerek szükségesek. Jelen dolgozatban a Teager-operátort és a Hilbert-Huang-transzformációt (HHT) ismertetjük, mint a pillanatnyi amplitúdó és a pillanatnyi frekvencia becslésére alkalmazható jelfeldolgozási eljárást. A HHT-vel előállítható pillanatnyi amplitúdó és pillanatnyi frekvencia paramétereket összehasonlítjuk a Teager-operátorra alapozott becslések eredményeivel mind vizsgálójel, mind beszédjel esetén.

1. Bevezetés

A gépi beszédfeldolgozásban számos feladat megoldásának alapja az úgynevezett kvázi-stacionárius jelmodell. Eszerint a beszédjel feldolgozható úgy, hogy elegendően rövid időtartamú szakaszok egymást időben átfedő sorozatain végezzük az adott feladat megoldása érdekében számításainkat. Feltételezzük, hogy a beszédszakasz időtartama alatt a beszédjel-modell paraméterei nem változnak. Az elegendően rövid időtartamot a hangszalagok nyitási-zárási ütemének megfelelő alapperiódus-idő 2...5-szöröseként határozza meg a szakirodalom, az átfedési idő 1...3 ugyanebben az időegységben [1].

A gépi beszédfeldolgozás fejlődése során felmerült az igény olyan elemző módszerek iránt, amelyekkel az alapperiódus időtartamánál rövidebb idő alatt lejátszódó változások is vizsgálhatók. Az ilyen változások alkotják a beszédjel finomszerkezetét. A nemlineáris módusú hangszalag-rezgés okozta kismértékű alapperiódus-idő ingadozás jelensége – sok egyéb mellett – olyan jelenség, aminek vizsgálatához a finomszerkezet leírására alkalmas módszerek szükségesek. A módszerekkel szembeni elvárás az, hogy néhány beszédmintányi adathoz tudjanak fizikailag is értelmezhető jellemzőket rendelni. Következésképpen erre a célra nem használható a kvázi-stacionárius jelmodell alapján kidolgozott gépi beszédfeldolgozási eszköztár [2].

A probléma lényegét tömören összefoglalva azt mondhatjuk, hogy az időfelbontás növelése a részletes frekvenciakép megtartása mellett nem lehetséges, mert fennáll az időpont és a frekvenciaérték együttes meghatározásának bizonytalanságát összekapcsoló Gábor Dénes-féle határozatlansági reláció, ezáltal a gördülő Fourier-transzformációra (STFT, Short-Time Fourier-Transform) alapozott – vagy azzal kapcsolatba hozható – módszerek a beszédjel finomszerkezetének leírására nem alkalmasak.

Ma már elterjedt a megnövelt időfelbontást igénylő alkalmazásokban a wavelet-transzformáció használata, de a wavelet-es beszédelemzés időfelbontását is korlátozza az, hogy a fentebb említett idő-frekvencia bizonytalanság helyére az idő-skála bizonytalanság lép.

Van olyan beszédábrázolás is, amelynél nincs jelen a határozatlansági reláció okozta korlát – ilyen például a Wigner-Ville-eloszlás vagy a Choi-Williams-eloszlás, ám itt más problémák jelentkeznek a finomszerkezet feltárásakor (például a transzformáltban megjelenő kereszt-tag elnyomása jelent megoldandó feladatot). Ezzel az izgalmas témakörrel jelen dolgozatban nem foglalkozunk, a részleteket [2,3] tartalmazza.

A beszédjel finomszerkezetének elemzésére szolgáló – az előző bekezdésben említettektől lényegesen eltérő – módszer a Teager-operátorra alapozott ES-algoritmus (Energy Separation algorithm) [2], amivel becslés adható a beszédjel pillanatnyi amplitúdójára és pillanatnyi frekvenciájára. A cikk hátralévő részében ezeket együtt pillanatnyi jellemzőknek nevezzük. Az előző bekezdésben foglaltakat is figyelembe véve talán nem meglepő, hogy a wavelet-es elemzés és a Teager-operátor összekapcsolása mára sikeres alkalmazásokhoz vezetett [4].

További lehetőség a pillanatnyi jellemzők meghatározására a Hilbert-Huang transzformáció [5] alkalmazása. Mivel a Teager-operátorra illetve a HHT-re alapozott módszerek összehasonlításáról a számunkra hozzáférhető – nyomtatásban, illetve elektronikusan megjelentetett – beszédfeldolgozási szakirodalomban nem találtunk közölt eredményeket, jelen dolgozatunk témájának ezt választottuk.

2. A Teager-operátor és az ES-algoritmus

2.1. A folytonos idejű Teager-operátor és a pillanatnyi jellemzők becslése

A címben szereplő operátor fogalmának megalkotása és a vele elvégzendő művelet meghatározása az emberi beszédkeltés közben fellépő nemlineáris fizikai jelenségek gondos vizsgálata után vált lehetségessé.

H. M. Teager és S. M. Teager először 1980-ban közöltek ilyen mérési eredményeket, majd 1985-ös publikációjukban a modellalkotásról számoltak be. Kiderült, hogy az alapperiódus-időn belüli gyors jelenergia-változás jelenségének leírásához célszerű meghatározni a jelet előállító rendszer összenergiáját. Ezen összenergia becslését kapjuk meg, ha a jelre egy alkalmasan megválasztott operátor hat – ma ezt az operátort Teager-operátornak nevezzük. A részleteket és a bőséges szakirodalmi forrást [2]-ben találhatjuk meg.

Azt, hogy miként lehet egy rendszer által előállított jelből a rendszer összenergiájára következtetni, a rugóra függesztett test harmonikus rezgőmozgásának példáján szokás bemutatni. Ezt a mozgást másodrendű differenciál-egyenlet írja le, ami ideális esetben a következő alakú:

$$\frac{d^2x(t)}{dt^2}+\frac{k}{m}\cdot x(t)=0\,,$$

ahol *x(t)* a kitérés-idő függvény, *k* a rugóállandó, *m* a harmonikus rezgőmozgást végző test tömege. A differenciál-egyenlet megoldása $x(t)=a\cdot\cos(\omega\cdot t+\varphi)$ alakú – a fentebbi szóhasználat szerint ez a rendszer által előállított jel. A rugóból és a harmonikus rezgőmozgást végző testből álló rendszer összenergiája a rugóban tárolt energia és a mozgási energia összege:

$$E=\frac{1}{2}\cdot k\cdot x^2(t)+\frac{1}{2}\cdot m\cdot\left(\frac{dx(t)}{dt}\right)^2.$$

Behelyettesítés után adódik, hogy $E=m\cdot a^2\cdot\omega^2$, ezáltal ha a kitérés-idő függvényből méréssel meghatározzuk az amplitúdót és a körfrekvenciát, akkor ezek szorzatának négyzete arányos a jelet előállító rendszer összenergiájával. A Kaiser által javasolt általánosítás alapja az, hogy – egy állandó szorzótényezőtől eltekintve – ugyanezt az eredményt kapjuk, ha a következő operátort alkalmazzuk a kitérés-idő függvényre, mint jelre [2]:

$$\Psi\{x(t)\}=\left(\frac{dx(t)}{dt}\right)^2-x(t)\cdot\frac{d^2x(t)}{dt^2}\,, \qquad (1)$$

ahol Ψ{.} a Teager-operátor. A fenti kitérés-idő függvényre alkalmazva a következőképpen számolhatunk:

$$\frac{dx(t)}{dt}=-a\cdot\omega\cdot\sin(\omega\cdot t+\varphi),$$
$$\frac{d^2x(t)}{dt^2}=-a\cdot\omega^2\cdot\cos(\omega\cdot t+\varphi), \qquad (2)$$

amivel

$$\Psi\{x(t)\}=a^2\cdot\omega^2 \qquad (3)$$

adódik. Ellenőrizhető, hogy ugyanezt az eredményt kapjuk, ha az operátort az $x(t)=a\cdot\sin(\omega\cdot t+\varphi)$ jelre alkalmazzuk – amint az várható is. Érdekességképpen megemlítjük még, hogy fennáll:

$$\Psi\left\{a\cdot e^{j(\omega t+\varphi)}\right\}=0\,. \qquad (4)$$

Az $x(t)=a\cdot\cos(\omega\cdot t+\varphi)$ jel egy lehetséges általánosítása az, amikor mind az amplitúdó, mind a fázis időfüggő, az így keletkező AM-FM jel alakja:

$$x(t)=a(t)\cdot\cos(\varphi(t)). \qquad (5)$$

Közvetlen számolással hamar belátható, hogy tetszőleges amplitúdó- és fázis időfüggvény esetén az (1)-ben megadott operátor nehezen kezelhető kifejezéshez vezet. Ám abban az esetben, ha mind az amplitúdó, mind a fázis lassan változik az időben, vagyis amikor fennállnak az alábbi közelítések:

$$\frac{da(t)}{dt}\approx 0,\quad \frac{d\varphi(t)}{dt}\approx\text{állandó},\quad \frac{d^2\varphi(t)}{dt^2}\approx 0, \qquad (6)$$

akkor az (5)-beli AM-FM jelre alkalmazva a Teager-operátort, a következőket kapjuk:

$$\frac{dx(t)}{dt}\approx -a\cdot\frac{d\varphi(t)}{dt}\cdot\sin\varphi(t),$$
$$\frac{d^2x(t)}{dt^2}\approx -a\cdot\left(\frac{d\varphi(t)}{dt}\right)^2\cdot\cos\varphi(t), \qquad (7)$$
$$\Psi\{x(t)\}\approx a^2(t)\cdot\left(\frac{d\varphi(t)}{dt}\right)^2.$$

Alkalmazhatjuk az operátort a jel deriváltjára is, ekkor:

$$\Psi\left\{\frac{dx(t)}{dt}\right\}=\left(\frac{d^2x(t)}{dt^2}\right)^2-\frac{dx(t)}{dt}\cdot\frac{d^3x(t)}{dt^3}\,. \qquad (8)$$

A (6)-beli közelítéseket figyelembe véve az (5)-ben szereplő AM-FM jelre a részletes számítás után adódik, hogy:

$$\Psi\left\{\frac{dx(t)}{dt}\right\}\approx a^2(t)\cdot\left(\frac{d\varphi(t)}{dt}\right)^4. \qquad (9)$$

A kapott közelítések segítségével becslést adhatunk az amplitúdó abszolút értékére, mivel fennáll:

$$\frac{\Psi\{x(t)\}}{\sqrt{\Psi\left\{\frac{dx(t)}{dt}\right\}}}=|a(t)|\,, \qquad (10)$$

valamint a fázis deriváltjának (a pillanatnyi frekvenciának) abszolút értékére:

$$\sqrt{\frac{\Psi\left\{\frac{dx(t)}{dt}\right\}}{\Psi\{x(t)\}}}=\left|\frac{d\varphi(t)}{dt}\right|. \qquad (11)$$

Az (1), (10) és (11) egyenletekkel tehát a jelből becsülhető az időben lassan változó *a(t)* burkoló, és a lassan változó pillanatnyi frekvencia. Ellenőrizhető, hogy az $x(t)=a\cdot\cos(\omega\cdot t+\varphi)$ jelre ezek a becslések megadják az (állandó) amplitúdó és az (állandó) körfrekvencia értékét.

2.2. A diszkrét idejű Teager-operátor és az ES-algoritmus

A gépi számítás alapjául is szolgálhat (1), (10) és (11) megfelelő mintavételezés valamint a differenciálás alkalmas diszkrét közelítése után. Mint numerikus eredményeink mutatják, ez utóbbi célra a Savitzky-Golay-féle 5 pontos simító deriválási algoritmus [6] megfelelő. Ezt a továbbiakban közvetlen számításnak nevezzük.

A diszkrét idejű Teager-operátort a folytonos idejű Teager-operátor (1)-ben megadott alakjából úgy tudjuk származtatni, hogy a differenciálást a $d(n)=x(n)-x(n-1)$ differenciával közelítjük.

Ezzel a diszkrét idejű Teager-operátor alakja a következő lesz:

$$\Psi_D\{x(n)\} = x^2(n) - x(n-1)\cdot x(n+1). \quad (12)$$

Némi számolás után adódik, hogy $x(n)=a\cdot\cos(\omega\cdot n+\varphi)$ mintasorozatra alkalmazva a diszkrét idejű Teager-operátort, az eredmény

$$\Psi_D\{x(n)\} = a^2\cdot\sin^2\omega, \quad (13)$$

ahol ω a digitális körfrekvencia.

A diszkrét idejű Teager-operátor esetén megmutatható, hogy az $x(n)=a(n)\cdot\cos(\varphi(n))$ mintasorozatból kiindulva a lassan változó pillanatnyi jellemzők becslésére a következő összefüggések érvényesek [2]:

$$a(n) \approx \frac{2\cdot\Psi_D\{x(n)\}}{\sqrt{\Psi_D\{x(n+1)-x(n-1)\}}}. \quad (14)$$

$$\omega(n) \approx \arcsin\left(\sqrt{\frac{\Psi_D\{x(n+1)-x(n-1)\}}{4\cdot\Psi_D\{x(n)\}}}\right). \quad (15)$$

A (12), (14) és (15) kifejezésekkel adott számítási eljárást nevezi a szakirodalom ES (Energy Separation)-algoritmusnak.

Az ES-algoritmusnak megvan az az előnye, hogy csak három mintát igényel a becslés meghozatalához, míg a közvetlen számítás a simító deriválás miatt öt mintát használ a becsléshez, ám ez utóbbi esetben nem szükséges az arcsin(.) függvény a digitális körfrekvencia értékének meghatározásához.

3. A Hilbert-Huang-transzformáció és a pillanatnyi jellemzők számítása

Az előző pontban láttuk, hogy a Teager-operátor alkalmazásával történő pillanatnyi jellemző-számítás meghatározott feltételek mellett lehetséges, amit például alkalmas sávszűréssel biztosíthatunk.

Felmerülhet a kérdés, hogy nincs-e ennél általánosabb módszer a fizikailag is értelmezett pillanatnyi paraméterek – a pillanatnyi frekvencia és pillanatnyi amplitúdó – becslésére? Az igenlő választ Norden E. Huang és munkatársai adták meg 1998-ban közölt dolgozatukban [5]. A cikkükben felvetett egyik első kérdés az, hogy mi jellemzi a fizikailag értelmezhető pillanatnyi frekvenciát? A természetes válasz az, hogy a pillanatnyi frekvencia legyen pozitív valós szám. Ezt követően felmerül, hogy olyan jel esetében, aminek nincs egyenáramú komponense, milyen jelbéli szerkezet az, ami negatív pillanatnyi frekvenciát ad? Ennek ismeretében ugyanis törekedni lehet az ilyen jelszerkezet elkerülésére a pozitív pillanatnyi frekvencia biztosítása érdekében. A szerzők érveléséből kiderül, hogy abban az esetben, ha két egymást követő pozitív helyi maximum között található pozitív helyi minimum, avagy két negatív helyi minimum között található negatív helyi maximum, a pillanatnyi frekvencia negatív lesz.

Tehát a feladat az, hogy a pillanatnyi jellemzők számítása előtt a meglévő mintasorozatból olyan összetevőket kell kinyerni, amelyekre az előző tulajdonság nem teljesül. Ezt követően már sor kerülhet a pillanatnyi jellemzők számítására is. A természetes módusfelbontás (EMD, Empirical Mode Decomposition) nevű algoritmust adták meg ezen összetevők előállítására, amelyeket benső módusfüggvényeknek (IMF, Intrinsic Mode Functions) neveztek el. A felbontást követően már a jelmodellezésből ismert módszerekkel, nevezetesen az egyes benső módusfüggvények kanonikus reprezentációjának segítségével lehet meghatározni a pillanatnyi amplitúdót és a pillanatnyi frekvenciát.

3.1. A természetes módusfelbontási eljárás és a benső módusfüggvények

A benső módusfüggvények tehát eleget tesznek az előző bekezdésben leírt feltételeknek, aminek következménye, hogy két alapvető tulajdonsággal kell rendelkezzenek [5]:

– a szélsőértékek és a nullaátmenetek száma vagy azonos, vagy eltérésük 1,
– rendre a helyi maximumok és minimumok által kijelölt burkolók középértéke zérus.

A benső módusfüggvények előállítása az [5]-ben közölt algoritmussal történik.

Az algoritmusban főszerepet játszik a leválasztási eljárás (sifting process), mert – szemléletesen szólva – ezzel fejtünk le a jelről rendre egy-egy benső módusfüggvényt. Mindeközben az eredeti jel (adatsor) helyi jellemzőivel kell számolni, így a benső módusfüggvények a jelhez igazítottak lesznek, vagyis az eljárás ebben az értelemben adaptív. A leválasztási eljárás ezen felül olyan, hogy az eredeti jel – egy maradékjeltől eltekintve – a benső módusfüggvények összegzésével állítható elő. A benső módusfüggvények számára [5] nem tartalmaz előírást, így azt többnyire tapasztalati úton kell meghatározni. A leválasztási eljárás után az eredeti valós mintasorozat tehát a következőképpen írható fel:

$$x(n) = r(n) + \sum_{k=0}^{K-1} m_k(n), \quad (16)$$

ahol $r(n)$ a maradékjel, $m_k(n)$ a k-adik benső módusfüggvény.

3.2. A jel kanonikus reprezentációja és a pillanatnyi jellemzők

Gábor Dénes részletes vizsgálatainak [7] eredménye, hogy az $x(t)=a(t)\cdot\cos(\varphi(t))$ alakú jelmodell, amit az előző pontban használtunk, nem minden esetben egyértelmű. Ha azonban a jelből és

$$\hat{x}(t) = H\{x(t)\} = \frac{1}{\pi}\cdot P\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau \quad (17)$$

Hilbert-transzformáltjából előállítjuk a

$$Z(t) = x(t) + j\cdot\hat{x}(t) = A(t)\cdot e^{j\Phi(t)} \quad (18)$$

komplex analitikus jelet, akkor az ebből származtatható

$$x(t) = A(t)\cdot\cos(\Phi(t)) \quad (19)$$

kanonikus reprezentáció már egyértelmű,

továbbá a pillanatnyi paraméterek is definiálhatók:

$$A(t)=\sqrt{x^2(t)+\hat{x}^2(t)} \tag{20}$$

$$\omega(t)=\frac{d\Phi(t)}{dt}=\frac{d}{dt}\left(\arctan\left(\frac{\hat{x}(t)}{x(t)}\right)\right). \tag{21}$$

Megjegyezzük, hogy a (17) egyenletben az improprius integrál főértéke, a létező, $\lim_{A\to\infty}\int_{-A}^{+A} f(x)dx$ alakú határérték szerepel – erre utal a P betű.

A (21) egyenletben a pillanatnyi körfekvencia az analitikus jel fázisának deriváltjaként áll elő, de számítható az

$$\omega(t)=\mathrm{Im}\left\{\frac{\partial}{\partial t}\ln(Z(t))\right\} \tag{22}$$

összefüggés alapján is, amivel

$$\omega(t)=\frac{x(t)\cdot\frac{d\hat{x}(t)}{dt}-\frac{dx(t)}{dt}\cdot\hat{x}(t)}{x^2(t)+\hat{x}^2(t)}, \tag{23}$$

akárcsak a (21)-ben kijelölt deriválás tényleges elvégzésével. Mind (21), mind (23) alapján származtathatunk algoritmust a pillanatnyi frekvencia becslésére. Fontos tulajdonság, hogy a jel és Hilbert-transzformáltjának Fourier-transzformáltja között fennáll az

$$\hat{X}(j\omega)=-j\cdot\mathrm{sgn}(\omega)\cdot X(j\omega), \tag{24}$$

összefüggés, továbbá teljesül, hogy

$$F\{x(t)+j\cdot\hat{x}(t)\}=\begin{cases}2\cdot X(j\omega) & \omega>0\\ 0 & 0\le\omega\end{cases}, \tag{25}$$

ahol F{.} a Fourier-transzformáció műveletét jelöli.

3.3. A diszkrét idejű Hilbert-transzformált számítása és a pillanatnyi jellemzők becslése

A diszkrét idejű Hilbert-transzformáltat előállíthatjuk (24)-ből kiindulva megfelelő digitális szűrővel [8], vagy (25) alapján FFT-re alapozott számítási eljárással, amit jelen munka során is használtunk.

A mintasorozat és a Hilbert-transzformált sorozat ismeretében következhet a pillanatnyi amplitúdó és a pillanatnyi frekvencia becslése. A pillanatnyi amplitúdó mintákat (20) alapján a következőképpen határozhatjuk meg:

$$A(n)=\sqrt{x^2(n)+\hat{x}^2(n)} \tag{26}$$

A pillanatnyi frekvencia minták számítására egyrészt (21), másrészt (23) alapján származtathatunk eljárást. A (21) alapján a fázis mintasorozat

$$\Phi(n)=\mathrm{arctg}\left(\frac{\hat{x}(n)}{x(n)}\right), \tag{27}$$

ám a jel időbeni fejlődése során a fázis úgy változik, hogy

$$\Phi_u(n)=\Phi(n)+r(n)\cdot 2\pi,\qquad \Phi(n)\in[-\pi;\pi], \tag{28}$$

ahol *r(n)* pozitív egész szám. A számítások során azonban közvetlenül a fázis főértékének *Φ(n)* mintái adódnak, ebből kell a tényleges fázis mintáit előállítani. A feladat az, hogy minden minta esetében ismert legyen a 2π ide tartozó egész szám-szorosa, vagyis elő kell állítani az *r(n)* sorozatot. Erre a célra például a mod(2π) fázis-visszahajtogatási (phase-unwrapper) eljárás használható [2].

Ha rendelkezésre áll a pillanatnyi fázis, a pillanatnyi digitális körfrekvencia meghatározásához szükséges deriválást az alábbi differencia kiszámításával közelíthetjük:

$$\omega(n)=\Phi_u(n)-\Phi_u(n-1). \tag{29}$$

Más eljárás adódik (23) alapján, ahol a deriválás alkalmas közelítése szükséges. Ahogy az előző pontban, itt is alkalmazható a Savitzky-Golay-féle 5 pontos simító deriválás.

4. A Teager-operátor alapján és a HHT-vel számított pillanatnyi jellemzők összehasonlítása

4.1. A jel visszaállítása a pillanatnyi jellemzőkből

A 2. pontban ismertettük, hogy a lassan változó jel pillanatnyi amplitúdójának és frekvenciájának abszolút értéke két algoritmus-párral is becsülhető, míg a 3. pontban a benső módusfüggvényekhez rendelt analitikus jel alapján becsültük a pillanatnyi amplitúdót, továbbá vagy közvetlenül, vagy a pillanatnyi fázis előállítását követően a pillanatnyi frekvenciát. Ezekre a becslésekre is megadtunk két algoritmus-párt. Mivel az előző két pontban tárgyalt algoritmusok megközelítési módja, az alkalmazott jelmodell lényegesen különbözik egymástól, felmerül a kérdés, hogy ugyanazon a jelen számolt pillanatnyi jellemzőik hogyan viszonyulnak egymáshoz? Esetleg valamilyen szempontból hasonlóak-e?

Ebben a pontban ezt a kérdést vizsgáljuk meg az alábbi négy összetartozó algoritmus-pár összehasonlításával (zárójelben az ezt követő táblázatokban szereplő elnevezések):

- közvetlen számítással becsült pillanatnyi amplitúdó és frekvencia (közvetlen számítás),
- a diszkrét idejű Teager-operátorral becsült pillanatnyi amplitúdó és frekvencia (ES-algoritmus),
- a kanonikus reprezentáció alapján számolt pillanatnyi amplitúdó és a fázis-visszahajtogatással kapott pillanatnyi frekvencia (HHT (fázis-differencia)),
- a kanonikus reprezentáció alapján számolt pillanatnyi amplitúdó és a simító deriválással kapott pillanatnyi frekvencia (HHT (simító deriválás)).

Az egyes algoritmus-párok összehasonlításának egy lehetséges módja az, hogy az adott jel esetén meghatározzuk velük a pillanatnyi jellemzőket, majd ugyanazon visszaállítási eljárással e pillanatnyi jellemzőkből becsüljük az eredeti jelet. Az *x(n)* eredeti jel, és az $\tilde{x}(n)$ becslés ismeretében az adott algoritmus-pár jóságát az

$$\mathrm{NSR}=10\cdot\log\left(\frac{\sum_{n=4}^{N-5}\left[x(n)-\tilde{x}(n)\right]^2}{\sum_{n=4}^{N-5}x^2(n)}\right) \tag{30}$$

zaj/jel viszonnyal jellemezzük.

Az indexek magyarázata az, hogy a közvetlen számítás során nemcsak a jelre, hanem deriváltjára is alkalmazzuk az 5 pontos simító deriválást, így a jel mindkét széléről elhagyunk 4-4 mintát. Emiatt mindegyik algoritmus-párnál az így adódó jelrészletet vettük figyelembe. A visszaállítási algoritmus alapja maga az adott algoritmus-párhoz tartozó jelmodell. Ehhez a pillanatnyi amplitúdó mindegyik esetben közvetlenül adódik. A saját jelmodelljének megfelelő pillanatnyi fázist azonban csak egy algoritmus állítja elő közvetlenül, a többi három eljárás a pillanatnyi frekvenciára ad becslést, ezért – az egységesség érdekében – mindegyik esetben a pillanatnyi frekvenciából indultunk ki, és ebből határoztuk meg a pillanatnyi fázist az alábbiak szerint:

$$\tilde{\Phi}(k) = \tilde{\Phi}(-1) + \sum_{n=0}^{k} \tilde{\omega}(n) \qquad k = 0,1,\ldots,N-1 \qquad (31)$$

Numerikus kísérleteink tanúsága szerint az egyes esetekben a visszaállított jel és az eredeti jel között fázisingadozás mutatkozik. Ezért mindegyik algoritmuspárnál kereséssel határoztuk meg a legjobb NSR-t adó $\Phi(-1)$ kezdőfázist $\pi/180$ (1°) fázisléptetés mellett.

Az összehasonlítást vizsgálójelen és egy szó bemondásából származó beszédjelen is elvégeztük.

4.2. A módszerek összehasonlítása vizsgálójel esetén

Vizsgálójelnek a szakirodalomban található AM-FM-jelet használtuk [2]:

$$s(n) = (0{,}998)^n \cdot \left[1 + 0{,}8 \cdot \cos(2 \cdot \pi \cdot f_3 \cdot n)\right] \cdot \cos\left[2 \cdot \pi \cdot \left(f_1 \cdot n + \frac{1}{2 \cdot \pi} \cdot \sin(2 \cdot \pi \cdot f_2 \cdot n)\right)\right] \qquad (32)$$

$$f_s = 10000\text{Hz} \quad f_1 = \frac{1000\text{Hz}}{f_s} \quad f_2 = \frac{100\text{Hz}}{f_s} \quad f_3 = \frac{50\text{Hz}}{f_s}$$

Időbeli alakja alapján ez a jel egyben benső módusfüggvény is, ezért azt várjuk, hogy az EMD-algoritmus egyetlen lényeges IMF-et ad vissza.

Ez így is van, amint az a túloldali *1. ábrán* is látható. A visszaállított jel eltérését mind az eredetitől, mind az IMF-től számszerűen jellemezve az *1. táblázatban* látható adatokat kapjuk.

Az *1. ábra* a számított eredményeket szemlélteti vizsgálójel-részleten. Az ábra *b)* részén kivehető, hogy az 1°-os fázisléptetés ellenére egyik-másik módszernél még marad kis fázishiba, ami nyilván rontja a zaj/jel viszonyt. Az elméleti pillanatnyi frekvenciát az egyes módszerek kis hibával közelítik, az elméleti pillanatnyi amplitúdó közelítése is közel azonosan jó.

4.3. A módszerek összehasonlítása sávszűrt beszédjel esetén

Az előző pontban a vizsgálójel – konstrukciójánál fogva – olyan volt, hogy pillanatnyi jellemzői lassan változtak, így a (6) feltétel teljesült, ami a pillanatnyi paraméterek becsléséhez szükséges mind a közvetlen számítás, mind az ES-algoritmus esetében. Ennek megfelelően a beszédjel esetében is gondoskodni kell arról, hogy a becsülni kívánt pillanatnyi jellemzők lassan változzanak.

Ez megfelelő sávszűréssel biztosítható. A megfelelő sávszűrő tervezésére beszédfeldolgozási feladatokhoz – ismereteink szerint – nincs általánosan elfogadott módszer, de a szakirodalom szerint az egységnyi kritikus sávszélességű szűrősor (valamely tagja) megfelelő a Teager-operátor alkalmazhatóságához [4], ami az említett két eljárás alapja. A sávszűrő alkalmazásának praktikus oka is van, hiszen a tapasztalat szerint megfelelő sávszűrés után a diszkrét idejű Teager-operátor sokkal ritkábban ad negatív, tehát fizikailag nem értelmezett értéket, mint a nélkül.

Ebben a pontban sávszűrt beszédjel pillanatnyi jellemzőinek becslését mutatjuk be. A beszédjelminták az *igen* szó férfi bemondótól származó megvalósításából származnak 8000 Hz mintavételi frekvencia és 16 bites lineáris kvantálás alkalmazásával. Az eredeti bemondást 300 Hz...3400 Hz áteresztő sávú lineáris fázisú FIR-szűrővel sávhatároltuk. A spektrogram megtekintése alapján az 500 Hz körüli erős formáns jelenléte miatt hallásmodell alapú wavelet-szűrősor egyik tagját alkalmaztuk további lineáris fázisú FIR-szűrésre [9]. Az így előállt jel amplitúdóját a $-1 \le x(n) \le 1$ egyenlőtlenség szerint normalizáltuk, ez lett a pillanatnyi jellemzők becslésének alapja.

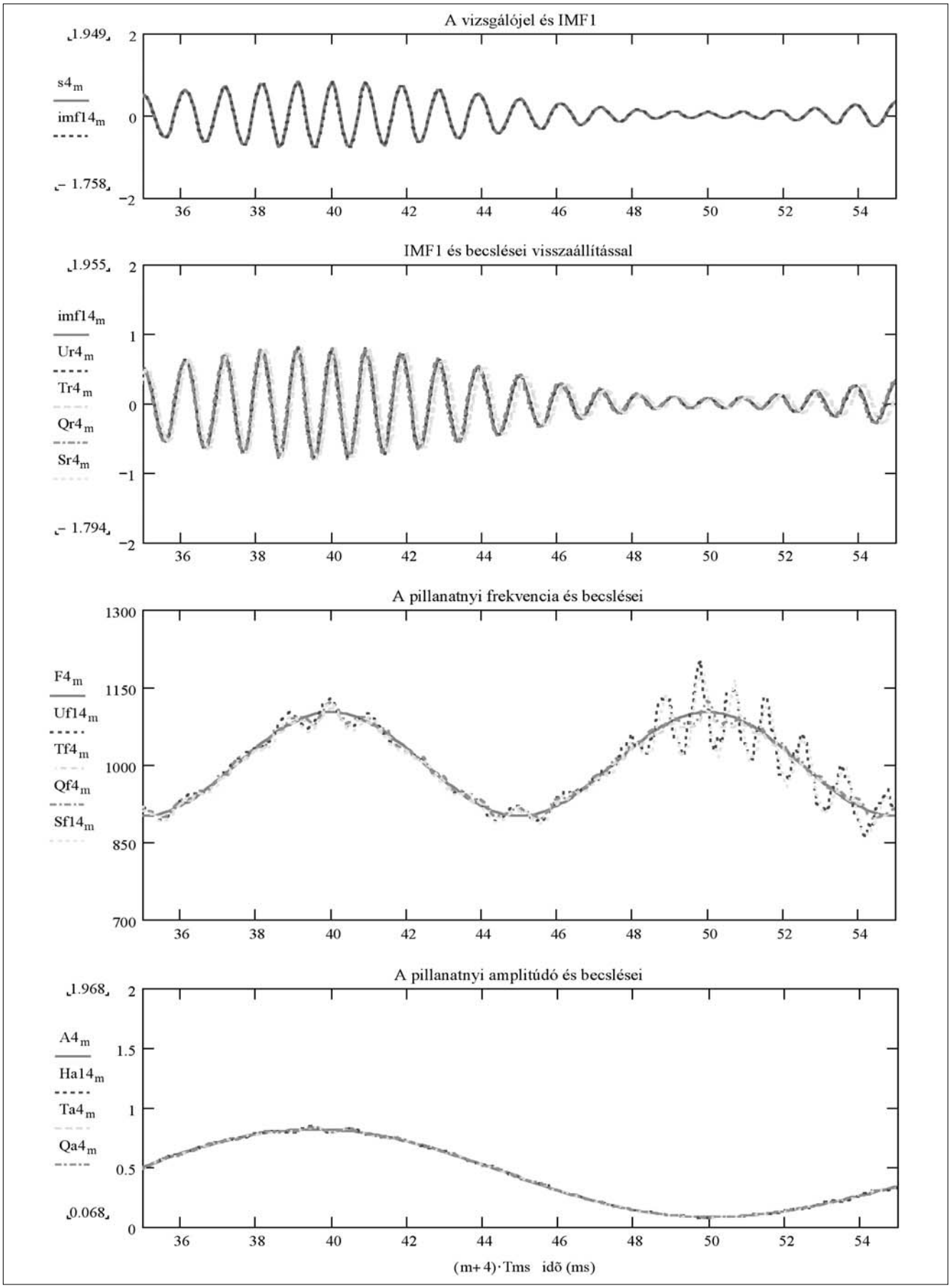
Megtekintve a sávszűrt jelet, látható, hogy ez is benső módusfüggvény, ezért azt várjuk, hogy az EMD-algoritmus egyetlen lényeges IMF-et ad vissza. Ez itt is így van, amint az a *2. ábrán* is látható. A visszaállított jel eltérését mind az eredetitől, mind az IMF-től számszerűen jellemezve a *2. táblázatban* látható adatokat kapjuk. A legjobb eredményt a HHT (fázis-differencia) módszer adja.

1. táblázat
Az algoritmus-párok jellemzése a vizsgálójel esetében

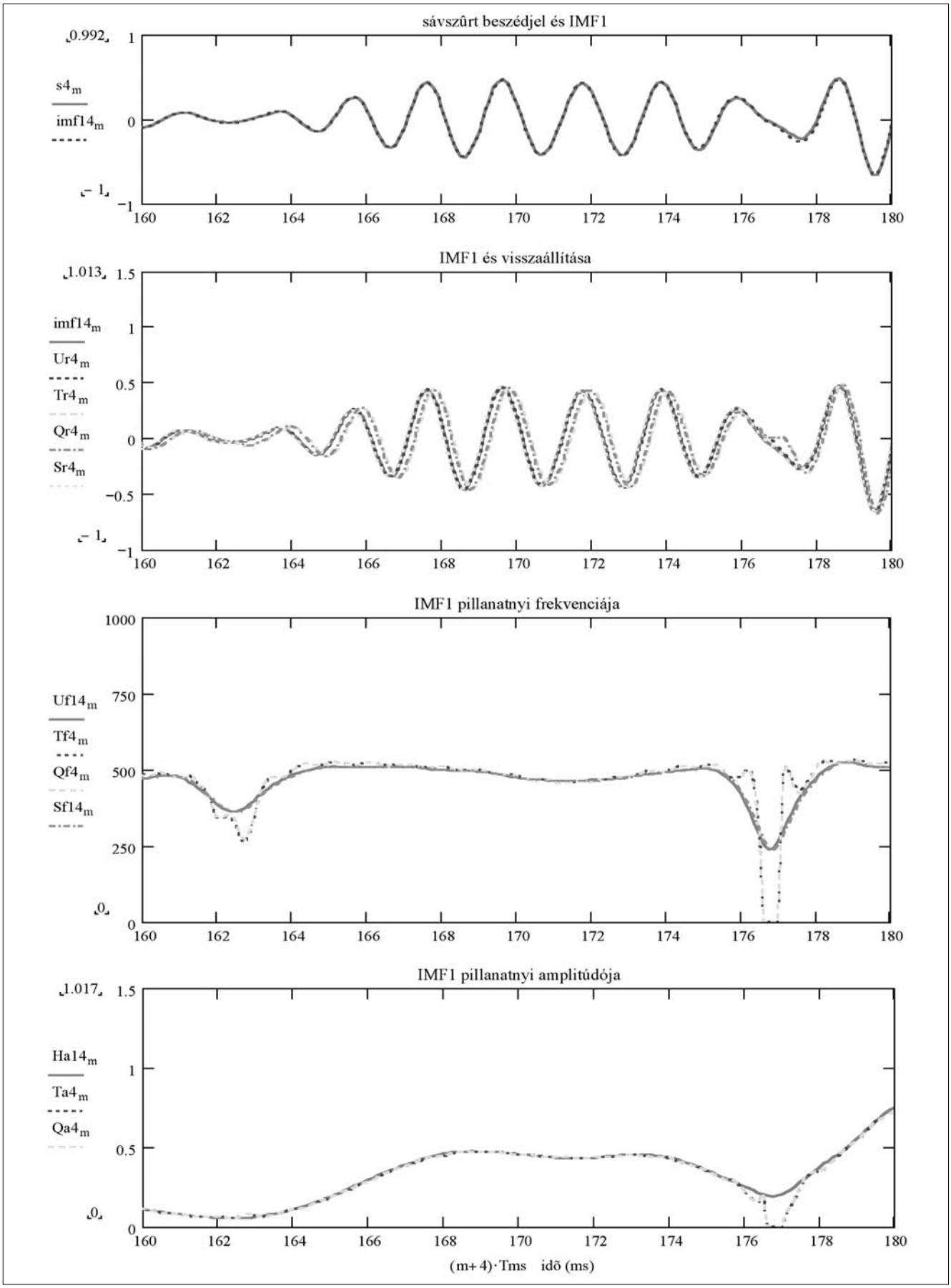
Módszer	Eredeti jel NSR (dB)	IMF1 NSR (dB)
Közvetlen számítás	-8	-8
ES-algoritmus	-18	-19
HHT (fázis-differencia)	-24	-27
HHT (simító deriválás)	-7	-7

2. táblázat
Az algoritmus-párok jellemzése a sávszűrt beszédjel esetén

Módszer	Eredeti jel NSR (dB)	IMF1 NSR (dB)
Közvetlen számítás	-5	-5
ES-algoritmus	-2	-2
HHT (fázis-differencia)	-29	-30
HHT (simító deriválás)	-14	-14



1. ábra A négy algoritmus-párral számolt eredmények szemléltetése vizsgálójel-részleten: a) a vizsgálójel és az első benső módusfüggvény (IMF1), b) az IMF1 és négyféle becslése, c) az elméleti pillanatnyi frekvencia és négyféle becslése, d) az elméleti pillanatnyi amplitúdó és háromféle becslése



2. ábra A négy algoritmus-párral számolt eredmények szemléltetése sávszűrt beszédjel-részleten: a) a sávszűrt beszédjel és az első benső módusfüggvény (IMF1), b) az IMF1 és négyféle rekonstruálása, c) a pillanatnyi frekvencia négyféle becslése, d) a pillanatnyi amplitúdó háromféle becslése

A *2. ábra* a számított eredményeket szemlélteti sávszűrt beszédjel-részleten. Az ábra *b)* részén látható, hogy az 1°-os fázisléptetés ellenére több módszernél is van fázishiba. (Megjegyezzük, hogy bár az NSR alapján az eredeti beszédhez képest nagy eltérésre következtetnénk három algoritmusnál is, a rekonstruált beszédet meghallgatva azt jónak találjuk.) A pillanatnyi frekvencia becslésénél együtt fut rendre a két HHT-s és a két Teager-operátoros algoritmussal számolt adatsor. Ez utóbbiaknál a 177 ms-nál lévő beszakadás oka az, hogy a megvalósított program 0 becsült frekvenciaértéket ad vissza, ha negatív számból kellene gyököt vonni (lásd (10),(11),(14),(15)).

Ez a megoldás az algoritmus vizsgálatakor fontos, a gyakorlati alkalmazásban a környező adatokból becsült helyettesítő értékkel élhetünk ilyenkor. A kis pillanatnyi amplitúdót és a jelrészletet megvizsgálva látható, hogy az efféle bizonytalan becslés a 0-hoz közeli jelamplitúdóknál fordulhat elő. Ettől eltekintve a négyféle módszer becslései jól egyeznek.

4.4. A módszerek összehasonlítása beszédjel esetén

Az előző pontban bemutatott eredmények egyrészt megerősítik azt a tapasztalatot, hogy a beszédjel Teager-operátoros feldolgozásához elegendő az egységnyi kritikus sávszélességű szűrőkből álló szűrősor alkalmazása, másrészt megfigyelhető, hogy a Teager-operátorra alapozott becslések nagyon hasonlítanak a HHT-vel kapható becslésekhez. Felvetődik a kérdés: helyettesítheti-e a szóban forgó szűrést a természetes módusfelbontás, és hogyan alakulnak a becsült pillanatnyi jellemzők? Erre a kérdésre megítélésünk szerint csak nagy beszéd-adatbázison történő részletes vizsgálat eredményei alapján lehet válaszolni. Az alábbiakban egyetlen szó bemondásának elemzésével kapott eredményeinket mutatjuk be.

Ezekben a vizsgálatokban tehát nem szerepel sávszűrés. Maga a természetes módusfelbontási eljárás viselkedik sávszűrőként, mégpedig az adott beszédjelhez igazodó, adaptív módon. Ugyanis a felső és alsó burkolók egymáshoz igazítása a helyi maximumokhoz és minimumokhoz kapcsolódik, vagyis az első benső módusfüggvény a jelamplitúdóban lévő, egymás szomszédságában található gyors változásokhoz, így a magasabb frekvenciájú spektrális részlethez igazodik. Utána azt a jelből levonva haladunk tovább a következő módusfüggvényekhez, vagyis a kisebb frekvenciájú spektrális részletek felé. (Az EMD-eljárás ilyetén viselkedése jól nyomon követhető az egyes IMF-ek spektrogramján is.) Emiatt az a kérdés, hogy az így megvalósuló adaptív szűrés elégséges-e a Teager-operátoros pillanatnyi jellemző-becslésekhez?

Jelen dolgozatban ezt a kérdést is az előző pontban szereplő *igen* szó bemondásából nyert mintasorozaton vizsgáltuk. A 3. pontban említettük, hogy az EMD-algoritmus alkalmazásakor nincs támpont arra, hogy mennyi a benső módusfüggvények elégséges száma. Numerikus kísérleteink azt mutatják, hogy az első három benső módusfüggvényből (16) alapján az eredeti beszédjel NSR= -22 dB jósággal állítható vissza, ezért a pillanatnyi jellemzőket az első három módusfüggvényre számítottuk ki a négyféle módszerrel, és a visszaállítást is rendre a három benső módusfüggvényre végeztük el, majd a rekonstruált beszédjelet ezek összegzésével határoztuk meg.

Az áttekinthetőség érdekében a *3. ábrán* csak a legjobban közelítő algoritmussal kapott beszédrészleteket mutatjuk be, a *3. táblázat* a számszerű eredményeket tartalmazza.

A 3. ábrához tartozó fontos megjegyzés, hogy nem szerepel a másik három módszerrel kapott visszaállított beszédjel, de itt is megfigyelhető a fázisingadozás jelensége, ami a zaj/jel viszonyt lerontja, viszont maga a beszéd jól érthető.

5. Következtetések

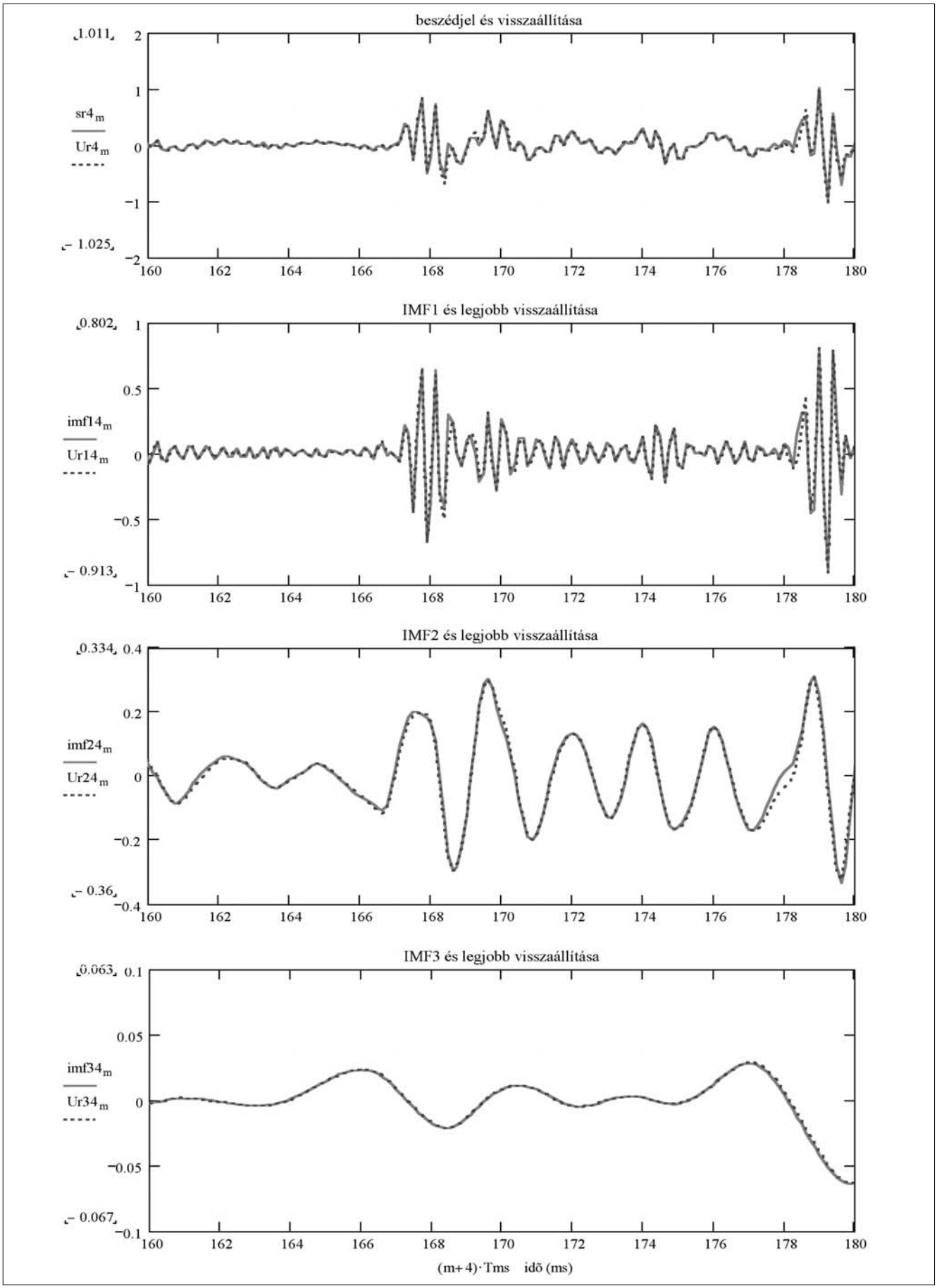
A dolgozatban a beszédjel pillanatnyi amplitúdójának és pillanatnyi frekvenciájának becslésére mutattunk be négyféle módszert. Ezek közül kettő a Teager-operátorra, kettő pedig a Hilbert-Huang-transzformációra épül. A vizsgálójel és beszédjel pillanatnyi jellemzőinek becslési példáin ábrákkal szemléltettük az egyes módszereket, és megadtunk egy visszaállítási eljárást is, amivel a beszédjel a becsült pillanatnyi jellemzőkből rekonstruálható. Ez alapján már alkalmas zaj/jel viszonnyal hasonlíthatók össze az egyes módszerek. A munka során szerzett tapasztalatainkat a dolgozatban több helyütt ismertettük.

Legfontosabb következtetéseink az alábbiak:

1. A Teager-operátorra alapozott módszerek esetében lényeges, hogy sávszűrt beszédjelen végezzük a becslést. Erre a célra szolgálhat valamely egységnyi kritikus sávszélességű szűrő, ennek kimenete dolgozható fel tovább a Teager-operátorra alapozott algoritmusokkal. Ez a Teager-operátorra épülő pillanatnyi amplitúdó és pillanatnyi frekvencia becslés esetében is igaz.
2. A beszédjelből a természetes módusfelbontási eljárással kapható benső módusfüggvények pillanatnyi jellemzőire mind a Teager-operátorra alapozott módszerekkel, mind a HHT alapján egymáshoz hasonló eredmények adódnak. Ez nem magától értetődő, és megítélésünk szerint érdemes nagy beszéd-adatbázison részletesen megvizsgálni, hogy általánosabb érvényű-e ez a megfigyelésünk.

3. táblázat
A legjobb visszaállítást adó módszer jellemző adatai

Módszer	Eredeti jel NSR (dB)	IMF1 NSR (dB)	IMF2 NSR (dB)	IMF3 NSR (dB)
HHT (fázis-differencia)	-12	-10	-19	-24



3. ábra A beszédjel valamint az első három módusfüggvény visszaállítása a pillanatnyi jellemzőkből: a) az eredeti jel és a becsült IMF-ek összegzésével kapott visszaállítás b-d) rendre az egyes IMF-ek és legjobb becsléseik

3. Az ES-algoritmusnál tapasztalható az a hiba, hogy néha negatív számból kellene a végrehajtás során négyzetgyököt vonni, amit a szomszédos becslések alapján javaslunk kiküszöbölni. Ugyanis – bár kézenfekvő lenne mediánszűrővel simítani a pillanatnyi jellemzőket becslő adatsorokat – tapasztalataink szerint az így simított változatból visszaállított beszédjel a meghallgatáskor rosszabb minőségű, mind mediánszűrés nélkül.
4. A visszaállítás során tapasztalható egy fázisingadozási jelenség, mely szerint a nullához közeli jelamplitúdót követő jelrészlet visszaállítása időben elcsúszik. Emiatt jobb visszaállítást várhatunk el, ha a rekonstruáló algoritmust úgy módosítjuk, hogy minden nulla-közeli jelrészlet után keresse meg a legjobb illeszkedést adó kezdőfázist.
5. Az EMD-algoritmus módosítható annak figyelembe vételével, hogy a vizsgált beszédjel eleve sávhatárolt. Így – például a maradékjel és az eredeti jel megfelelően előírt zaj/jel viszonya alapján – automatikusan kaphatjuk meg a szükséges számú benső módusfüggvényt.

További feladatunk a jelen dolgozatban tárgyalt algoritmusok alkalmazási lehetőségeinek feltárása.

Köszönetnyilvánítás

A szerző ezen a helyen is megköszöni Gordos Gézának, Németh Gézának és Tatai Péternek (BME VIK TMIT) a gépi beszédfeldolgozási célú algoritmusfejlesztési munkái során kapott segítséget, támogatást és bíztatást.

Irodalom

[1] Gordos G., Takács Gy.:
Digitális beszédfeldolgozás. Műszaki Könyvkiadó, 1983.
[2] Quatieri, T. F.:
Discrete-time Speech Signal Processing:
Principles and Practice. Prentice-Hall, 2001.
[3] Abbate, A., DeCusatis, M. C., Das, K. P.:
Wavelets and Subbands:
Fundamentals and Applications. Birkhäuser, 2002.
[4] Chen, S-H., Wang, J-F.:
„Speech Enhancement Using Perceptual Wavelet
Packet Decomposition and Teager Energy Operator”,
Journal of VLSI Signal Processing 36, pp.125–139.,
Kluwer Academic Publishers, 2004.
[5] Huang, N. E., Shen, Z., Long, S. R.,
Wu, M. C., Shih, H. H., Zheng, Q., Yen, N-C.,
Tung, C. C., Liu, H. H.:
„The empirical mode decomposition and
the Hilbert spectrum for nonlinear and non-stationary
time series analysis”.
Proc. R. Soc. Lond. A (1998) 454, pp.903–995.
[6] Valkó P. Vajda S.:
Műszaki-tudományos feladatok megoldása
személyi számítógéppel. Műszaki Könyvkiadó, 1987.
[7] Gábor, D.:
Theory of communication.
J. Inst. Electr. Eng., Vol. 93. (1946), pp.429–457.
[8] Simonyi E.:
Digitális szűrők – a digitális jelfeldolgozás alapjai.
Műszaki Könyvkiadó, 1984.
[9] Pintér, I.,
„Perceptual wavelet-representation of speech signals
and its application to speech enhancement”,
Computer, Speech and Language,
Vol. 10. No.1. pp.1–22., Academic Press, 1996.