

A beszédszintetizálás fejlődése és a számítógép szerepe

Napjainkban mindjobban előtérbe kerülnek a beszéd mesterséges előállításának kérdései. A technika legfejlettebb országokban a tudományos kutatás célkitűzései között szerepelnek olyan „beszélőgépek”, gépi beszédet előállító számítógépes megoldások, amelyek ki tudják elégíteni a fejlődő technika, az ipar és a társadalom egyre növekvő igényeit. Magyarországon is folytak ilyen kísérletek az MTA Nyelvtudományi Intézetének fonetikai laboratóriumában, dr. Bolla Kálmán kandidátus irányításával. Cikksorozatunkban szeretnénk megismertetni az olvasókat a gépi beszéd előállításával, a számítógép ez irányú szerepével, az alkalmazások lehetőségeivel és a magyar beszéd számítógépes előállításával.

A mechanikus gépektől a VODER-ig

Bevezetőben szólunk néhány szót a mesterséges beszéd előállításának kezdetéről. A beszédnek, mint akusztikai produktumnak a létrejötté, előállításának fiziológiai mechanizmusa az embert már régóta izgatja, de a mesterséges beszédre eddig csak a 18. században kezdtek el foglalkozni. A mesterséges beszédhang előállítására tervezett első szerkezeteket az orosz Kratzenstein professzor (1779) és a magyar Kempelen Farkas (1791) konstruálta. Utóbbiról részletesebben Tarnóczy Tamás: „Kempelen Farkas beszélőgépe” című cikkében (1946) olvashatunk. Ezek a szerkezetek mechanikus elven működtek, az ember hangképző mechanizmusát próbálták durván utánozni. Kempelen gépét 1837-ben Wheatstone tökéletesítette (lásd az ábrát).

Kempelen beszélőgépének Wheatstone által tökéletesített változatában megtalálható volt a tüdőt utánozó fújtató (1), a hangszalagrezgést imitáló rezgőnyelvszerkezet (2), az orrot képviselő két lyuk (3), az artikulációs csatornát, vagyis a szájüreget és az ajkakat utánozó elasztikus, bőrből készült csőszerű csatorna (4), a zörejhagokat előállító nyílás (5) és egy zöngés-zöngétlen kapcsoló (6). A fújtatót működtetve, és a bőrrezonátor alakját kézzel változtatva — a nyíl irányába — különböző hangok hagyták el a szerkezetet.

A későbbiekben Bell (a telefon feltalálója) is foglalkozott a mechanikus rendszerű beszélőgép kifejlesztésével. Beszélőgépével Bell a magánhangzók és a nazális hangok mellett már egyszerűbb szavakat is meg tudott szóltatni.

A mechanikus konstrukciók a 20. század elején még tovább fejlődtek, de az igazi forradalmat a mesterséges beszéd előállításában előbb az elektronika fejlődése, majd a számítógép alkalmazása jelentette.

Folyamatos beszédet első ízben a VODER elnevezésű beszédszintetizátorral állítottak elő. A készüléket az 1939-es New York-i vilákiállításán mutatták be. A VODER a beszédhangok akusztikai szerkezetét, vagyis a spektrumanalí-

zis útján kapott átviteli jellemzőket elektromos rezgőkörök segítségével hozta létre. A rezgőköröket — zöngés hang esetén — oszcillátor, zöngétlen hang esetén zajgenerátor gerjesztette. Működtetésére klaviatúra szolgált. A képzett kezelő a rezonatórköröket, a zöngés-zöngétlen kapcsolót a mechanikus billentyűzet segítségével működtette, az oszcillátor frekvenciáját — azaz a hangmagasságot — lábpedállal változtatta. A VODER-nél a számítógép szerepét a kezelő személy töltötte be, hiszen ő tárolta agyában az adatokat, ő hajtotta végre — a számítógéphez képest hihetetlen lassúsággal — a programot, tehát a billentyűk lenyomását, a lábpedál és a zöngés-zöngétlen kapcsoló működtetését.

A számítógép szerepe

A jó minőségű mesterséges beszéd előállításához szinte nélkülözött a számítógép: műveletgyorsaságával, nagy adattároló képességével és rugalmas adatkezelési rendszerével.

Hogyan lehet a számítógép segítségével beszédet előállítani? A beszéd nem más, mint egy folyamatosan és gyorsan változó nyomáshullám, amelyet az ember beszélőszerveivel hoz létre. Ha a beszédet mikrofon segítségével elektromos jellé alakítjuk, és ebből az analóg jelből számítógéppel, megfelelő sűrűséggel mintát veszünk, a mintákat tároljuk akkor a beszéd analóg jelt digitális formává alakítottuk át. Ezekkel a digitális jellel a számítógép különféle műveleteket tud végezni. Ha mesterséges beszédet akarunk előállítani, akkor ennek a folyamatnak a fordítottjára van szükségünk. A számítógéppel olyan jelsorozatot kell előállítani, amelyet megfelelő, min. 8000 jel/s sebességgel AD konverterre vezetve a beszéd burkológörbéjéhez nagyon hasonló folyamatos elektromos hullámot kapunk.

Kezdeti megoldások

Hogyan lehet a számítógép segítségével ilyen jelsorozatot előállítani? A „beszélőgépek” fejlesztése során erre többféle megoldás is született, s napjainkban is rohamosan fejlődnek és tökéletesednek a beszélő számítógépeket működtető hardware és software megoldások.

A hatvanas évek végén a Bell-laboratórium dolgozott ki egy formánszintetizálás alapuló megoldást, amely Honeywell DDP 56 típusú számítógéppel és egy szintetizátorral valósította meg a következő rendszert: a hangszalagok rezgését (zöngés hangokhoz) egy

számláló jelsorozata, a zörejjel forrást random generátor jelei szolgáltatták, az artikulációs csatornát pedig egy digitális szintetizátorral imitálták. (Mint ismeretes, az artikulációs csatorna a beszédképzés folyamán mint fizikai üregrendszer működik. A rezonanciahelyek a zöngé felharmonikusait kiemelik. Az ilyen felerősített felhangnyalabokat formánsnak nevezzük. A formánsok értéke az artikulációs szervek pillanatnyi állásától függenően állandóan változik.)

A beszédszintetizáló rendszer alapját formánskódolt szavakból álló szótár képviselte. Ezzel a számítógépes megoldással számokat szintetizáltak, és a rendszert telefonközpontokban alkalmazták telefonszámok közlésére. A formánszintetizálás alapuló megoldás igen gazdaságosnak bizonyult, csupán 500 bit/s jeltovábbítást kellett alkalmazni, hogy a beszédhullámot jó közelítéssel megkapják, amíg a természetes hullámalak közvetlen előállításához kb. nyolcvanszor ekkora sebességre és adatmennyiségre lett volna szükség.

Egy más elven működő rendszerrel — ugyanezzel a számítógéppel — a nyomtatott szövegek megszólaltatását, az ún. szövegszintetizációt valósították meg. A rendszerben a beszédhullám előállítására az artikulációs csatorna legjellemzőbb hét fizikai paraméterének változtatását használták fel. Ebben a megoldásban az alapot a kiejtési szótár alkotja, amely a következőket tartalmazza: egy átlagos szótár szavait, a szavak fonetikus átírását, a szóhangsúly helyét, néhány elemi nyelvtani szabályt, valamint a szótövek és a toldalékok kapcsolódási szabályairól betáplált néhány információt. A program ezen adatok alapján vezérli a dinamikus artikulációs modellt, amelynek kimenő jelei szolgálnak az előbbi szintetizátor vezérlésére. Ennek a rendszernek a határossága, ha az eredeti beszédhullám előállításához viszonyítjuk, kb. ezerszer jobb.

A beszédszintetizálás más elvei is ismertek, mint pl. az optikai leolvasóval működő számítógépes szintetizálás, amelyben a beszédanyagot jellemző formánsmeneteket, zajgócsokat, azok intenzitását és időstruktúráját a gép előre megrajzolt ábrák alapján állítja elő és szóltatja meg.

Legmodernebb a hullámalak-előállítás

Az egyik legmodernebb eljárás nem a beszéd akusztikai szerkezetét vagy az artikulációs folyamat fizikai paramétereit használja fel a beszéd mesterséges előállítására, hanem közvetlenül a beszédhullám alakját állítja elő matematikai módszer segítségével. Ezzel a módszerrel a számítógép a természetes beszédet szinte a megtevésztésig képes utánozni. A rendszer határossága is elég jó, mivel működtetéséhez kb. harmincszor kevesebb adat szükséges, mint a természetes beszédhullám közvetlen, digitális úton történő tárolásához és visszajátszásához.

Az előbbiekből — a röviden bemutatott négy eljárásból — látható, hogy milyen széles körű a számítógép felhasználhatósága a mesterséges beszéd létrehozása terén is.

Következő cikkünkben a mesterséges beszéd szükségességéről, alkalmazásának lehetőségeiről és az MTA Nyelvtudományi Intézet fonetikai laboratóriumában folyó beszédszintetizálási kutatásokról számolunk be.

KISS GÁBOR
OLASZY GÁBOR

Üzemeltetés

Szinkronizációs hiba elhárítása az ESZ 1022-es gép megszakítási rendszerében

Intézetünk, az Építészgazdasági és Szervezési Intézet az építésügyi ágazaton belül regionális számítógép-hálózatot épített ki és üzemeltet. Jelenleg a hálózat ESZR tagjai a következők: 4 db ESZ 1020 B, 3 db ESZ 1022 B és 1 db ESZ 1040.

A gépek műszaki állapota jó, amit a műszaki kihasználási tényező 1979. évi adatai bizonyítanak:

ESZ 1020 B	88,2%	(átlag)
ESZ 1022 B	87,9%	(átlag)
ESZ 1040	88,8%	

(Műszaki kihasználási tényező: az éves bekapcsolott gépidő aránya az ugyanezen üzemeltetési periódus alatt karbantartással, javítással és tévműködések elhárításával töltött összes időhöz.)

A még biztonságosabb üzemeltetést gátolja gyakori „rendszerlemeredezés” vagy „kemény várakozás” (hard-wait), amely véletlenszerűen, de jellemzően a terhelési csúcsokban jelentkezik. A felhasználónak figyelembe kell vennie, hogy a job futási idejével arányosan megnő a „hard-wait” veszélye. Másik kellemetlen következmény, hogy a terhelési csúcsok egyben feldolgozási csúcsokat is jelentenek.

A hibakijelzés nélküli leállások a környezet részletes analízisét teszik szükségessé. A továbbiakban egy ilyen konkrét hibajelenség feltárását és elhárítását ismertetjük.

A hiba leírása

Az üzemeltetők jelezték, hogy az ESZ 1022-es gépeken, multiprogramos környezetben, az adathibás mágnesszalag olvastatása, központi egység-gépből okoz. A vizsgálatok, a javítás során mintafeladat-ként egy kritikusságnak ítélt példát használtunk: az egyik partícióban PL/I fordítást, a másikban adathibás mágnesszalag-másolást kíséreltünk meg. Monoprogramos üzemmódban a fordítás hibátlan, a másolás pedig a követelménynek megfelelően, az olvasási hiba kijelzésével futott le. Multiprogramos környezetben kétféle hibát tapasztaltunk:

- gépi stop (TO) és az operatív memória adatregiszterének paritáshibája (RDOP), de úgy, hogy statikusan a regiszterek paritása helyes;
- zárt mikroutasítás-lánc, amelyből csak rendszertörléssel lehet kilépni.

A kísérletet DOS+POWER operációs rendszer környezetében végeztük, de a későbbiekben, kísérleteink során megállapítottuk, hogy a hiba az operációs rendszertől független.

A mágnesszalagos alrendszeren túl, a hiba a mágneslemez alrendszeren is jelentkezett, ahol a probléma feltárását a ROBOTRON által az ESZ 1040-eshez szállított műszaki teszt (DGSS) segítette — ugyanis az adathordozó felcserélhetőségét vizsgáló szelekcióval ugyanezt a hibát sikerült előállítani.

A hibakörnyezet

A két hibajelenség közül a mágnesszalagos a b) pontban leírta vezethető vissza.

A második szelektorsatornán (itt működnek a mágnesszalagegységek) az egység egy ún. „végső állapot-byte”-ot ad le. Az interface adatszerűen megjelenő „végső állapot-byte”, „csatorna befejezte”, „egység befejezte” jelentésű bitkombinációt tartalmaz.

Az interface-azonosítók közül a „kiválasztás a csatornától”, a „periféria működik” és

a „vezérlés a perifériától” (VBR-K, RAB-A és UPR-A) logikai szintje magas, amelyből az következik, hogy olyan I/O megszakítás volt, amelyre a szelektorsatornát kiszolgáló mikroprogramnak (SERVS) kellett volna elindulni. A SERVS mikroprogram fix kezdőcímenek betöltése akkor történhet meg, ha a vezérmű a megszakítási igényt elfogadta és engedélyezte a kezdőcím betöltését a mikroprogramtár címregiszterébe. A közös csatornában levő mikroprogramos megszakításblokkoló regiszter negyedik bitje magas logikai szintű, és jelzi, hogy a SERVS mikroprogram végrehajtása folyik. Ez az állapot blokkolja a további I/O mikroprogramos megszakítást egészen addig, míg véget ér, vagy olyan állapotba kerül, amikor megszakíthatóvá válik egy magasabb prioritású igény számára. A hibás mikroprogram-ciklusában SERVS mikroprogram nem szerepel, annak ellenére, hogy a csatornavezérlés állapota erre utal. A mikroprogram ugrási feltételei a csatornában alakulnak ki, mivel az ezt meghatározó flip-flop is magas logikai szintű. Függetlenül attól, hogy processzor- vagy csatorna-mikroprogramok futnak, az ugrási feltétel a csatorna-állapot függvénye. Ennek következtében lehetséges, hogy az operatív tár adatregisztere paritáshiba lesz, ha a rosszul kialakuló feltétel miatt az RDOP regisztere integrálja a megszakítási igények regiszterének (GRP) tartalmát. Az adatregiszterben az egységek párosságától függően — mely ilyenkor véletlenszerű — paritáshiba jelentkezik.

Összegezve: a hibajelenség oka az, hogy a csatornavezérlés állapota a SERVS mikroprogram általi működtetésre utal, de az még nem indult el.

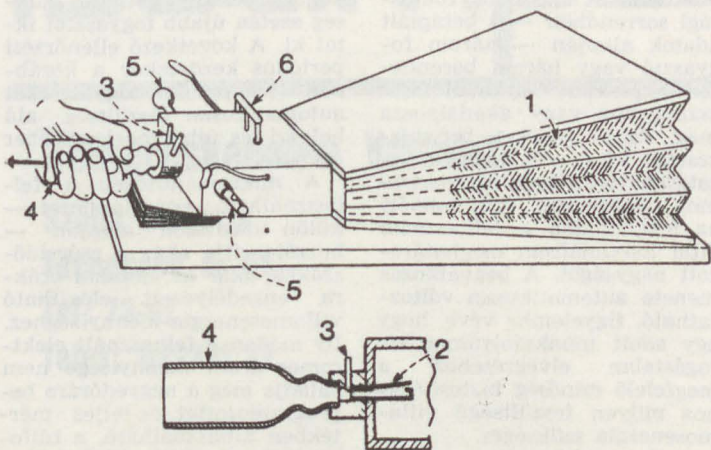
A hibás állapot kétféleképpen alakulhat ki: a címbetöltési intervallumban a mikroprogramtár címregiszterét valamely más információ felülírta, avagy a fix cím be sem töltődött.

A mérés egy olyan segédáramkörrel történt, amely a címbetöltés engedélyezésének letárolásával, a betöltési idő alatt letiltotta a címregiszter átírását. Ennek eredményeként viszont a címregiszter üresen maradt, tehát a fix címbetöltés engedélyezési ideje alatt az információ, azaz a SERVS mikroprogram kezdőcíme nem érkezett meg. A jelenség szinkronizációs hibára utal. A hiba az eredetivel eltérő szinkronimpulzusok használatával és az adott áramkör logikai átalakításával kiküszöbölhető.

Üzemeltetési tapasztalatok

Intézetünk bolgár és szovjet gyártmányú ESZ 1022 B gépeknél ugyanezt a hibajelzést tapasztaltuk. A felhasználók a számítógépeken a jelzett hibát az átalakítás után már nem észlelték. Olyan észrevétel sem érkezett, amiből arra lehetne következtetni, hogy más körülmények között a módosított áramkör működése kritikussá vált volna. A bolgár gyártó vállalatot a feltárt hibáról értesítettük, és átadtuk a módosítás műszaki dokumentációját. A hazai ESZ 1022-es számítógépeket üzemeltetők számára — az átalakítás részleteiről — készségesen adunk felvilágosítást.

RUTTKAY GYÖRGY



Kempelen beszélőgépének Wheatstone által tökéletesített változata