ITT AZ ÖN SZOFTVERE BESZÉL...

Beszélő programok fejlesztése egyszerűen

Ki ne hallotta volna a mobiltelefonjában azt a „kedves, szőke, kék szemű női hangot”, aki közli velünk számlaegyenlegünket, esetleg útbaigazít a szolgáltató menürendszerében.Beszédét általában kellemesnek ítéljük meg, néha azonban kicsit gépiessé válik, esetleg kattan, vagy „ugrik” egyet… Miért van ez? Aki végigolvassa a sorozatunkat, meg fogja érteni. Sőt, arra is kap útmutatást, hogy a fenti hibábat miként kerülheti el. Amikor a gondolkodó ház elterjedése napi kérdéssé érett, a háztartási berendezések egyre intelligensebbek lesznek, a háztartási robotok pedig egyre többször tünnek fel -elsősorban a Japán újságcikkekben, de lassan a boltok polcain is-, akkor egyre égetőbbé válik, hogy megtanuljunk néhány apró, szakmai fogást, melyekkel mi is tudunk beszélő programokat készíteni. Ha ehhez hozzátesszük az LPT-s cikksorozatban megismerteket, illetve a webprogramozással foglalkozó cikkeket, akkor határtalan lehetőségek nyílnak meg előttünk; Számítógépünkkel a kommunikációnk sokkal emberibbé, közvetlenebbé válik, ugyanakkor a gépet beköltöztethetjük lakásunk, autónk vezérlőegységébe, stb. S így egy meglepően kellemes, intelligens személyiség benyomását keltő segítőtársunk lehet az a PC, melyet korábban sokan az íróasztalon tudtak csak elképzelni. Ebben a cikksorozatban röviden összefoglalom a beszédelőállítási módszereket, majd ennek egy fajtáját részletesebben ismertetem.

Kis történelem:

Talán meglepőnek hangzik, azonban a beszédelőállítás korántsem mai találmány, sok forrásban lehet olvasni éneklő, beszélő automaták leírásairól már a középkorban is. Vélhetőleg korábban szintén kísérletezhettek ezzel. Ellenben először, -építési leírással dokumentáltan- Kempelen Farkas készített beszélő gépet, mely mechanikus úton képes volt az emberi hangokat előállítani. Természetesen nem volt egyszerű dolog „beszéltetni” ezt a berendezést! Szinte zongoraművész tehetség kellett ahhoz, hogy valami érthető, emberi hangot kicsiholjon belőle. Azonban a gép ennek ellenére mégiscsak beszélt! Bizonyítva ezzel, nem kizárólag az élőlények privilégiuma a beszéd előállítása, hanem technikai úton is lehetséges ez. Fontos megmlíteni, hogy nem hangrögzítőt alkotott, hanem -mai szóval- „hangszintetizátort”, ami az egyszerű rögzítésnél sokkal előremutatóbb dolog!

A következő lépést Edison találmánya, a fonográf, majd ennek tökéletesített változata, a gramofon jelentette. Edison évekig táviratozással kereste a kenyerét, s megfigyelte, hogy a távíró tűje leérkezéskor a papíron sercegő hangot ad, mert a folyamatosan elhaladó papírszalag barázdáit követve rezgésbe jön. Arra gondolt, mi lenne, ha zenével, emberi hanggal próbálná meg a szalagra felvinni ezeket a barázdákat. Ugyanis ezt visszahallgatva, talán a távírótű megszólalna… Sztaniollal bevont fahenger lett a hordozó, ami egy kézzel tekert menetes orsón forgott, s a tű egy fémlemezkét rezgetett, illetve annak a rezgéseit karcolta a lemezbe. Jellemző, hogy az első szöveg, amit visszajátszottak a bemutatón a szerkezettel egy gyerekdal pár sora: „Mérinek kicsiny barikája volt…” Természetesen a hangrögzítés forradalma újabb távlatokat nyitott a beszélő gépek számára is. A telefon elterjedésével híreket, fontos eseményeket, gyerekeknek szóló esti meséket mondott a telefonközpont. Azonban számunkra igazán érdekes a pontos időt bemondó automata volt!

Eleinte fonográfhengerekre, majd gramofonlemezekre, később pedig egy optikai tárcsára, -fénnyel-, (mint a filmeknél a „fényhang”) vették fel a szavakat. Egy mechanikus óra tengelye forgatta a tárcsákat, az időnek megfelelelően. Amikor be kellett mondani az időt, akkor a tárcsák állásának megfelelően egyik tárcsát a másik után bejátszották. Ez az első olyan gépi beszéd volt, amikor külső eseménytől (pontosidőtől) függően kellett változó szöveget bemondani.

A 30-as években ezek a berendezések gombamódon elszaporodtak, divattá vált a „házirobot”, illetve a kiállításokon feltűntek a mágneses hangrögzítőkkel felszerelt beszélő robotok, idegenvezetők. Ezekben egy, vagy több magnetofon állította elő a beszédet. Bár hatalmas szerkezetek voltak, de itt már külső eseménytől (hőmérséklet, idő, szenzorok jelei, stb…) függő, emberi beszéd szólalt meg. A technika fejlődésével ezek a gépek is egyre fejlettebbé váltak. Azonban a fenti alaptípusok, vagyis a szintetizált szöveg, illetve a tárolt beszéd variálásával működő módszer lényegében a két, ma is alapvetően elterjedt beszédgenerálási eljárás. Nézzük meg tehát ezek alapvető sajátosságait:

Szintézisen alapuló módszer

Az emberi beszéd egy nagyon összetett függvénnyel leírható jelsorozat. Mivel egyénenként változó az egyes hangok ejtése, -magassága, a felharmonikusok összetétele, a beszédtempó, a remegés (vagyis „intonáció”), stb,- szintetikus úton valakinek a hangját pontosan leutánozni szinte lehetetlen. Az is hatalmas erőfeszítést igénylő feladat, ha egy semleges, ellenben már nem gépies hangzású, monoton beszédet szeretnénk így kapni. A szintetizált beszédnél egy bemondóval felolvastatnak egy előre meghatározott, kellően változatos szöveget. Ezután a szövegen Fourier transzformációt hajtanak végre. Ekkor megkapják a frekvencia-idő diagramját. (Aki nem tudja miről van szó, gondoljon a népszerű WINAMP program „spectrum bar” kijelzőjére, csak sokkal több oszloppal.) Ezután elkezdik elemezni, hogy a kimondott szöveg ismeretében az egyes hangok milyen frekvencia és amplitúdóösszetevőkkel rendelkeznek. Összegyűjtik az egyformákat, különválasztják az eltérőket. A hangok azonban nem egyszerűen hangonként bírnak jellemzőkkel. Másképp szól, ha másik hanghoz kapcsolva ejtünk ki egy mássalhangzót. (pl. „ED és ÖD kapcsolat) Vannak felfutó, lefutó ívek, ívdarabok, réshangokhoz, zörejhangokhoz kapcsolódó tagok, stb. Ezeket mind, mind el kell különíteni. Végül létrejön egy ún. „hangszelettár”, melyből építkezve már tetszőleges szövegek építhetők fel. A hangszelettárnak jobb esetben is többszáz, de gyakran több ezer önálló eleme lesz. Ezeket az elemeket részletesen megvizsgálják. Feljegyzik az erősség változásaikat, vagyis az „amplitúdó- burkológörbét”, illetve az egyes frekvenciaösszetevők arányait. A kapott adatokat táblázatos formában eltárolják. Így jön létre a „hangszelettár-paramétertáblázat”. A további eljárás már a technológiától függ. Régebben célhardvert építettek programozható frekvenciájú oszcillátorból, (zönge generátor) zajgenerátorból, (zörejgenerátor), illetve programozható frekvenciájú szűrőkből. /Általában a 4. harmonikusig szintetizálják a hangot, (ekkora pontosság a gyakorlatban elegendőnek bizonyult) ehhez tehát 4 szűrő kell. / A zajt, illetve az alapfrekvenciát, s a szűrőket egy keverőre vezették, melyet szintén programozni lehetett. Ennek kimenetén áll elő a kívánt beszédhang. A megoldás hátránya, hogy célhardvert igényel, előnye, hogy a számítógép teljesítménye viszonylag alacsony lehet. Gyakran a beszédszintetizátort egyetlen integrált áramkörben valósították meg. Az IC néha tartalmazta a hangszelettárszintéziséhez szükséges adattáblázatot, néha pedig helyette a gép adta a paramétereket is. Előző esetben egyszerűbb volt a szoftver, utóbbiban tetszőlegesen változtatható a hang minden paramétere. Az idő múlásával azonban megjelentek előbb a nagyobb teljesítményű asztali számítógépek, majd a gyorsabb mikrovezérlők, illetve a DSP-k (digitális jelfeldolgozó processzorok). Ezeknél már szoftverből a teljes szintetizátor hardvere helyettesíthető, méghozzá valós időben! Vagyis a szintetizált beszéd előállítása teljesen folyamatosan, szoftverből történik. A számítógép a digitális jeleket végül egy D/A átalakítóra küldi, s ezzel előállt a hang. Az ilyen szoftver már jóval bonyolultabb természetesen, hiszen jelalak szinten mindent a programnak kell kiszámolni. Azonban a programot csak egyszer kell megírni, így végül a termék nagyon olcsó lehet. Fontos előnye a rendszernek, hogy tetszőleges szavak kimondására alkalmas, vagyis ma még nem létező, új szavakat is képes lesz a szoftver a jövőben kimondani, ha szöveges formában (pl. TXT) elé tesszük. Vannak direkt olyan rendszerek, melyek úgynevezetett „text to speech”, (vagyis szöveg-beszéd) átalakítók. Szokás ezeket a rendszereket emiatt „kötetlen szótáras beszélő rendszernek” is nevezni. Egyébként a programok a nyelvtani szabályok, hasonulások, kiejtési kivételek, stb. ismeretében meglepően kevés kiejtési hibával tudnak beszélni, folyamatosan. Vannak vakok, illetve gyengénlátók számára felolvasó szoftverek, melyek jó ideje kifogástalanul üzemelnek. Számtalan előnye mellett meg kell azonban említeni a módszer hátrányát is: sajnos a kapott beszéd ma még általában monoton, gépies hangzású. Habár egyre jobb és jobb szoftverek látnak napvilágot… Bátran mondhatjuk tehát, hogy Kempelen Farkas egykori találmánya a mai napig is kihat ránk, mitöbb megjósolható, hogy a távoli jövőben szinte kizárólag ez fog megmaradni.

Mintavételezésen alapuló módszer (sampler)

Ennél a módszernél a feladatot alapvetően más oldalról közelítjük meg. Egy beszélővel a létező összes mondatot, vagy szót felmondatjuk, amit szükségesnek tartunk s a megfelelő minőségben rögzítjük. Ezután az egyes szavakat, mondatokat, mondattöredékeket variálva, a megfelelő sorrendben történő szakaszok lejátszásával állítjuk elő a beszédet. Fontos észerevenni, hogy ennél a módszernél csak azt tudjuk kimondatni a géppel, amit a beszélő is kimondott egyszer. Vagyis ezek szókészlete kötött. Ezért ezeket a rendszereket szokás „kötött szótáras beszélő rendszernek” is nevezni. A kapott hangminőség általában kitűnő, hiszen a beszélő saját hangján szólal meg a rendszer. Olyan helyeken, ahol viszonylag keveset kell beszélni, azonban igen fontos az érthetőség, szinte kizárólag ezeket használják. Sok program alternatív szókészletet tartalmaz, s kiválaszthatjuk a számunkra legkellemesebben beszélő hangját. A mobiltelefonok, az ébresztőórák, a felvonók, ipari berendezések beszélő rendszerei ma ilyen módon szólnak hozzánk. Amellett, hogy a jövőt vélhetően a szintetizált beszédelőállítás fogja jelenteni, a jelen egyértelműen a mintavételezésé.
Mi is az ilyen programok készítését fogjuk áttekinteni, mert megdöbbentően kevés munkával lehet a mai eszközökkel igen látványos eredményeket elérni.

Hibrid beszélő rendszerek

Ezeknél a fenti két módszer kombinációját alkalmazzák. Igen sokfajta megoldás elképzelhető, így csak címszavakban említünk meg néhányat. A tárolt beszédből kivett darabok adják a hangszelettárat, melyből építkezünk. Ekkor a beszélő hangján fog megszólalni a rendszer. Azon változtatni nem tudunk, csakis másik hangminta alkalmazásával. Ellenben ugyanolyan kötetlen szótáras rendszert tudunk felépíteni belőle. Természetesen a beszélő eredeti hangját nem fogja megközelíteni a kapott hangzás, annál jóval „robotosabb” lesz.

A másik érdekes kombináció, amikor szintetikus úton állítunk elő beszédet, de csak pár szót. Ilyenkor lehetőség van az igen finom kidolgozásra, ami szinte csilingelően szép hangzást adhat. Mintha egy földöntúli lény beszélne kristálytiszta hangon az emberhez. Ilyenkor -bár szintézist végzünk-, mégis kötöttszótáras rendszerünk van. A módszert csak nagyon ritkán alkalmazzák, (pl. filmekben) mert meglehetősen gazdaságtalan eljárás. Gyakorlatban olcsóbb egy jóhangú narrátort megfizetni és a stúdióban elektronikusan „kiszinezni” hangját.
Ezzel a bevezetőnk végére értünk. A sorozat következő részeiben megismerkedünk egy módszerrel, melynek segítségével házilagosan is elfogadható minőségű hangokat leszünk képesek előállítani. Annyit már most is elárulhatunk, hogy számtalan szakmai fogást kell bevetni a kívánt eredmény eléréséhez. Megtanuljuk, hogyan lehet megtervezni egy beszélő szoftver szókincsét, összeállítani a szavakat, helyesen beolvasni, normalizálni, majd megalkotni a hang-adatbankot.
Végül pedig egy gyakorlati példát mutatunk be Visual Basic nyelven írt beszélő órára. A program messze nem tökéletes, mert viszonylag gyorsan kellett összedobnom VB vizsgára. Ellenben nagyon jól lehet belőle tanulni, mert szépen látszanak, hallatszanak a helyes, illetve helytelen megoldások hatásai is.

Kis Norbert norbimagan@freemail.hu