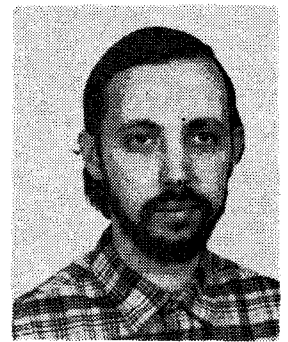


Kötött szótárú beszédszintetizátor távbeszélő-hálózathoz illesztett közszükségleti alkalmazása

NÉMETH GÉZA
BME HEI (BEAG)*



ÖSSZEFOGLALÁS

A cikkben ismertetett készülék a beszédszintézis három lehetséges felhasználói alkalmazását valósítja meg. A Z-80 mikroprocesszorral vezérelt készülék — a Philips MEA 8000 szintetizátorára alapozott magyar fejlesztésű MINIVOX beszédszintetizáló rendszer segítségével — riasztójelek telefonvezetéken való automatikus átvitelére, telefonválaszadásra és hívásfelépítésre használható.

Bevezetés

Napjainkban zajlik az a folyamat, melynek során az automatikus beszédfeldolgozás több területe a laboratóriumokból kikerülve a mindennapi élet részévé válik [10]. Az egyre kisebb bitsebességet elérő digitális beszédátvitelen kívül az ún. beszélő perifériák a legismertebbek. Ma már a komolyabb személyi számítógépekhez kb. 140 dollártól kezdődő áron jó minőségű beszédet előállító egységek kaphatók. Az ágazat fejlődését jól tükrözik a következő adatok [1]: a beszédfeldolgozás területén az USA piaci forgalma 1982-ben mintegy 25 millió dollárt tett ki, ami a BEAG azévi teljes termelése kétszeresének felel meg, de az előrejelzések szerint ez az összeg 1987-re megközelítően 800 millió dollárra nő. Ez már több, mint 1982-ben a Philips vagy a Control Data adatfeldolgozó berendezésekből származó bevétele a nyugat-európai piacon [2].

A beszédfeldolgozás következő területeit különböztetjük meg [3, 4, 10]:

- digitális beszédátvitel: kis adatsebesség, titkosítás.
- beszédszintézis rendszerek: digitális vezérlésű eszköz emberi hangon ad információt. A máig piacra került gyártmányok többsége ide tartozik. Pl.: Speak and Spell, tolmácsgép stb.
- beszélő igazoló és azonosító rendszerek: egy személy az-e, akinek állítja magát ill. több lehetséges beszélő közül ki beszél.
- beszédfelismerő rendszerek: üzenet akusztikus alakról írott alakra hozása.
- testi fogyatékosok segítése: pl. vakok olvasni süketek beszélni tanítása.

* A szerző a BEAG alkalmazottjaként a BME HEI-n nappali szakmérnöki tanulmányokat folytatott. Jelen munkája ott készült.

Beérkezett: 1987. XI. 18. (H)

DR. NÉMETH GÉZA

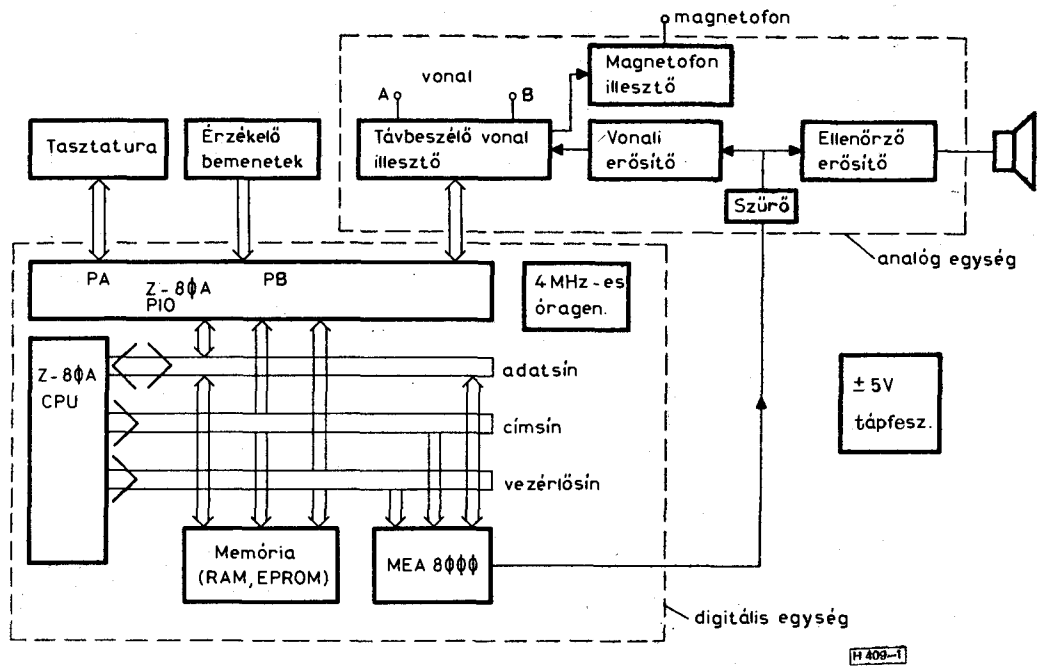
Dr. Németh Géza 1983-ban végzett a BME Villamosmérnöki Karának Híradástechnikai Szakán. Diplomaterve a HTE 1983-as pályázatán

I. díjat nyert. 1985-ben szakmérnöki diplomát, 1986-ban egyetemi doktori fokozatot szerzett. Szakmai érdeklődési területe a digitális beszédfeldolgozás és alkalmazásai.

A fentiekben már említettük a beszédszintézis rendszerek jelentőségét. A közeljövőben még szélesebbkörű alkalmazásuk várható. A piacért folyó versenyben több nagy IC-gyártó cég fejlesztette ki saját egyehipes beszédprocesszorát. A Philips ilyen eszköze, a MEA 8000-es beszédszintetizátor 1982 tavaszán jelent meg a világpiacra.

A BME HEI Átvitel- és Rendszertechnika Osztályán dolgozták ki a MINIVOX fantázianevű programrendszert, ami képes a MEA 8000-et magyar nyelven megszólaltató adatok meghatározására [5]. Mivel ezzel az áramkörrrel igen nagy adattömörítés érhető el (1 másodperc beszédhez kb. 1 kbit adat szükséges a 8 kHz-es PCM átvitel 64 kbit/s-os adatsebességével szemben), ez lehetővé teszi az eddig alkalmazott magnetofonok kiváltását azokban az esetekben, ahol perces nagyságrendű fix szókincsre van szükség. A továbbiakban ismertetésre kerülő Voxalarm fantázianevű berendezés ezt automatikus riasztóként, üzenetrögzítőként és telefonhívóként képes megvalósítani. Az eszköz a következő funkciókat tudja ellátni:

- a valamelyik riasztó bemenetére érkező indító jel hatására automatikusan felhív egy előre beprogramozott telefonszámot, és ha a hívott felvette a kagylót, többször megismételve bemoond egy szintén előre beprogramozott üzenetet. pl.: a bemenetre tűzérzékelőt csatlakoztatva, ha az jelez, felhívja a 05-ös számon a tűzoltókat és kétszer bemoondja (például:) „A kettőszázegyves számú automatikus riasztó beszél. A XI. kerület Fehérvári út ötvenkettes számú épületben a tűzérzékelő jelez”.
- utasítható, hogy ha az adott számra hívás jön, zárja a vonalat, mondjon be egy, a tasztatúráról előre megadott üzenetet, majd indítsa el a hozzá csatlakoztatott magnetofont. Bizonyos idő eltelté után a berendezés figyelmeztető jelzést küld, majd leállítja a magnetofont és megszakítja a vonalat.



1. ábra. Teljes rendszerterv

— a tasztatúráról megadható egy felhívandó telefonszám és egy bemondandó üzenet. A Voxalarm automatikusan felhívja az adott számot, és ha a hívott jelentkezik, bemondja az üzenetet.

A berendezés felépítése

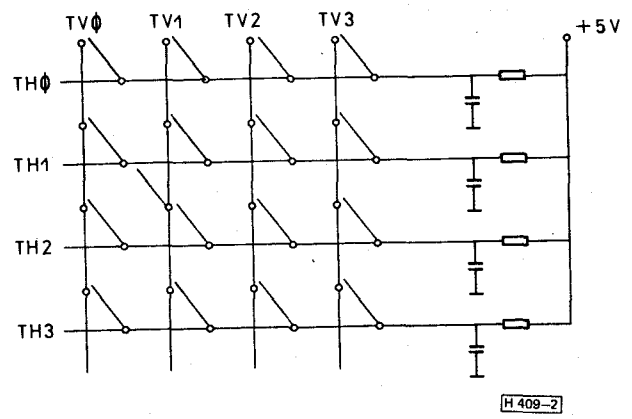
A Voxalarm rendszerterve az 1. ábrán látható. A funkcionálisan összetartozó, szaggatott vonallal körülvett egységek egy nyomtatott áramköri kártyán helyezkednek el.

A digitális egység egy minimálkiépítésű mikrogép. Mindössze egy Z-80 központi egységet és PIO-t, 1 kbyte RAM és 12 kbyte EPROM memóriát, valamint a beszédszintetizátort tartalmazza. Az EPROM-ban 8 kbyte-ot foglal el a kb. 70 másodpercnyi alapszókincs, a fennmaradó 4 kbyte-ban a program és az alapszókincs elemeinek összefűzésére szolgáló kifejezéstáblázat helyezkedik el. A RAM-ban a tasztatúráról beírt adatok, a program változói és a stack található. A beszéd-szintézis teljesen digitálisan vezérelhető, így a központi egység egyszerűen memóriaként kezelhető. A MEA 8000 4 MHz frekvenciájú órajelet igényel és mivel egyik kimenetén megjelenik annak hárommal leosztott változata, ez a jel lassú processzorok órajeleként szolgálhat.

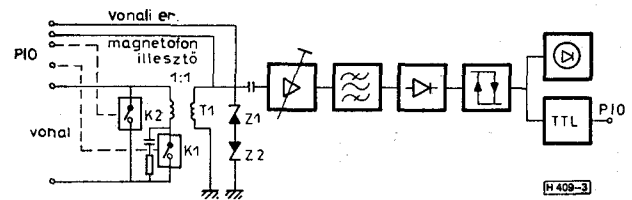
A CPU a PIO egyik portján keresztül kezeli a tasztatúrát, a másik port segítségével pedig az érzékelő bemenetek leolvasását valamint az automatikus telefonhívás és üzenet rögzítés vezérlését végzi. A tasztatúra elvi kapcsolása a 2. ábrán látható [6]. A 4x4-es mátrixelrendezésű 16 kapcsoló minden TH-jelű sora alapállapotban +5V feszültség szinten van. A TV-jelű oszlopokra egymás után alacsony szintet adva, és figyelve, hogy melyik sor kerül erre a szintre, megállapítható,

hogy melyik kapcsolót nyomták le. A 16 kapcsoló közül 10 számjegymegadásra, 5 parancsmegadásra, egy pedig parancslezárára szolgál.

Az analog egység dolgozza fel a szintézis D/A átalakítójának kimenetén és a távbeszélő vonalon megjelenő jeleket. A MEA 8000 kimenő jele egy korrekciós szűrőre kerül, onnan pedig a visszahallgatásra szolgáló ellenőrző erősítő és a távbeszélő vonalra megfelelő szintű beszédjelet adó vonali erősítő bemenetére. A magnetofon illesztő a távbeszélő vonalon megjelenő jelből állít elő az üzenet rögzítéshez megfelelő szintű és impedanciájú kimeneti jelet. A távbeszélő vonal illesztő rendszerterve a 3. ábrán látható. A T1 1:1 át-



2. ábra. A tasztatúra elve



3. ábra. A távbeszélő vonal illesztő rendszerterve

tételű transzformátor a galvanikus leválasztást, a Z1 és Z2 zener-dióda a nagyfeszültségű zavarok elleni védelmet biztosítja. Az egyenáramú horuk zárását és bontását valamint a tárcsázást az eszköz a PIO-n keresztül vezérelt K1, K2 reed-relék segítségével végzi. A K1 reed-relével párhuzamos RC-tag az egyenáramú vonal zárása nélkül juttatja a leosztott csengető jelet a további egységek bemenetére. Ezeknek a feladata az, hogy a 200–500 Hz közti frekvenciájú, legalább –30 dBm szintű jel megjelenését a digitális egység számára jelezzék. Az erősítő fokozat erősítése $-\infty$ és +50 dB között állítható be. A szűrőt egy 200 Hz határfrekvenciájú max. lap., többszörös negatív visszacsatolású másodfokú aktív felüláteresztő és egy 500 Hz határfrekvenciájú elsőfokú aluláteresztő kaszkád kapcsolása valósítja meg. A szűrő kimeneti jele aktív csúcseyenirányítóra, majd 1,9 V és 2,9 V billenési feszültségű hiszterézises komparátorra kerül. A komparátor kimenetének állapotát vizuálisan mutatja a LED-kijelző és a TTL-illesztő teszi a PIO számára leolvashatóvá. Az egyszerűen megvalósítható elrendezés kimenetét jelentő bitet figyelve a program határozza meg a vonal állapotát.

Természetesen a telefon megszületése óta számos megoldás született az automatikus hívásmegvalósítására. A fenti elrendezés nagy előnye, hogy az alkatrészigény jelentős csökkentése mellett képes a hívásfelépítést úgy megvalósítani, hogy hibát gyakorlatilag csak a központ helytelen működése okozhat.

A vezérlőprogram

A vezérlőprogram három részre bontható:

- a MEA 8000 vezérlése,
- automatikus hívásmegvalósítás és hívásfogadás,
- az érzékelő bemenetek és a tasztatúra figyelése és kezelése.

A MEA 8000 vezérlése

A beszéd szintetizátorokat kódolási alapelvük szerint a hullámforma- vagy a forráskódolók közé sorolhatjuk. Az előbbi esetben a tömörített adatokból az eredeti jelalakot kívánjuk minél nagyobb hűséggel visszaállítani, az utóbbiban pedig az agy számára az eredetivel azonos információt hordozó jel előállítását a cél [7]. A MEA 8000 a forráskódolású ún. formánsszintetizátorok közé tartozik, azaz a beszéd első néhány formánsát, a beszéd spektrum burkolójának csúcsait állítja elő [8]. Három változtatható és egy állandó (3500 Hz) frekvenciájú formánst képez. Mindegyik sáv szélessége változtatható. Zöngés beszéd részlet esetén az alaphang frekvencia (pitch) értékét is meg kell adni. A szintetizátorban a jobb adattömörítés érdekében két eljárást alkalmaztak. Egyrészt az alaphang frekvenciát csak a kimondandó üzenet elején kell megadni, a további beszédkereteknél a kevesebb bittel leírható alaphang frekvencia-megváltozásra van szükség, másrészt egy keret hossza 8 és 64 ms között —

2 hatványainak szorzóival — változtatható, így változatlan vagy lineárisan változó jellemzők esetén ugyanaz az adat hosszabb beszéd részletet határoz meg. Egy beszédkerethez négy byte adat tartozik. Ez a következő jellemzőket tartalmazza:

- három formánsfrekvencia,
- négy sáv szélesség,
- amplitúdó,
- a keret időtartama,
- az alaphang frekvencia megváltozása.

A MEA 8000 belső interpoláció segítségével folytonos átmenetet biztosít két keret adatai között és az effektív mintavételi frekvenciát 64 kHz-re hozza, így egyszerű visszaállító szűrő alkalmazható. A szintetizátor a STOP parancs hatására kerül alaphang állapotba. Ebből kezdhető a szöveg kimondatása. A kezdeti alaphang frekvencia megadása után az IC egy státuszbit 0-ról 1-re állításával jelzi, hogy kéri az első beszédkeretet jellemző 4 byte adatot. Ezután kb. 8 ms alatt felkészül az első keret kimondására. Az első keret kimondásának megkezdésekor megkéri a második keret adatait, annak kimondásakor pedig a harmadikét stb. Az utolsó keret elején már nem létező adatot kér, ekkor pl. csupa nullát adhatunk meg. Mikor ennek az „utolsó utáni” keretnek az elején újra adatot kér, akkor beírhatjuk a STOP parancsot.

Nevezzük szavaknak azokat a beszéd részleteket, melyeknek kódjait EPROM-ba programoztuk! Természetesen ezek nem feltétlenül nyelvtani értelemben vett szavak, hanem olyan építőelemek, amikből értelmes üzenetek állíthatók össze. A szavak kombinációit hívjuk kifejezéseknek! Pl.: a „jó napot kíván”, „Szabó”, „Kovács”, „Tibor”, „János”, és a „beszéd szintetizátor” szavakból „Szabó Tibor jó napot kíván”, „Kovács Tibor és Szabó János jó napot kíván” valamint „a beszéd szintetizátor jó napot kíván” kifejezések egyaránt összeállíthatók.

Automatikus hívásmegvalósítás és hívásfogadás

Helyi hívás esetén a következő vonali jeleket kell megkülönböztetni [9]:

Tárcsázási (T) hang:

425 ± 25 Hz; ütemezés: folytonos
–10 ± 5 dBm

Foglaltsági hang:

425 ± 25 Hz; ütemezés: 312 ms ± 20%
jel
312 ms ± 20%
szünet

–10 ± 5 dBm

Csengetési visszhang:

425 ± 25 Hz; ütemezés: 1250 ms ± 10%
jel
3750 ms ± 10%
szünet

–10 ± 5 dBm

Csend: impulzusszerű zavarok, áthallás.

Egyetlen végződésen végzett nagyszámú tájékoztató jellegű mérésünk szerint a fenti névleges értékektől mind frekvenciában, mind szintben

jelentős eltérések találhatók. A jelalakok szintén nagyon változók. A hardver felépítésben ismeretett reedrelés vezérlés segítségével a tárcsa impulzusokkal szemben támasztott követelmények teljesíthetők. A pontos időzítést a kvarcvezérlésű órajel biztosítja. A hívásfelépítés folyamata a következő:

- egyenáramú hurok zárása,
- T-hang várása,
- tárcsázás,
- csengetési visszhang várás,
- csengetési visszhang megszűnés várás,
- üzenetbemondás,
- egyenáramú hurok bontása.

A program a fenti lépéseket végzi el egymás után. Először a KI reed-relé segítségével zárja a hurkot. Ezután a fentebb felsorolt valamennyi vonali jel megjelenhet. Itt azonban csak arra van szükség, hogy a T-hangot nagy biztonsággal kiválasszuk ezek közül. A vonali jelzőbitet 200 ms-ként mintavételezve, ha egymás után 15-ször jel jelenlétét tapasztaljuk, azt T-hangnak tekintjük. Ha 30 másodpercig nem detektálunk T-hangot, bontjuk a hurkot, majd 0,8 s múlva újra zárjuk. Ha volt T-hang, a K2 reed-relé segítségével KI nyitott állapotában beadjuk a tárcsaimpulzusokat, és a csengetési visszhangra várunk. Itt 34 ms-ként mintavételezzük a vonali jelzőbitet. Ha egy 0,8—4 s közötti hosszúságú impulzust detektálunk, azt csengetési visszhangnak tekintjük. Ha ötször 200—800 ms közti impulzust találunk, azt foglaltsági hangnak tekintjük és újra kezdjük a hívást. Az ennél rövidebb impulzusokat zajnak tekintjük. Négy másodpercnél hosszabb impulzus esetén is új hívást kezdeményezünk. Erre kerül sor akkor is ha 30 s-ig sem csengetési visszhangot sem foglaltsági hangot nem érzékelünk. Ez soknak tűnhet, azonban előfordulhat, hogy a kapcsolás lassúsága ilyen hosszú kivárást tesz szükségessé.

Az algoritmus hibája, hogy ha egy csengetési visszhang sem jön a hívott jelentkezése előtt, akkor a hívott beszédét foglaltsági hangnak érzékeli. Ez az eset azonban elég kis valószínűséggel fordul elő. Ha egy csengetési visszhangot érzékelünk, annak megszűnését kezdjük várni. Ez a legalább 0,8 s hosszú impulzusok 4,2 s-nál hosszabb ideig történő kimaradását jelenti. Sajnos, ez a megoldás jelentősen megnöveli az eszköz válaszidejét. Cserébe viszont nagyon megbízható működést érünk el. Mintegy félévi vizsgálat során a berendezés egyszer sem érzékelt tévesen a csengetési visszhang megszűnését. Ha az eszköz 30 s-ig nem érzékeli a csengetési visszhang megszűnését, egy jelzőbit beállításával tér vissza a főprogramba. A hívott jelentkezése után az eszköz bemondja a beprogramozott üzenetet, majd bontja a hurkot.

Üzenetrögzítés során a program 2 ms-ként minta vételezi a vonali jelzőbitet. Ha folyamatosan 1 s-ig jelet észlel, azt csengető jelnek tekinti. Ezután zárja a hurkot, majd kétszer megismételve bemondja az erre az esetre beprogramozott üzenetet és elindítja a magnetofont. Mintegy 10 s eltelte után a szintézerrel egy figyelmeztető hangot

adat ki, s 10 s-al ezután leállítja a magnetofont és bontja a hurkot.

A fenti rövid ismertetés is jól mutatja, hogy a mikroprocesszorhoz illesztett egyszerű hardver segítségével egy olyan meglehetősen bonyolult érzékelési algoritmust alkalmazhattunk, aminek tisztán hardver megvalósítása igen rugalmatlan, nehezen módosítható és költséges lenne.

Az érzékelő bemenetel: és a tasztatúra figyelése és kezelése

A program legfontosabb feladata az, hogy riasztás esetén a lehető leghamarabb érzékelje és továbbítsa azt. Nem riasztási feladatok végzése közben max. 1 s-ként olvassa le a szintérezkeny, prioritási sorba rendezett bemenetek állapotát. Minden bemenet-hez hozzárendelhető egy felhívandó telefonszám és egy bemondandó üzenet. Ha a bemenetek közül egyszerre több válik aktívvá, a program először a legmagasabb prioritásúhoz tartozó számot hívja fel és bemondja a megfelelő üzenetet. Ezután a következő prioritásút szolgálja ki stb. A bemenetek egyszeri leolvasásához tartozó riasztások elvégzése után újra megvizsgálja a bemeneteket és mindaddig folytatja a riasztást, amíg aktív bemenetet talál.

A tasztatúra működési elvét a hardver felépítés leírása során ismertettük. A tasztatúráról a következő utasítások adhatók:

- max hatszámjegyű telefonszám megadása,
 - az alapszókinccs egy szavának kimondása.
- A max. kb. 70 s-nyi alapszókinccs elemeit 0 és 127 közti számkódokhoz rendeljük. A számkód megadásával mondható ki az adott szó.

— a megadott telefonszám felhívása és sikeres hívás esetén adott szó vagy kifejezés bemondása. Sikertelen hívás után az eszköz „A hívás nem sikerült” üzenet kimondásával tér vissza parancsváró állapotba.

— egy kifejezés kimondása. Az alapszókinccs elemeiből is 0 és 127 közti számkódokhoz rendelt szófüzés, kifejezés definiálható és a megfelelő számkód megadásával kimondatható. A kifejezések lehetnek előre beprogramozottak és a tasztatúráról megadottak.

— új kifejezés definiálása. Az alapszókinccs elemeiből és a már rendelkezésre álló kifejezésekből új kifejezés hozható létre.

— általunk definiált kifejezés törlése.

— üzenetrögzítő állapot és üzenetrögzítéskor bemondandó üzenet megadása.

Összegzés

Az ismertetett eszköz kis alkatrészigény, alacsony ár, kis méretek és fogyasztás mellett teszi lehetővé nagyméretű, alkatrészigényes, mechanikus elemeket tartalmazó, nagy fogyasztású berendezések kiváltását. Mivel csaknem teljesen elektronikus felépítésű (elektromechanikus elemként csak reed-relét tartalmaz), megbízhatósága is jóval nagyobb. Ezeket az előnyöket a mikroprocesszoros vezérlés és a beszédszintetizátor alkalmazása biztosítja.

A pillanatnyi kiépítésben elsősorban a riasztáshoz tartozó funkciókat fejlesztettük ki, azonban

kis módosítással a berendezés például vakok számára „beszélő telefonként”, „beszélő óraként” stb. is szolgálhatna. Ezek alapján bátran állíthatjuk, hogy hazánkban is megvan a lehetőség a beszédszintézis közszükségleti alkalmazására.

Köszönetnyilvánítás

Végezetül szeretném megköszönni a MINIVOX szabadalom megalkotóinak továbbá konzulenseimnek — Podoletz Györgynek és Takács Györgynek — és a BME HEI Átvitel- és Rendszert Osztály dolgozóinak munkám során nyújtott segítségét.

IRODALOM

- [1] Voice Input Output, Electronics, April 21, 1983
- [2] Elektronikai kutatás az EGK-ban, HVG, 1984. I. 7. pp 13—15

- [3] *Rabiner L. R., Schafer, R. W.*: Digital Processing of Speech Signals, Prentice Hall, 1978. pp. 6—8
- [4] *dr. Gordos Géza*: Digitalizálás a hangtechnikában: új távlatok az ember-gép kapcsolatokban, Kép- és Hangtechnika, Vol. XXX. No. 1, 1984 febr. pp. 15—23.
- [5] *dr. Békési S., dr. Gordos G., Glaszy G., Podoletz Gy., Takács Gy.*: Eljárás formánsszintetizátor vezérlés, sére mesterséges beszéd és speciális hangjelenségek létrehozása céljából, lastkonsz: OTH 189.337 (bejelentés: 1983).
- [6] *U. Tietze, Ch. Schenk*: Analóg és digitális áramkörök, Bp. M.K. 1981. p. 632
- [7] *Gordos, Takács*: Digitális beszédfeldolgozás, Bp. M.K. 1983. pp. 140, 141
- [8] Philips Data Handbook, Integrated Circuits, Part 11. April 1983. pp. 535—546
- [9] KPM: Távbeszélőtechnikai mérések, Bp. Közdok. 1979. pp. 121—125
- [10] *dr. Gordos Géza*: Szolgáltatásközlés gépi beszédfeldolgozással Híradástechnika, XXXV. évf. (1984) 11. sz., pp. 512—518