

Beszédátviteli rendszerek objektív minősítése

HANZÓ LAJOS—HINSENKAMP LÁSZLÓ—UHERECZKY LÁSZLÓ

Távközlési Kutató Intézet



ÖSSZEFOGLALÁS

A kódolt beszédátviteli rendszerek minősítésével kapcsolatosan más problémák vetődnek fel, mint a hagyományos analóg rendszereknél. A kódolás hatására a csatorna transzparenciája mindenképpen „gyengül”, s ezért a vizsgáló jelek a konfidens minősítés érdekében a lehető legjobban hasonlítanak az átviendő jelekhez. Ez beszédátviteli rendszereknél eddig csak beszédjellel való széles körű szubjektív vizsgálatokkal volt lehetséges. A CCITT vizsgálataiban azonban azt bizonyítják, hogy a beszéd lineáris predikciós modelljén alapuló objektív cepstrumtávolságmérték a szubjektív vizsgálatokkal egyező eredményt szolgáltat, s gyors minősítést tesz lehetővé.

1. A probléma felvetése

A beszédátvitel minőségének legfőbb jellemzője, hogy azok, akik használják, milyen véleménnyel vannak róla [3]. Ezt a szubjektív véleményt a berendezések és rendszerek specifikálásához, tervezéséhez és minősítéséhez számszerű, mérhető, könnyen ellenőrizhető adatokra kell „lefordítani”. Legkorábban az érthető beszédátvitelhez tartozó sávzélességet határozták meg. Széles körű kutatások alapján állapították meg az összefüggést a mondat-, szó-, szótagérthetőség (a megértett mondatok, szavak, szótagok százaléka) és az ezen tulajdonságokat egységesen leíró érthetőségi index között [3]. Az érthetőségi index az „átlagos beszéd” érthetőségére jellemző százalék, míg azonos körülmények között a nagyobb redundancia miatt a mondatérthetőség mindig jobb, mint a szóérthetőség. Az érthetőségi index alapján úgy találták, hogy a még érthető (telefon minőségű) beszéd átviteléhez legalább a 300...3400 Hz-es sáv átvitele szükséges.

Itt célszerű megjegyeznünk, hogy a beszéd átviteléhez (extrém esetektől eltekintve) nem szükséges megszorítást tenni az átvitel fáziskarakterisztikájára, ami plauzibilis, ha arra gondolunk, hogy a beszéd érthetőségét nem befolyásolja lényegesen az akusztikus környezet (a visszhangokat visszaverő felületek formája és távolsága, a visszhangok keletkezésének módja stb.). A tapasztalat szerint az átvitel fáziskarakterisztikájának megváltozását a beszéd színezetének megváltozásaként érzékeljük, de az érthetőség változatlan marad [12]. Hogy a hétköznapi beszédben mennyire jelen vannak a visszhangok, azt akkor érzékeljük, ha visszhangmentesített (és akusztikusan árnyékolt) ún. süket-szobában próbálunk beszélni. A fázismentesre vonatkozó előírás a másodlagos hírsanyagok (távíró, adat stb.), valamint a beszédnél eleve jóval szélesebb frekvenciasávot igénylő zene átviteléhez szükséges.

Beérkezett: 1986. 1. 21. (□)

DR. HANZÓ LAJOS

A BME Híradástechnika Szakán végzett 1976-ban, s diplomatervével, valamint TDK dolgozatával pályázatot nyert. Azóta a TKI tudományos munkatársa. 1980-ban egy évet dolgozott az erlangeni egyetemen (NSZK), 1982-ben szakmérnöki diplomát, 1983-ban

egyetemi doktori fokozatot szerzett, 1984-ben Pollack-Virág-díjjal, 1985-ben pedig OMFB szabadalmi pályadíjjal tüntették ki. Szakmai érdeklődési körébe az információátvitellel kapcsolatos jelfeldolgozási és rendszertechnikai problémák tartoznak. Hobby: zene, sport, fajamegmunkálás.

Új helyzet állt elő a digitális átvitel (PCM) megjelenésével. Szabályozni kellett a linearitás (ami alatt a be- és kimeneti szint arányos változásának toleranciáját értjük), valamint a kvantálási zaj mértékét is.

Véleményünk szerint a PCM-től eltérően kódolt (általában csökkentett sebességű) beszédátvitel terjedésével ismét új helyzet áll elő. A differenciális kódolási eljárások ui. kihasználják, hogy az egyes beszédminták nem függetlenek egymástól, s a jel változásának csak véges meredekségét képesek átvinni. A jelmeredekség korlátozása még az adaptív eljárásoknál is jelentkezik. Egyrészt azért, mert az adaptív eljárások csak bizonyos késéssel képesek követni a jel természetének megváltozását, másrészt azért, mert realizálási okok miatt mindenféleképpen jelentkezik egy határmeredekség. A jel alakja sokszor durván eltorzul, de ez a tény nem biztos, hogy lényegesen rontja a beszédátvitel minőségét, hiszen a lineáris fázistorzítással kapcsolatban beláttuk, hogy érthető beszédátvitelhez nem feltétlenül szükséges a beszéd hullámformájának alakhíj átvitele, ami egyébként a torzítatlan átvitel szokásos feltétele.

A jelalak torzulása olyan mértékű lehet, hogy ezek a rendszerek közelítőleg sem tekinthetők lineárisnak, de azt sem állíthatjuk, hogy tipikusan nemlineárisak. Az ilyen tulajdonság jellemzésére megkülönböztetésül a szokásos „gyengén nemlineáris” jelzőről a „kvázi lineáris” meghatározást tartjuk célszerűnek. A PCM átvitel jellemzésére szokásos linearitás és kvantálási zaj (torzítás) alkalmazása a kódolt átvitelre erősen vitatható, ui. egyrészt a hullámalak torzulása kérdésessé teszi a szint értelmezését, másrészt a kvantálási zaj mérésekor használt hullámforma által keletkező zavaró jelek nem biztos, hogy jellemzőek a minőség romlására. Általában a kvázilineáris rendszerek vizsgálatához a linearitáson alapuló Fourier felbontás (szinuszos mérőjelek) nem alkalmazható. Az újszerű feladat megoldásához kétféle dologra kell tenni:

1. A mérőjellel „közelebb kell menni” ahhoz a jellehez, aminek átvitelére a rendszert terveztük. Ez mindenféleképpen a minősítés általánosságá-

nak romlásával jár. Csak olyan típusú jelek átviteli minőségére tehetünk megállapításokat, amilyen típusú mérőjeleket alkalmaztunk. Ez a tény a linearitás korlátozottságának a következménye.

2. Újra kell fogalmazni azt a minősítési rendszert, aminek segítségével a felhasználó szubjektív véleményét számszerű adatokra „fordítottuk le”.

Mivel az átvendő jelek véletlenszerűek, célszerű, ha a mérőjel is véletlenszerű. Véletlenszerű (sztochasztikus) mérőjelek torzításmérésre való felhasználását [11] tárgyalja.

A csillapítás-torzítás fogalma esetleg átmenthető, ha azt véletlen jelek spektrális burkolójának torzulásként értelmezzük. Méréstechnikai okokból célszerű álvéletlen (véges periódushosszal és vonalas spektrummal rendelkező) mérőjel alkalmazása, ekkor ui. a mérendő átlagértékek hamarabb „kialakulnak”. Egy ilyen generátorral (mivel több szinuszelet állít elő, sokhangú generátornak nevezhető) a különböző szinuszelek kezdőfázisának és amplitúdójának beállításával különböző eloszlások közelíthetők.

Észre kell venni, hogy a beszédátvitel minősítése és a beszélő személy azonosítása rokon feladatok. Az az átviteli rendszer ui., amelyen keresztül a beszélő személy felismerhető, jó minőségűnek mondható. A beszélő személy azonosításának jól ismert módszereit képezik a különböző távolságmérések [1]. Nyilvánvaló, hogy közvetlenül a be- és kimeneti jeleken képzett távolságok nem alkalmasak, ui. a beszéd értelmezéséhez nem szükséges a jel szigorú alakhűségét megtartani. A CCITT vizsgálatai szerint [5], [6], [9] a be- és kimeneti jeleken képzett cepstrumtávolság és a felhasználó szubjektív értékítélete jól egybe esik. A cepstrumtávolság mérésének elvi alapjait a 2. fejezetben tárgyaljuk. A lineáris predikció és a cepstrumanalízis leírásunkhoz szükséges alapvető fogalmait az F.1. és F.2. függelékben foglaltuk össze.

2. Objektív minősítés a cepstrumtávolság alapján [1], [6], [7]

A beszédfeldolgozásban olyan $d(x, y)$ távolságfogalmakat célszerű alkalmazni két beszédminta (x és y) között, melyekre teljesül, hogy:

1. $d(x, y) = d(y, x)$, azaz szimmetrikus,
2. $d(x, y) \geq 0$, ha $x \neq y$, és $d(x, x) = 0$ } azaz pozitív definit,
3. $d(x, y)$ fizikailag értelmezhető a frekvenciatartományban, s
4. hatékony eljárás adható $d(x, y)$ meghatározására.

Az első két kritérium (amelyek meglétét szigorúan megköveteljük), a távolság fogalmához szokás szerint hozzá tartozik. A szimmetria azt mondja ki, hogy nem különböztetjük meg a referens és a mérendő jelet, a pozitív definit követelmény pedig azt, hogy két jel akkor hasonlít legjobban egymásra, ha megegyeznek. A harmadik és negyedik kritérium a gyakorlati alkalmazhatóságot célozza. Nem kell feltétlenül megkövetelni három minta egymás közötti távolságának viszonyára, azaz nem követeljük meg, hogy a távolságok halmaza metrika legyen, vagyis hogy az (1) háromszög egyenlőtlenség teljesüljön:



HINSENKAMP LÁSZLÓ

A BME Híradástechnika szakán végzett 1970-ben. Három évig a BME Mikrohullámú Híradástechnika Tanszékén dolgozott, azóta a Távközlési Kutató Intézet munkatársa. 1974-ben szakmérnöki diplomát szerzett. 1983-ban háromhóna os tanulmányutat tett a bochumi egyetemen (NSZK). Szakmai érdeklődése a digitális hírközléssel kapcsolatos rendszertechnika problémákra terjed ki. Hobby: zene és irodalom.

$$d(x, y) \leq d(x, z) + d(y, z). \quad (1)$$

A [6] dolgozat szerint a kodek be- és kimeneti mintái közötti ún. cepstrumtávolság és a felhasználó szubjektív értékítélete (Mean Opinion Score — MOS [9]) szoros korrelációban van egymással.

Az x és y jel cepstrumtávolságát a

$$CD = \left(2 \sum_{i=1}^K |C_x(i) - C_y(i)|^2 \right)^{1/2} \quad (2)$$

összefüggés definiálja [6], ahol $C_x(i)$ és $C_y(i)$ az x és y -jel cepstrumegyütthatói. A cepstrumegyütthatókat a beszédet modellező (lásd F.1. függelék) tisztán pólusokat tartalmazó, z^{-1} -ben M -ed fokú $A(z)$ polinom segítségével az:

$$\ln \left(\frac{1}{A(z)} \right) = - \sum_{k=1}^{\infty} C(k) z^{-k} \quad (3)$$

Taylor sorfejtéssel számíthatjuk, s belátjuk, hogy a (3) összefüggésben a sorfejtés $C(k)$ együtthatói épp a cepstrumegyütthatók.

A $z = e^{j\theta}$ helyettesítéssel és abszolútérték-négyzet képzéssel:

$$\ln |1/A(e^{-j\theta})|^2 = \ln(1/Ae^{-j\theta}) + \ln(1/Ae^{+j\theta}). \quad (4)$$

A (3) egyenlet (5) szerinti alakjával a (4) összefüggés a (6) szerinti formára hozható:

$$\ln(1/Ae^{-j\theta}) = - \sum_{k=1}^{\infty} C(k) e^{-jk\theta}, \quad (5)$$

$$\ln(\sigma^2 |A(e^{-j\theta})|^2) = - \sum_{k=-\infty}^{\infty} C(k) e^{-jk\theta}, \quad (6)$$

ahol $C(0) = \ln(\sigma^2)$, és $C(k) = C(-k)$.

A (6) egyenletben $z = e^{j\theta}$ helyettesítéssel:

$$\ln |a/A(z)|^2 = - \sum_{k=-\infty}^{\infty} C(k) \cdot z^{-k}, \quad (7)$$

majd inverz z transzformáció után:

$$Z^{-1} \{ \ln |a/A(z)|^2 \} = -C(k), \quad (8)$$

hiszen (7) jobb oldalán $-C(k)$ z transzformáltja áll. Feltételezve, hogy

$$a/A(z) = Z \{ y(m) \}, \quad (9)$$

a (8) összefüggés (10) szerinti lesz:

$$Z^{-1} \{ \ln |Z \{ y(m) \}|^2 \} = -C(k), \quad (10)$$

amit az (F.2.13) egyenlet szerinti cepstrummal összevetve látható, hogy attól abszolútérték-négyzetképzéssel tér el. Ezért nevezik az (F.2.13) szerinti teljesítménycepstrumnak. Beláttuk tehát, hogy a beszédet modellező $A(z)$ polinom ismeretében meghatározhatók a $C(k)$ cepstrumegyütthatók.

Annak érdekében, hogy az $A(z)$ polinom pólusai és $C(k)$ között explicit számítási formulát kapjunk, írjuk fel $A(z)$ -t gyöktényezős alakban:

$$A(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) \dots (1 - z_M z^{-1}), \quad (11)$$

majd logaritmus képzés után:

$$\ln [A(z)] = \ln (1 - z_1 z^{-1}) + \ln (1 - z_2 z^{-1}) + \dots + \ln (1 - z_M z^{-1}). \quad (12)$$

Felhasználva (12)-ben, hogy:

$$\ln (1 - e) = - \sum_{k=1}^{\infty} \frac{e^k}{k}, \quad (13)$$

a (14) egyenletet kapjuk:

$$\ln [A(z)] = - \sum_{n=1}^M \sum_{k=1}^{\infty} \frac{(z_n z^{-1})^k}{k}. \quad (14)$$

Az összegzés sorrendjét felcserélve:

$$\ln [A(z)] = - \sum_{k=1}^{\infty} \frac{1}{k} \sum_{n=1}^M z_n^k z^{-k}. \quad (15)$$

A (15) összefüggést a (3)-mal összehasonlítva:

$$C(k) = - \frac{1}{k} \sum_{n=1}^M z_n^k \quad (16)$$

adódik, ami a polinomegyütthatók és a cepstrumegyütthatók átszámítási formuláját adja.

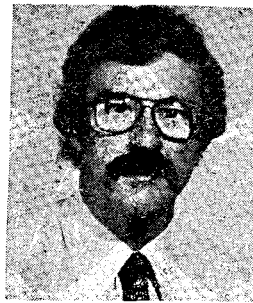
A cepstrumtávolságot tehát az alábbi módon határozhatjuk meg. A be- és kimeneti jeleken a lineáris predikció valamelyik eljárásával (F.1 és [1, 4, 10]) meghatározzuk a modellező $A(z)$ polinomot. $A(z)$ ismeretében a (16) és a (2) összefüggésekkel egyszerűen meghatározhatjuk a cepstrumtávolságot. A kodek be- és kimeneti jeleinek cepstrumtávolsága [6] szerint viszont jól egybeesik a beszédfeldolgozás minőségének szubjektív ítéletével. Laboratóriumi mérésekhez valamilyen alkalmas mérőjelet kell használni. Ezt a mérőjelet ismert együtthatójú lineáris prediktor generálhatja, és így csak a kimenőjelen kell elvégezni a prediktor együtthatók meghatározását.

A cepstrumtávolság (2) definíciójában csak véges számú cepstrumegyütthatót vettünk figyelembe. A véges sok cepstrumegyüttható a (3), illetve az (F.2.10) szerinti végtelen sok cepstrumegyütthatón alkalmazott ablakolással (cepstrum simítás, liftering) határozható meg. A [4] irodalom egyszerű négyzetleges ablakot alkalmaz.

A cepstrumegyüttható tulajdonságaiból következik, hogy a (2) összefüggéssel definiált CD cepstrumtávolság kielégíti a 2. fejezet elején tett matematikai (első és második) és gyakorlati (harmadik és negyedik) követelményeket.

3. Konklúzió

A CCITT néhány cég bevonásával széles körű vizsgálatokat végzett beszédátviteli rendszerek minősítésére



UHERECZKY LÁSZLÓ

A BME Híradástechnika Szakán végzett 1966-ban.

1966—1977 között a Telefongyárban dolgozott a fejlesztésén, 1973-tól a számítástechnikai Fejlesztési Főosztály vezetőjeként. 1977-től a TKI tudományos osztály-vezetője. 1970-ben ösztöndíjként dolgozott Japánban a Fujitsu Ltd-nél és a tokiói egyetemen folytatott tanulmányokat. 1978—79-ben a National Physical Laboratóriumban, Angliában, vendégkutatóként adatátviteli protokollal jellemzőinek vizsgálatával foglalkozott. Szakmai érdeklődése számítógépes kommunikáció, mikroprocesszoros rendszerek.

vonatkozóan, és azt találta, hogy a felhasználók szubjektív megítélése jól egyezik a dolgozatban ismertetett cepstrumtávolság objektív mérőszámaival [5], [6], [9]. Ezért a mérőjelet ismert együtthatójú lineáris prediktorral generálva csak az átviteli rendszer kimenetén kell a folyamatot leíró prediktor-együtthatókat, majd segítségükkel a cepstrumtávolságot meghatározni, hogy a beszédátviteli minőségére objektív mutatót kapjunk. Egy ilyen módszer kifejtése tehát nagyban segítené a különbözőképp kódolt beszédátviteli rendszerek nagyarányú elterjedése előtti objektív minősítését, és időrabló, drága szubjektív vizsgálatokat takarítana meg.

4. Köszönetnyilvánítás

Köszönettel tartozunk dr. Gordos Gézának, Korda Tibornak és dr. Lajtha Györgynek, akik irodalmi forrásokkal és ötletekkel voltak segítségünkre a beszéd-kódoló minősítésével kapcsolatos útkeresésünk során.

F.1. A lineáris predikció alapjai [1, 2, 4, 10]

A lineáris predikció jelek olyan modellezési eljárása, amelynél a jel jelenlegi értékét múltbeli értékeinek és a modellező rendszer gerjesztésének jelenlegi és múltbeli értékeinek lineáris kombinációjával becsüljük. A tapasztalat azt mutatja [1], hogy $p \approx 8-14$ fokú prediktorral a legtöbb hangzó esetében a predikciós hiba elegendően kicsi lesz, tehát pl. az 1. ábra szerinti predikciós beszédmodell a beszédet jól írja le.

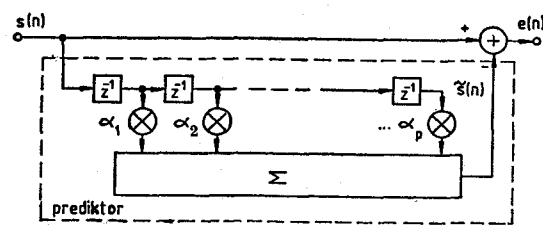
Ha ismert $e(n)$ és az α_k együtthatóhalmaz (valamint a G erősítés), akkor az $s(n)$ jel az 1.b. ábra szintézismodelljének kimenőjele. Ha ismert az $s(n)$ jel és az α_k együtthatóhalmaz (valamint a G erősítés), akkor az $e(n)$ jel az 1.a. ábra analízismodelljének kimenőjele.

Az α_k együtthatók (és a G erősítés) meghatározása a predikció alapvető feladata. Az együtthatók meghatározása a modellezendő $s(n)$ sorozatból történik oly módon, hogy a becsült $\hat{s}(n)$ sorozat n . elemét az $s(n)$ sorozat előző p darab elemének lineáris kombinációjával becsüljük:

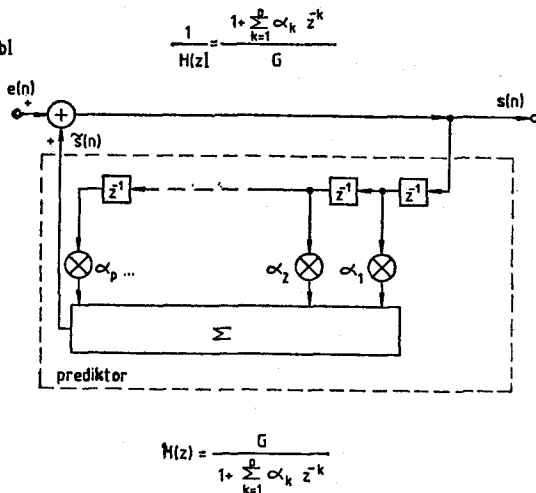
$$\hat{s}(n) = \sum_{k=1}^p \alpha_k s(n-k). \quad (F.1.1)$$

A becsült $\hat{s}(n)$ és a becsülendő $s(n)$ sorozat eltérése a predikciós hiba:

a)



b)



1. ábra Beszédeltés a lineáris prediktor közvetlen struktúráival
a) analízis modell, b) szintézis modell

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k). \quad (F.1.2.)$$

A predikciós hiba általában mintáról mintára változik. A predikció feladata éppen abból áll, hogy az α_k együtthatókat úgy határozzuk meg, hogy a predikciós hiba valamilyen értelemben (általában négyzetes középértékben) kicsi legyen. Ekkor az α együtthatók az $s(n)$ sorozat jó, tömör jellemzésére használhatók. A különböző típusú $s(n)$ sorozatok (determinisztikus, sztochasztikus stb.) és a különböző hibakritériumok (négyzetösszeg, abszolútérték maximum stb.) a predikciós feladat alcsoportjait adják.

A beszédjel bizonyos részletei egy-egy rövidebb időszakaszban jól modellezhetők olyan lineáris időváriáns rendszer súlyfüggvényével, amely tisztán pólusokat tartalmaz [1]. A $H(z)$ akkor fogadható el a beszédeltés mechanizmusának jó modelljeként, ha kis $e(n)$ hibajel, vagyis az $s(n)$ idősort (a beszédet) jól közelítő $\hat{s}(n)$ jelet produkál. Ha $s(n)$ a súlyfüggvény, akkor $e(n)$ (a $t=0$ időpillanattól eltekintve) nulla, és ha $s(n) \approx \hat{s}(n)$, akkor a predikciós együtthatók a beszéd jó, tömör jellemzésére használhatók.

Maga a teljes beszéd folyamat a hangzók állandósult időszakazaiból és az azokat összekötő átmenetekből áll. Az időszakaszok hosszában megnyilvánuló időstruktúra és az egyes időszakaszokban érvényes spektrális amplitúdó-karakterisztika (melyet a predikciós együtthatók meghatároznak) kellő részletességgel jellemzi a beszédet [1]. A beszéd modellezéséhez nem szükséges a hangzók hullámformájának pontos leírása! Az állandósult időszakok jellemzése a spektrum fázisinformációja nélkül is kellő mélységben elvégezhető.

F.2. A cepstrumanalízis alapjai [1, 2, 8, 13]

Az $y(t)$ folytonos idejű, Fourier transzformálható jel (teljesítmény) cepstruma:

$$C_y(q) = |F^{-1}\{\ln |F\{y(t)\}|^2\}|^2, \quad (F.2.1.)$$

azaz a Fourier transzformált abszolútérték négyzete, logaritmus, inverz Fourier transzformáltjának abszolútérték négyzete.

A cepstrum fogalmát Bogert és társai [13] 1963-ban alkották arra a célra, hogy az időfüggvényen „összekeveredett” echók felismerhetők legyenek. Könnyű belátni ugyanis a cepstrum első lényeges tulajdonságát, nevezetesen azt, hogy az echók, amelyek a jel spektrumán a 1. é. leltetéssel fordítva arányos hullámhosszúságú hullámok formájában jelentkeznek, a cepstrumon éles kiemelkedések (közelítőleg Dirac-szimbólumok) formájában figyelhetők meg. Legyen u_i az $y(t)$ jel és τ idejű késleltetettjének (az echónak) összege:

$$y(t) = x(t) + ax(t-\tau), \quad (F.2.2.)$$

ahol a az echo csillapodása ($a < 1$). A Fourier transzformált abszolútértékének négyzete:

$$|F\{y(t)\}|^2 = G_y(\omega) = |X(\omega)|^2 |1 + ae^{-j\omega\tau}|^2. \quad (F.2.3.)$$

A második tényezőben az abszolútérték négyzetet kiszámolva:

$$G_y(\omega) = G_x(\omega)(1 + 2a \cos \omega\tau + a^2). \quad (F.2.4.)$$

Kis a értékek esetén logaritmus képzés után:

$$\ln G_y(\omega) = \ln G_x(\omega) + \ln(1 + 2a \cos \omega\tau + a^2),$$

$$\ln G_y(\omega) \approx \ln G_x(\omega) + 2a \cos \omega\tau. \quad (F.2.5.)$$

Az $\ln G_y(\omega)$ -ből F^{-1} transzformációval és abszolútérték-négyzet képzéssel számíthatjuk a $C_y(q)$ cepstrumot, amin az $F^{-1}\{2a \cos \omega\tau\}$ tag miatt Dirac impulzusok adódnak $F^{-1}\{\ln G_x(\omega)\}$ -hoz:

$$C_y(q) = |F^{-1}\{\ln G_y(\omega)\}|^2 \approx C_x(q) + 2a\delta(q+\tau) + 2a\delta(q-\tau). \quad (F.2.6.)$$

A cepstrum második fontos tulajdonsága, hogy segítségével jelek periodicitása (vagy kvázi periodicitása) könnyen felderíthető.

Legyen az $y(t)$ periodikus jel:

$$y(t) = p(t) * \sum_{n=-\infty}^{\infty} \delta(t - nT_p), \quad (F.2.7.)$$

alakú, ahol $*$ konvolúciót jelent, T_p a periódusidő és $p(t)$ a jel egy periódusának időfüggvénye. Az (F.2.7.) összefüggés mindkét oldalán $|F\{\cdot\}|^2$ -et számolva:

$$G_y(\omega) = |P(\omega)|^2 \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2n\pi}{T_p}\right), \quad (F.2.8.)$$

ahol felhasználtuk, hogy a Dirac-impulzus sorozat Fourier transzformáltja is Dirac-impulzus sorozat.

Az (F.2.8.) összefüggésben elvégezve az abszolútértéknégyzet képzését, valamint a cepstrumszámítás további műveleteit (\ln , F^{-1} , $|\cdot|^2$), hosszadalmas levezetés után azt kapjuk, hogy a periodikus jel cepstrumában a periodicitás ténye élesen kiemelkedő maximumok (Dirac-impulzusok) formájában megfigyelhető.

A cepstrum transzformáció harmadik hasznos tulajdonsága, hogy két, diszjunkt tartóval rendelkező cepstrumú jel konvolúciójából a cepstrumszámítás segítségével az egyes összetevő jelek jellegére könnyen következtetni lehet, azaz dekonvolúciót végezhetünk. Legyen:

$$f(t) = x(t) * y(t). \quad (F.2.9.)$$

Alkalmazva (F.2.1.)-et:

$$\begin{aligned} C_f(q) &= |F^{-1}\{\ln(G_x(\omega)G_y(\omega))\}|^2 = \\ &= |F^{-1}\{\ln G_x(\omega)\} + F^{-1}\{\ln G_y(\omega)\}|^2 = \\ &= [F^{-1}\{\ln G_x(\omega)\} + F^{-1}\{\ln G_y(\omega)\}] \cdot \\ &\quad \cdot [F^{-1}\{\ln G_x(\omega)\} + F^{-1}\{\ln G_y(\omega)\}]^* = \\ &= C_x(q) + C_y(q) + [F^{-1}\{\ln G_y(\omega)\}] \cdot \\ &\quad \cdot [F^{-1}\{\ln G_x(\omega)\}]^* + [F^{-1}\{\ln G_x(\omega)\}] \cdot \\ &\quad \cdot [F^{-1}\{\ln G_y(\omega)\}]^*. \quad (F.2.10.) \end{aligned}$$

Feltételezve, hogy C_x és C_y tartója diszjunkt, azaz ha valamely q -ra:

$$C_x(q) \neq 0, \text{ akkor } C_y(q) = 0, \text{ és ha} \quad (F.2.11.)$$

$$C_y(q) \neq 0, \text{ akkor } C_x(q) = 0,$$

akkor:

$$C_f(q) = C_x(q) + C_y(q), \quad (F.2.12.)$$

azaz a konvolúciós jel cepstruma az összetevő jelek cepstrumának összegeként adódik. Ha az (F.2.11.) feltétel csak közelítőleg teljesül — azaz a cepstrumok csak közelítőleg diszjunktak — (ez a helyzet beszédjel gerjesztésekor), akkor (F.2.12.) is csak közelítőleg igaz.

A cepstrum három lényeges tulajdonsága (echók felderítése, periodicitás felderítése, konvolúció felbontása) fontos szerepet játszik a beszédfeldolgozásban való alkalmazásoknál [1].

Magánhangzók ejtések a beszédjel kváziperiodikus rezgés, amelynek spektruma közelítőleg vonalas. A vonalas spektrum burkolójának helyi maximumai a formánsfrekvenciák. Általában az első három formánsfrekvencia a hangzóra, a negyedik, ötödik és hatodik a beszélő személyre jellemző [1]. A további formánsfrekvenciák nem jelentősek.

A formánsfrekvenciák nem feltétlenül esnek egybe a spektrumvonalakkal. A formánsjellemzők (a maximum helye, értéke, a kiemelkedés szélessége) meghatározásához a spektrumot simítani kell, és meg kell határozni a spektrum burkolóját. Ha a simítás túl erős, akkor egyes formánsok elveszhetnek, ha túl gyenge, akkor olyan helyi maximumok maradhatnak, melyek nem formánsok. A spektrum simításának nehézségei a cepstrumon való műveletvégzéssel megkerülhetők [1].

A beszédkezelés mechanikai rezonátorral (az ún. akusztikus csőmodellel) modellezhető. A modell megalkotásához a gerjesztő jel és a rezonátor súlyfüggvényének konvolúciójából álló beszédjelből külön kell tudni választani a gerjesztő jelet és a rezonátor súlyfüggvényét. Pontosabban az egyik jellemző ismeretében a beszédből az (F.2.12.) összefüggés segítségével a másik meghatározható.

A cepstrum fogalma bizonyos rokonságot mutat a spektrum fogalmával. A független változó (q) idő dimenziójú ugyan, de a Fourier transzformált frekvencia változójával rokon tulajdonságokat mutat (pl. periodikus jelek cepstruma vonalas). A cepstrumanalízis új fogalmainak tárgyalásához szükséges szakvakat ezért a spektrumanalízis hasonló fogalmaiból származtatják többnyire az első szótag megfordításával (1. táblázat).

1. táblázat

A cepstrumanalízis kifejezése

A spektrumanalízis kifejezései	A cepstrumanalízis kifejezései
spectrum	cepstrum
frequency	quefrequency
magnitude	gamnitude
phase	saphe
harmonic	rahmonic
filtering	liftering
period	repiod

Ha a feldolgozandó jel diszkrét idejű, akkor a hozzá tartozó cepstrumot a:

$$C_y(n) = |Z^{-1}\{\ln |Z\{y(m)\}|\}^2 \quad (F.2.13.)$$

kifejezés, ha stacionárius sztochasztikus folyamat, akkor a:

$$C_y(q) = |F^{-1}\{\ln S_y(\omega)\}|^2 \quad (F.2.14.)$$

kifejezés definiálja, ahol $S_y(\omega)$ az $y(t)$ stacionárius sztochasztikus folyamat spektrális sűrűsége.

IRODALOM

- [1] Gordos G.; Takács Gy.: Digitális beszédfeldolgozás, Műszaki Könyvkiadó, Budapest, 1983.
- [2] N. Hesselmann: Digitális jelfeldolgozás, Műszaki Könyvkiadó, Budapest, 1985.
- [3] Lajtha Gy.: Távközlő hálózatok elmélete és tervezése, Műszaki Könyvkiadó, Budapest, 1971.
- [4] M. L. Honig; D. G. Messerschmitt: Adaptive Filters. Structures, Algorithms, and Applications. Kluwer Academic Publishers, Boston, 1984.
- [5] NTT: Calculation Method of OPINE. CCITT Contribution, CCM—XII—235—E, 1984.
- [6] NTT: Proposal of Objective Quality Measure for Voiceband Codecs. CCITT Contribution, COM—XII—8—E, 1985.
- [7] A. H. Gray; J. D. Markel: Distance Measures for Speech Processing. IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP—24, No. 5, Oct. 1976, pp. 380—391.
- [8] D. G. Childers; D. P. Skinner; R. C. Kemerait: The Cepstrum: A Guide to Processing. Proc. of the IEEE, Vol. 65, No. 10, Oct. 1977.
- [9] NTT: Laws of Addition for Opinion Equivalent Q in Low-Bit-Rate Speech Coding. CCITT Contribution, COM—XII—9—E 1985.
- [10] J. Mokhou: Linear Prediction: A Tutorial Review. Proc. of the IEEE, Vol. 63, No. 6, April, 1975, pp. 561—579.
- [11] Gordos G.; Korda Ti.; Kovács P.: Szűrő nélkül működő automatikus torzításmérő. 5. Országos Elektronikus Műszer- és Méréstechnikai Konferencia. Budapest, 1980. március, 338—347. oldal
- [12] Gordos G.: Személyes közlés. 1985.
- [13] B. P. Bogert; M. J. Healy; J. W. Tukey: The Quefrequency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking. Time Series Analysis, M. Rosenblatt, Ed. New York, Wiley, 1963. Chap. 15, pp. 209—243.