

Kiss Gábor:

Magyar nyelvű szöveg-beszéd átalakítás az MTA
Nyelvtudományi Intézetében

Hat évvel ezelőtt az I. HLUG szimpóziumon számoltunk be az MTA Nyelvtudományi Intézetében megkezdett beszédelőállítási munkálatokról. Akkori előadásunkban a KÉPLET nevű rendszert mutattuk be, amely begépelte számokat (1-999.999.999) és számítási műveleteket mondott ki magyarul. Jelen előadásunkban a további munkálatok eredményeképpen megszületett magyar nyelvű beszédelőállító rendszert ismertetjük, amely nemcsak számokat, hanem bármilyen magyar helyesírással megadott szöveget meg tud hangosítani. A fejlesztés főbb lépései a következők voltak:

Az 1970-es évek végén az MTA Nyelvtudományi Intézetének fonetikai osztályán a beszéd akusztikumának vizsgálatára Bolla Kálmán vezetésével egy beszéd szintetizáló rendszert hoztak létre. Ezt OVE III formánsszintetizátor és PDP 11/34-es szg alkotja. A szg központi memóriája 32 Kbyte, valamint 2 floppy egységgel van ellátva.

A rendszerhez az 1980-as évek elején készült a FOPRO elnevezésű, a fonetikai kutatásokat segítő program, amelynek segítségével hatékonyan lehet interaktív módon hangokat, hangkapcsolatokat, szavakat bármilyen nyelven létrehozni úgy, hogy a felhasználó (fonetikai kutató) aprólékos munkával határozza meg a beszéd szintetizátor 15 paraméterének értékét átlagosan 20 msec-onként.

(Vö.: Kiss G.-Olaszy G.: Interaktív beszédszintetizáló rendszer számítógéppel és OVE III beszédszintetizátorral, Magyar Fonetikai Füzetek 10, Bp. 1982.) A fonetikai kutatásokban a FOPRO program használhatóságát a gyakorlat bizonyította, ezt számos tudományos publikáció mutatja.

Az ipari alkalmazások felé tekintve a FOPRO programmal határoztuk meg azon alapszavak (pl. egy, kettő, ezer, minusz, egyenlő stb.) akusztikai paramétereit, amelyeket a számítási műveleteket kimondó KÉPLET programban használtunk. (Vö.: Kiss G.-Olaszy G.: A magyar beszéd automatikus szintézisének első lépcsője, MFF 10.) A későbbiekben egy, a pontos időt hangosan jelző programot is készítettünk.

Mivel az előbb említett mindkét program kötött szótárral dolgozik, felmerült egy szótár nélküli rendszer megalkotásának is a szükségessége, amely már bármilyen magyar nyelvű szót, mondatot ki tud mondani. Ennek megalkotása során létre kellett hozni egyrészeztől egy akusztikai elemtárat, amelyből a szintetizáló rendszer felépíti a beszédet, másrészeztől meg kellett írni egy olyan programot, amely a magyar helyesírással begépett szövegből egy szabályrendszer segítségével többszörös transzformációval eljut a szintetizált beszédig. (Vö.: Kiss G.-Olaszy G.: A HUNGAROVOX magyar nyelvű, szótár nélküli, valós idejű, párbeszédes beszédszintetizáló rendszer, Információ Elektronika, 1984/2.) Az akusztikai elemtár alapváltozatát Olaszy Gábor villamosmérnök kollegám 1981-ben dolgozta ki, és fejlesztésén, bővítésén a további

években is dolgozott. A rendszer programját FORT-RAN nyelven írtuk meg, amelyet a szűkös memória kapacitás miatt overlay-technikával kellett ellátni.

A program a RUN parancs után először az akusztikai elemtár 370 elemét, az un. hangszeleteket tölti be a központi memóriába. Ezt befejezve a tilde (~) karakter kiírásával jelzi, hogy kész a meghangosítani kívánt szöveg beolvasására. Ekkor kezdheti a felhasználó a szöveg begépelését, amelynek hossza maximum 70 karakter lehet (ez kb. 5 sec-nyi beszédnek felel meg). A szöveget a mondatvégi írásjellel kell lezárni. A rendszer ezek után 4 lépésben állítja elő a mesterséges beszédet a karakter sorozatból:

1. betű-beszédhang transzformáció,
2. beszédhang-hangszelet transzformáció,
3. dallamgenerálás mondatvégi írásjel alapján,
4. a beszéd szintetizátor vezérlése.

A 4. lépés végrehajtása közben hallható a beszéd, amely objektív tesztelések eredményeként is jól érthetőnek minősíthető. A rendszert a nagy nyilvánosságnak az 1983. évi tavaszi BNV-n mutattuk be. A szöveg-beszéd előállító rendszert a Budapesti Elektroakusztikai Gyár megvásárolta és SYSTER számítógépre átdolgozva kívánja forgalmazni.