

# Beszéd-detektorok paraméterei, realizálása és alkalmazása SCPC rendszerekben

BÁCS ERNŐ—HANZÓ LAJOS—UHERECZKY LÁSZLÓ  
(TKI)



## ÖSSZEFOGLALÁS

Az SCPC (Single Channel per Carrier) rendszerekben a beszéd-detektor feladata, hogy ha a kijelölt csatorna bemenetén beszéd van, akkor engedélyezze, ha pedig beszédszünet van, akkor tiltsa le a vivő kiadását. Ezáltal értékes szatellit energiát takaríthatunk meg és a teljesítménykorlátos rendszer terhelését csökkentjük. Mivel a beszéd-detektor időosztásos kihasználása SCPC rendszerekben nem valósítható meg egyszerűen, a lehető legolcsóbb real-time realizációt céloztuk meg. A beszéd-detektor-paraméterek áttekintése után ismertetjük az adaptív beszéd-detektor folyamatábráját, amelyik közvetlen realizálásra alkalmas a kiválasztott I 8085 alapú multiprocesszor rendszerben. A detektort objektív és szubjektív tesztekkel minősítjük.

## 1. Bevezetés

Beszéd-detektorokat (a továbbiakban *BD*) a távközlés számos területén alkalmaznak, de a legtöbb alkalmazásban — függetlenül az adott rendszerkörnyezettől — alapvető feladatuk, hogy megkülönböztessék a beszédet a zajtól [1]–[12]. A dolgozatban olyan — lehetőségeinkhez mérten — csekély ráfordításigényű, I 8085 alapú multiprocesszorral realizált real-time *BD*-t ismertetünk, amelyik előnyösen alkalmazható rögzített csatornakiosztású (SCPC = Single Channel per Carrier) szatellit rendszerekben és érzékenyebb beszédre, mint zajra.

A *BD* specifikációk 2. fejezetbeli áttekintése után a 3. fejezetben összefoglaljuk a *BD* főbb paramétereit és a beszédérzékenységre gyakorolt hatásait [7], a 4. fejezetben a *BD* realizálási folyamatábráját ismertetjük, majd az 5. pontban a *BD* Intercsat SCPC rendszerben történő megvalósításáról számolunk be. Végül a 6. pontban a realizált *BD* objektív és szubjektív teszteredményeiről számolunk be.

## 2. A specifikációk összefoglalása

Az SCPC rendszerekben a *BD* feladata, hogy felismerje, ha beszéd érkezik a csatorna bemenetére, és engedélyezze a vivő kiadását. Így annak köszönhetően, hogy beszédszünetekben a vivő le van tiltva, értékes szatellit-energiát takarítunk meg és javítjuk a teljesítménykorlátos rendszer energiamérlegét.

A *BD* „A”-karakterisztikával kompendált 7 bites PCM mintákon kell, hogy működjön és a működési küszöb-szinteket és időket [16] specifikálja részletesen.

A [15] dolgozat vizsgálatai azt mutatják, hogy software és hardware fejlesztési munkaigény, költség, gyártási ráfordítások és távlati szolgáltatások miatti flexibilis továbbfejlesztettség szempontjából

## BÁCS ERNŐ

A BME műszer- és irányítástechnika szakán végzett 1974-ben. Azóta a Távközlési Kutató Intézetben dolgozik. Szak-

mai érdeklődése mikroprocesszoros rendszerek hardware és software kérdéseire, valamint digitális áramkörök hibafelderítésére terjed ki.

I 8085-ös MOS processzor alapú multiprocesszoros realizáció tűnik számunkra optimálisnak. Így ezzel a struktúrával harmonizáló, gyors és konfidens döntésű real-time algoritmust keresünk.

## 3. Beszéd-detektor paraméterek és a beszédérzékenységre gyakorolt hatásai

A *BD* paramétereit úgy kell megválasztani, hogy kompromisszumot teremtsenek az alábbi, alapvetően ellentmondásos kritériumok között:

1. minimálni a zaj hatására történő hamis triggerelések számát;
2. akkor is átvinni a beszédet, ha teljesítménye rövid időre a *BD* küszöbszintje alatt marad.

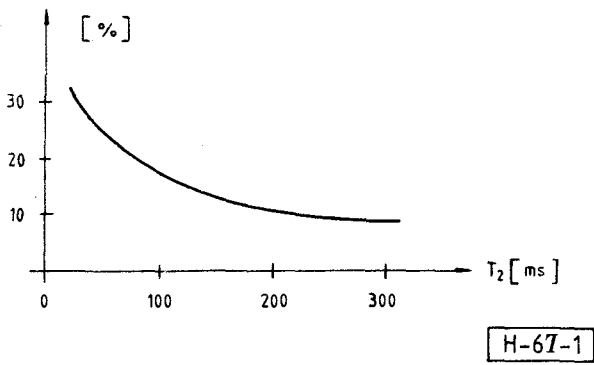
Így a legfontosabb *BD* paraméterekre vonatkozó követelmények már plauzibilisek:

- a) A beszéd-szegmensek kezdeti levágásának minimalizálása érdekében a *BD* a lehető leggyorsabban meg kell, hogy különböztesse a beszédet a zajtól, s a  $K_B$  beszédküszöb a lehető legérzékenyebb kell, hogy legyen.
- b) Másrészt a  $K_B$  beszédküszöb túl alacsony értéke zaj hatására történő hamis triggerelését okozhat.

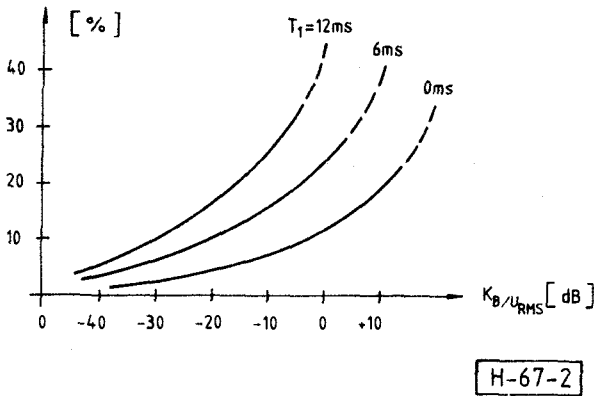
Ezek alapján olyan kompromisszumot köthetünk [4], hogy a  $K_B$  beszéd-küszöböt kissé a tipikus csatornazaj fölé, de a legalacsonyabb beszéd-szint alá választjuk. A beszéd-szegmensek végének csonkolása kikapcsolási késleltetés ( $T_2$ ) segítségével akadályozható meg, a hamis zajtriggerelés pedig úgy minimalizálható, ha a vonalon detektált energiát akkor tekintjük csak átviendő beszédnek, amennyiben időtartama meghaladja a  $T_1$  bekapcsolási késleltetés értékét.

A [8] dolgozat mássalhangzó—magánhangzó—mássalhangzó felépítésű logatomos vizsgálatai azt mutatják, hogy a magánhangzók érthetőségét lényegében nem érinti a *BD*, ezért a  $K_B$ ,  $T_1$  és  $T_2$  paraméterek értékét lényegében a szókezdeti és szóvégi mássalhangzók érthetősége determinálja. Mivel a paraméterek kölcsönhatásban vannak egymással, célszerű egyszerre csak egynek a hatását vizsgálni.

Beérkezett: 1985. V. 12. (□)



1. ábra. Szókezdeti mássalhangzóhibák valószínűsége (%) a  $K_B$  beszédküszöb (dB) függvényében



2. ábra. Szóvégi mássalhangzóhibák valószínűsége (%) a  $T_2$  kikapcsolási késleltetés (ms) függvényében

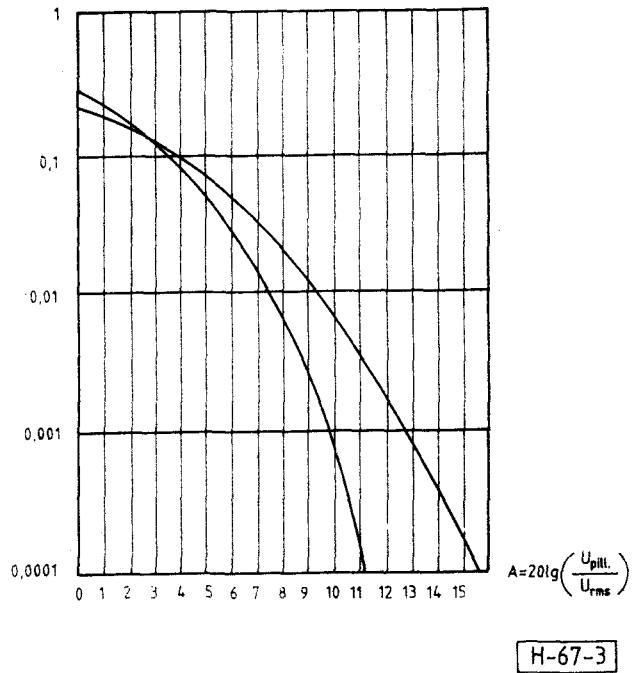
A  $K_B$  beszédküszöb és a  $T_1$  bekapcsolási késleltetés megválasztásához az 1. ábra ad útmutatást, a  $T_2$  kikapcsolási késleltetés pedig a szóvégi mássalhangzóhibák alapján méretezhető a 2. ábra segítségével.

A fentiek összefoglalásaként azt mondhatjuk [8], hogy a  $K_B$  beszédküszöb értékétől szinte függetlenül  $T_2 < 50$  ms esetén a zárómássalhangzók hibavalószínűsége nagy, ugyanakkor kis  $T_1$  érték és  $K_B < -15$  dB mellett  $T_2 > 150$  ms választással az érthetőség már nem javul.

Annak érdekében, hogy a zajimmunitás és beszédérzékenység ellentmondó követelménye miatti tervezési kompromisszumtól megszabaduljunk, kihasználhatjuk a zaj és a beszéd eltérő statisztikai tulajdonságait.

Így a  $BD$  érzékenyebb lesz beszédre, mint zajra [1]. Ennek érdekében modellezzük az emberi beszédet exponenciális valószínűsűrsűségű folyamattal, a csatornazajt pedig fehér Gauss-zajjal (l. 3. ábra).

A 3. ábra alapján megállapíthatjuk, hogy a két sűrűségfüggvény kb. a pillanatnyi rms=3,0 dB értéknél metszi egymást. Ettől kezdve a beszéd esetében valószínűbb, hogy pillanatnyi értéke nagyobb, mint az rms érték. Például 10 dB-lel az rms érték fölött már egy nagyságrenddel „gyakrabban tartózkodik” a beszéd, mint a zaj, ha azonos rms értékűek. Ráadásul a [13]-ban ismertetett mérések azt tanúsítják, hogy a beszéd eloszlása a gyakorlatban az alapul vett exponenciális eloszlásnál még „csúcsosabb”. Így a fentiek alapján válasszuk a  $BD$



3. ábra. Az exponenciális eloszlásúnak feltételezett beszéd és a fehér Gauss-zaj komplementer kumulatív eloszlásfüggvénye

küszöbszintjét ebben a számunkra kedvező tartományban, és vizsgáljuk több (pl. 3–4)  $PCM$  minta együttes túllépését annak érdekében, hogy eldöntsük, hogy beszéd van-e a  $BD$  bemenetén. Így olyan  $BD$ -t kapunk, amelyik kb. 8–10 dB-lel érzékenyebb beszédre, mint zajra, és gyorsabb reakcióidejű mint az energia-detektorok.

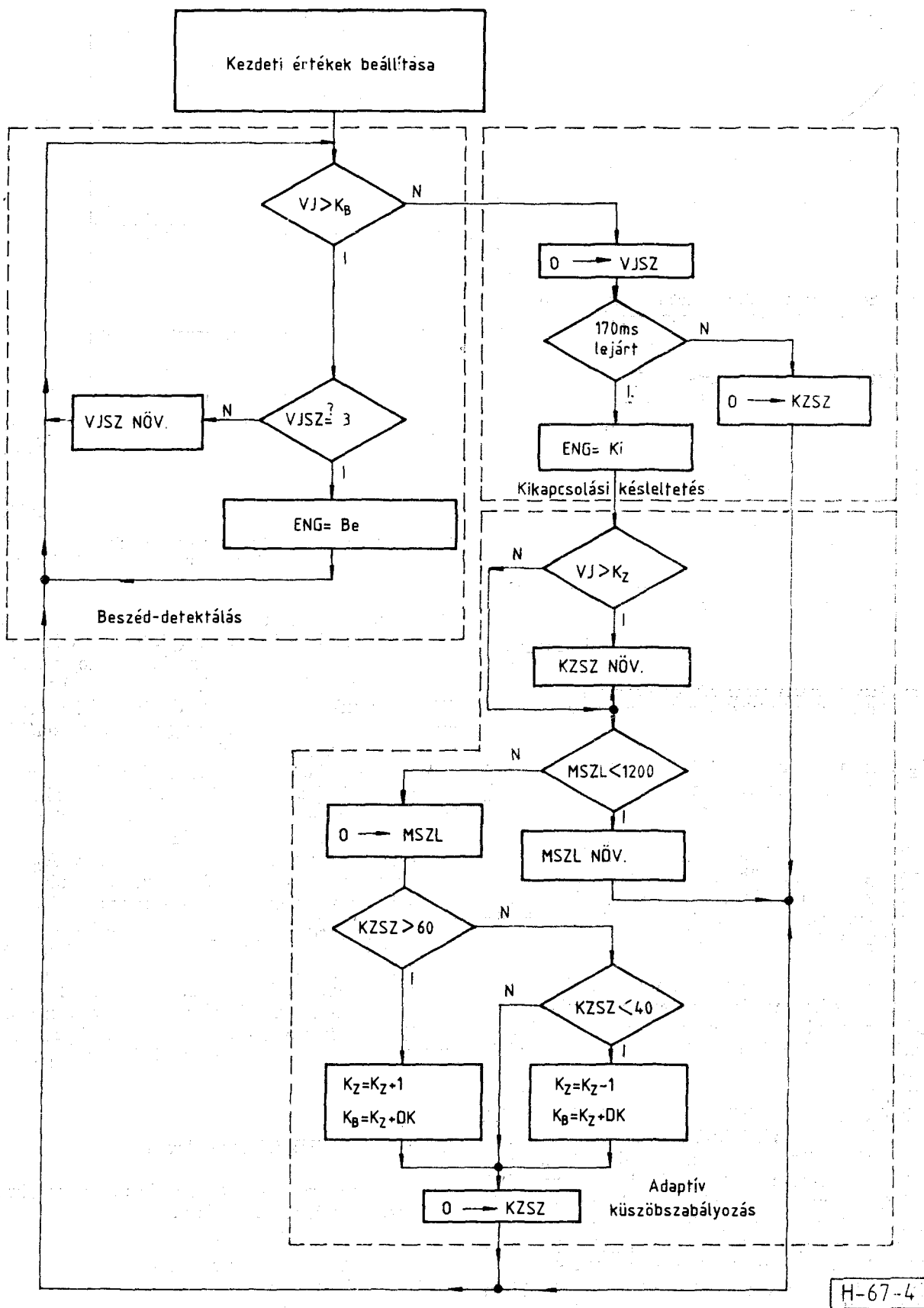
A  $BD$  hatékonysága tovább fokozható azáltal, ha küszöbszintjét az aktuális csatornazaj szintjéhez igazítjuk, azaz adaptívvá tesszük [8]. A küszöb-adaptációs folyamat lényege, hogy ha épp nincs beszéd a csatorna bemenetén, akkor bevezetve a  $K_Z$  zajküszöböt, ennek értékét úgy szabályozzuk, hogy a zajminták adott hányada — a nagy zajcsúcsok — lépje túl. A  $K_B$  beszédküszöböt ezek után egy kis biztonsági távolsággal a  $K_Z$  zajküszöb fölé kell választani, mégpedig úgy, hogy a  $K_B$  — a korábbi megfontolások alapján — a 3. ábra említett kedvező tartományába essen. Ha viszont beszéd van a csatornában, akkor a  $K_Z$  zajküszöb szabályozását letiltjuk, amennyiben az átviteli közegre jellemző maximális zajszint értékét elérte.

#### 4. Az adaptív $BD$ folyamatábrája

Az eddigiek alapján a  $BD$  realizálási folyamatábrája már megtervezhető (l. 4. ábra).

Az alkalmazott rövidítések:

- $VJ$ : A vett jel egy mintája
- $K_B$ : Beszédküszöb
- $K_Z$ : Zajküszöb
- $K_M$ : Maximális zajküszöb
- $VJSZ$ : Vett jel minta-számláló
- $Eng$ : Vivőkiadás engedélyezése
- $KZSZ$ :  $K_Z$ -t meghaladó minták számlálója



4. ábra. Az adaptív real-time BD realizálási folyamatábrája

MSZL: Mintaszámláló

DK:  $DK = K_B - K_Z$ , azaz a beszéd és zajküszöb távolsága.

A *BD* folyamatábrája alapvetően három fő részből áll:

1. Beszéddetektálás
2. Kikapcsolási késleltetés kezelése
3. Adaptív küszöbszabályozás.

1. A *BD* akkor engedélyezi a vivő kiadását, ha adott számú — pl. esetünkben három — egymást követő vett jel-beli *PCM* minta túllépi a  $K_B$  beszédküszöböt.
2. A *BD* akkor tiltja le a vivőt, ha a 170 ms-os kikapcsolási késleltetés lejárt és időtartama alatt egyszer sem érkezett három egymást követő *PCM* minta a  $K_B$  beszédküszöb fölé.
3. Ha beszédet nem detektáltunk, akkor pl. 1200 zajmintából statisztikát készítünk, és számláljuk a  $K_Z$ -t meghaladó mintákat. Ha ezek száma — *KZSZ* — nem esik pl. a  $40 < KZSZ < 60$  intervallumba, akkor  $K_Z$ -t úgy szabályozzuk, hogy a feltétel teljesüljön és a  $K_B$  beszédküszöböt szintén módosítjuk ( $K_B = K_Z + DK$ ).

A *BD* hatékonysága néhány ötlet segítségével még tovább fokozható [17], ezek tárgyalására azonban az olcsó real-time realizáció igénye miatt most nem térünk ki.

## 5. A *BD* alkalmazása az Interscat SCPC rendszerkörnyezetben

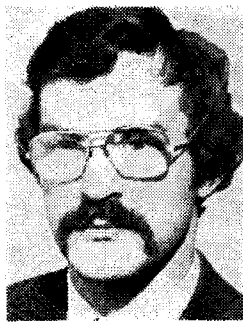
A [15] dolgozat árral, karbantarthatósággal, hardware és software fejlesztési befektetéssel, távlati szolgáltatásbeli bővítéssel kapcsolatos megfontolásai alapján az alapsávi csatornaegység realizálására optimális megoldásként *I 8085* alapú multiprocesszor struktúra adódik. A teljes *PCM* csatornaegység négy *I 8085*-re épül, amelyek közül egy elegendő a *BD* feladatának ellátására. Ezt úgy értük el, hogy a folyamatábrára és a rendszertechnika tervezésénél a feladatot kedvezően szegmentáltuk, azaz ha beszéd van jelen és  $K_Z$  elérte a küszöb maximumot, akkor nem kell és nem szabad a küszöböt szabályozni, és fordítva, azaz a küszöböt szabályozni kell, ha nincs beszédátvitel. Ezzel az optimális feladatstrukturálás-



HANZÓ LAJOS

A *BME* híradástechnika szakán végzett 1976-

ban, s diplomátvételt harmadik, *TDK* dolgozatát első díjjal jutalmazták. 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 tüntették ki. Szakmai érdeklődési körébe az információátvitellel kapcsolatos jelfeldolgozási és rendszertechnikai problémák tartoznak.



UHERECZKY  
LÁSZLÓ

A *BME* híradástechnika szakán szerzett diplomát 1966-ban. 1966—1977 között a Telefon-

gyárban dolgozott a fejlesztésben, 1973-tól a Számítástechnikai Fejlesztési Főosztály vezetőjeként. 1977-től a *TKI*-ban tudományos osztályvezető. 1970-ben ösztöndíjként Japánban a *Fujitsu Ltd*-nél és a *Tokio Egyetemen* folytatott tanulmányokat. 1978—79-ben a *National Physical Laboratórium*ban Angliában vendégkutatóként adatátviteli protokollal jellemzőinek vizsgálatával foglalkozott. Szakmai érdeklődése: számítógépes kommunikáció, mikroprocesszoros rendszerek.

sal elértük, hogy a *BD* processzorának minimális időt kell vesztegetnie a multiprocesszor közös rendszerbuszának lefoglalásával, és költség-effektív realizáció adódik. A *BD* software-jét a magasszintű *CORAL 66* nyelven írtuk és utána konvertáltuk *I 8085 Assembler*-be, ami gyors, jól dokumentálható programfejlesztést eredményezett. A *BD* software terjedelme 0,5 kByte.

## 6. A *BD* mérési tapasztalatai

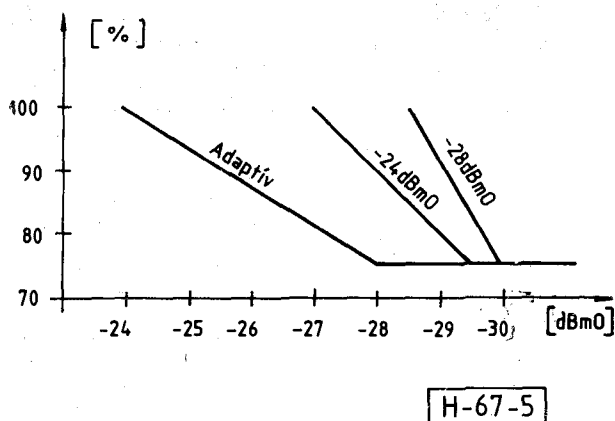
Vizsgáló jelek:

1. 1020 Hz-es szinusz jel.
2. Sávhatárolt fehérzaj (300—3400 Hz).
3. Monológ magyar szöveg.
4. Kisfrekvenciás ( $f < 1$  Hz) négyszög- és fűrészjelek a fehérzaj modulálására.

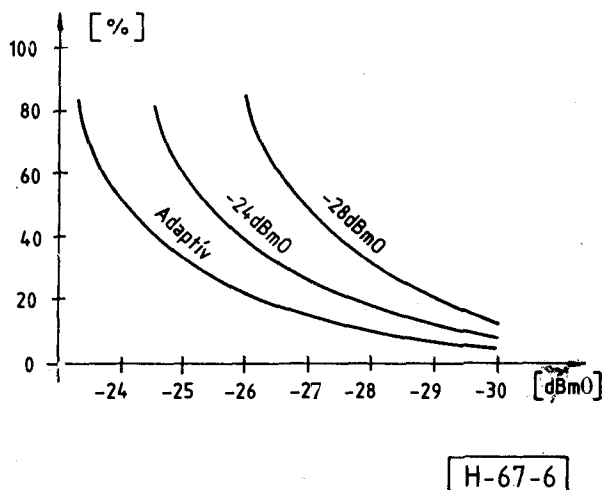
Objektív tesztek:

1. Fix-küszöbű üzemmódban 840 Hz-cel mérve négy küszöbbeállítás lehetséges:  $-20$ ,  $-24$ ,  $-28$ ,  $-32$  dBmO. Adaptív küszöbű üzemmódban a küszöb  $-20$  és  $-32$  dBmO között változhat az aktuális csatornazaj függvényében.
2. A zajimmunitás azzal jellemezhető, hogy a zajteljesítményt a  $K_B$  beszédküszöb alá 8 dB-lel felengedve a zaj a *BD*-t mindössze 1/5 s gyakorisággal aktiválja.
3. A magyar monológ beszéddel és additív fehérzajjal mért beszédaktivitást a zajszint függvényében az 5. ábra mutatja.
4. A fehér Gauss-zaj hatására történő hamis triggerelési rátát a 6. ábrán láthatjuk a zajszint függvényében.

Az 5. és 6. ábra alapján megállapíthatjuk, hogy mind a beszédaktivitás, mind a hamis triggerelés csökken, ha a csatornazaj csökken, ami tetszőleges küszöbszint mellett igaz. Ugyanakkor adott zajszintnél mindkettő csökken, ha a küszöbszintet emeljük, ami szintén kedvező volna számunkra. Ezzel szemben minél magasabb a küszöbszint, annál rosszabb az érthetőségi index, amit a küszöbszint



5. ábra. A mért beszédaktivitás a zajszint függvényében különböző küszöbök mellett (monológ beszéd+fehér Gauss-zajjal mérve)



6. ábra. A hamis triggerelés relatív időtartama a zajszint függvényében különböző küszöbszinteknél (fehér Gauss-zajjal mérve)

alá eső halk beszédrészek elvesztése okoz. Ezért az optimális kompromisszumot az adaptív küszöbszabályzás biztosítja, így ezt az üzemmódot célszerű preferálni.

#### Szubjektív tesztek:

1. Monológ magyar szöveget hallgatva azonosíthatók olyan halk beszédrészek, amelyek csak adaptív küszöbű üzemmódban hallhatók.
2. Lassan változó fehérzajjal terhelt beszédet a  $BD$  kb. 10 dB jel-zaj viszony mellett még éppen érthetően visz át.

#### 7. Összegezés

A  $BD$ -paraméterek és beszédérthetőségre gyakorolt hatásaik rövid áttekintése után megadtuk a  $BD$  köz-

vetlen real-time realizálásra alkalmas folyamatábráját. Beszámoltunk az Intercsat rendszerben való konkrét alkalmazásáról és teljesítőképességét objektív és szubjektív tesztekkel értékeltük.

#### 8. Konklúzió

Konklúzió gyanánt megállapíthatjuk, hogy a fentiek alapján — a kedvező feladat-strukturálásnak köszönhetően — a  $BD$  olcsó  $MOS$  processzorral realizálható [15], eleget tesz a nemzetközi előírásoknak [16], és jó beszédérthetőséget szavatol. Ez azt jelenti, hogy a lassú  $MOS$  processzor képes bizonyos hangfrekvenciás jelfeldolgozási funkciók betöltésére is.

#### I R O D A L O M

- [1] E. Fariello: A Novel Digital Speechdetector for Improving Effective Satellite Capacity. IEEE. Transactions on Communications, Febr. 1972, pp. 55—59.
- [2] A. Busala: Fundamental Considerations in the Design of a Voice-Switched Speakerphone BSTJ, March 1960, pp. 265—294.
- [3] W. F. Clemency—W. D. Goodale: Functional Design of a Voice-Switched Speakerphone, BSTJ, May 1961, pp. 649—668.
- [4] H. Miedema—M. G. Schachtman: TASI Quality — Effect of Speech Detectors and Interpolation, BSTJ, July 1962, pp. 1455—1473.
- [5] P. T. Brady: A Technique for Investigating On-Off Patterns of Speech. BSTJ, January 1965, pp. 1—22.
- [6] W. Y. Chan—D. D. Falconer: Speech Detection for a Voice/Data Radio Terminal. Proc. ICC 1983, pp. F8. 7.1.—F8. 7.5.
- [7] G. J. Barnes: Voice Switching Parameters in Telephony, Electrical Communication. Vol. 47, No. 3, 1972, pp. 186—196.
- [8] J. A. Jankowski: A New Digital Voice-activated Switch. COMSAT Techn. Review, Vol. 6, No. 1, Spring 1976, pp. 159—170.
- [9] K. Bullington—J. M. Fraser: Engineering Aspects of TASI. BSTJ, Vol. 38, pp. 353—364, 1959.
- [10] E. F. O'Neill: TASI-Time Assignment Speech Interpolation. BSTJ, Vol. 37, pp. 83—87. 1969.
- [11] J. M. Fraser—D. B. Bullock—N. G. Long: Overall Characteristics of a TASI System. BSTJ, Vol. 41, pp. 1439—1454, 1962.
- [12] A. M. Werth: SPADE: A PCM FDMA Demand Assignment System for Satellite Communications, presented at the London. Conf. Digital Satellite Comm., Nov. 1969.
- [13] R. L. Richard: Statistical Properties of Speech Signals. Proc. Inst. Electr. Eng. (London), Vol. 3, pp. 944—945, May 1964.
- [14] H. Fletcher: Speech and Hearing in Communication. Princeton, N. J.: Van Nostrand, 1953.
- [15] Bács E.—Gubányi M.—Hanzó L.—Hinsenkamp L.—Uhereczky L.: Az Intercsat rendszer PCM-csatornaegységének rendszertechnikai és realizálási problémáiról. Hir. t., megjel. alatt.
- [16] BG-9-21E (Rev. 3) 31, March 1982, System Specification, pp. 22—24.
- [17] E. Bács—L. Hanzó: A simple real-time adaptive speechdetector for SCPC systems. To be published in the Proc. of ICC 85, Chicago.