J. D. Markel—A. H. Gray J. R.:

Linear Prediction of Speech (Communication and Cybernetics Vol 12) Springer Verlag, Berlin—Heidelberg—New York 1976. XII+288 lap, 129 ábra, ára kötve 73 DM

Nem túlzás azt állítani, hogy a modern technikai törekvések között csak az űr meghódítása hasonlítható mind bonyolultságában, mind sokoldalúságában ahhoz a feladathoz, amit röviden a beszéd gépi megértése címmel foglalhatunk össze. Az összehasonlítás egyetlen gyönge pontja, hogy míg az űr meghódítása technikailag jól megoldhatónak mutatkozik, az utóbb említett feladat nehézségei rendkívüliek. Ez azért is meglepő, mert még néhányszor öt évvel ezelőtt is voltak olyan hangok a tudományos világban, hogy közel járunk a megoldáshoz. Ma azonban nem vagyunk ennyire optimisták. S ehhez a fölismeréshez nagyban hozzásegítettek a korszerű számítógéptechnikai módszerekkel végzett beszédelemzési kísérletek eredményei is.

Az előttünk fekvő könyv ezeknek a korszerű módszereknek szinte első összefoglalása. Címét talán leghelyesebben: „*A beszédjelek lineáris megfejtése*" alakban fordíthatnánk le. Ezúttal ugyanis az angol kissé pongyola. Hiszen a könyv tartalma világosan utal arra, hogy a feldolgozás a beszéd akusztikailag megfogható képével indítható el, továbbá, hogy egyelőre csak az állandósultnak tekinthető hangelemekre ad többékevésbé biztos ítéleti lehetőséget, s végül, hogy a nyert adatokból a visszaalakítás ellenőrzése még mindig csak emberi úton oldható meg.

Két dolgot kell tisztáznunk. A „lineáris predikció" Wiener által 1949-ben bevezetett módszere vagy állandósult (esetleg lassan változó), vagy pedig teljesen rendezetlen folyamatok adatainak megoldására alkalmas. A beszédfolyamat vizsgálatára 1967-ben kezdték alkalmazni, s az első használható eredményeket 1970/71-ben közölték. A beszéd folyamata azonban sem az állandósult, sem a rendezetlen folyamatok feltételeit nem elégíti ki, illetőleg csak nagyon kis részben. A második tisztázandó pont, hogy a predikció esetünkben nem „előrejelzés", hanem legföljebb a beszéddel közel azonos idejűséget biztosítja. Ez a „közel azonos idejűség mintegy 150—250 ms késést jelent, ami természetesen nem lebecsülendő előny a korábbi módszerek időigényességével szemben, a ez az előny a számítógépek nagy sebességének köszönhető.

Végül is, anélkül hogy a közös címmel összefoglalt sokfajta eljárás egyikét is részleteznénk, arra kívánunk rámutatni, hogy mit is oldanak meg a lineáris megfejtő eljárások a sok megoldandó feladat közül. A beszéd akusztikai megjelenésének keletkezését G. Fant. 1960 óta elfogadott modelljével így írjuk le: *1.)* hangforrás (zönge, vagy zörejhang vagy a kettő együttesen), *2.)* egy változó keresztmetszetű, de mindig kb. egyforma hosszú (17 cm) cső módosító hatása, *3.)* amely részben visszahat magára a hangforrásra is, *4.)* a szájnyílás és környezete sugárzása a térbe. A négyféle függvény vagy paramétersor együttesen határozza meg a beszéd hangfolyamatának akusztikai megjelenését.

Ennek az akusztikai jelsornak az elemzése képezi ez idő szerint az egyetlen lehetőséget a beszéd lényegének megismerésére. Az első nehézség a négyféle folyamat adatainak szétválasztása, méghozzá a megjelenéssel azonos időben. Ehhez a feladathoz segít hozzá — elég nagy lépéssel — a számítógéptechnika s ezen belül a lineáris megfejtés módszereinek bekapcsolása. A módszerek közös jellemzője, hogy a jelsor néhány előzetes adatának lineáris kombinációjával a gép kiszámítja a valószínű következő adatot, majd ezt összehasonlítja a valóságban bekövetkező adattal. Ezután nemcsak a hozzáigazítást végzi el, hanem megjegyzi a saját hibáját is, és ebből a saját hibarendszerből megadja külön-külön, de egyidejűleg az *1.)*, *2.)*, és *4.)* függvényeket (a maradék visszahatást elhanyagolja).

A módszerek kb. 0,1 ms mintavételi időközzel dolgoznak, és a számítási eljárást általában 10—20 ms elteltével elölről kezdik. Ez az idő férfihangon ejtett magánhangzó 1—2 periódusát jelenti, s általában a beszédhangok időtartamának harmadát-negyedót teszi ki. Így kívánják a lassúbb változási folyamatokat is figyelembe venni. Nem jó az eljárás a p-*t*-k jellegű hangok megismerésére és nem tökéletes a gyors átmenetek tisztázására sem. Pl. az angol „*how are you?*" szövegben az *á*-a-o-u, majd az u-*á* átmenet 10—30 ms alatt zajlik le, vagyis a formáns áthelyeződés néhol 40 Hz/ms sebességű. Az eredmények pontossága amellett bizonyos számítási ciklusok (előrejelzési együtthatók) számától függ. Ha ezen együtthatók száma 10—12, az előrejelzések hibája hangos beszédre 16—18%, suttogottra 45—50%. Mindez az eredmény kb. 200 000 bit/s számítási sebességet követel meg, vagyis elég nagy számítógépet igényel.

Mi az akkor, amit nem old meg a módszer alkalmazása? Ném oldja meg az átmenetek követését az egyes beszédhangok között, nem oldja meg a folyamatos jelsor kvantumozását (egyszerű jelrendszerbe való átírását), nem képes megoldani az agyi rövid idejű emlékezés folyamatát és nem oldja meg az agy óriási tárolási és visszakorrigálási lehetőségeinek gépi megvalósítását. Az agy megoldási sebessége az előbb említett értéknek még mintegy százezerszerese.

Visszatérve a könyv tárgyára: minden negatívum ellenére is *a jövő valószínű útját tartalmazza.* A szerzők nagy gondot fordítottak arra, hogy a látszólag teljesen különböző alapgondolatú megfejtő módszereket szerves egységű rendszerré formálják. Fölényesen kezelik a matematikai módszereket (a könyv lényegében inkább matematikai-információelméleti, mint nyelvészeti-fonetikai) és látszik, hogy tökéletesen otthonosak a számítógép-technika szellemi (programozási) részében.

Ma azt mondjuk, hogy nehéz és idegen a téma, a könyv tartalma mégis nagyon messzire világít. Talán 20—30 évvel „előrejelez", de — vagy éppen ezért — fizikusnak, nyelvésznek, híradástechnikusnak és programozónak már most egyaránt fel kell figyelnie rá, mert csak együttes fölkészülésükkel és összefogott munkájukkal lesz a feladat a 2000-es évek elején megoldható.

Tarnóczy Tamás