

Hangátmenetek a beszédfelismerésben

Sejtes Györgyi – Zsigri Gyula

Szegedi Tudományegyetem
sejtes@hung.u-szeged.hu, zsigri@hung.u-szeged.hu

Abstract. A beszédhangok többsége jól felismerhető a nekik tulajdonított szakasz közepéről, de ez nem mondható el a zárhangokról. Az alábbiakban a zárhangok kétféle szegmentálását hasonlíttjuk össze. Az egyik megőrzi a felismerési mechanizmus egységességét, de az anyanyelvi beszélők intuícójától eltérően a zárhangok zár részét és felpattanását két különböző szegmentumként elemzi. A másik bonyolultabb felismerőmechanizmust igényel, de jobban modellálja a beszélők intuícóját, és a fel nem pattanó zárhangok felismerését is lehetővé teszi.

A beszédfelismerő programok úgy jutnak el oda, hogy a bejövő hanghullámokat képesek legyenek beszédhangok sorozataként elemezni, hogy előtte rengeteg olyan digitalizált hangfelvételt bocsátunk a rendelkezésükre, amelyekben emberek által beiktatott határjelölők osztják a hangfolyamot diszkrét szakaszokra, és az így kijelölt szakaszok mindegyikéhez egy-egy fonetikus jel van hozzárendelve. A határjelölők a hangfolyam természetéből következően csak körülbelüliek lehetnek. A beszédhangok nem úgy követik egymást, hogy egyszer csak véget ér az egyik, és csak azután kezdődik a másik, hanem úgy, hogy az egyik még tart, amikor a másik már elkezdődik. A határjelölők nem is arra valók, hogy pontosan megadjuk, hogy melyik beszédhang mettől meddig tart, hanem arra, hogy tudassuk a programmal, hogy hány beszédhangot kell felismernie, és hogy melyiknek melyik szakaszon belül keresse az azonosító jegyeit. A magánhangzók és a folyamatosan hangoztatható mássalhangzók széleit a program levágja, és a középső rész alapján próbálja kiszűrni, hogy melyik beszédhangot miről lehet felismerni. Ez a módszer azonban a zárhangokkal már nem működik.

Papp István (1966: 86), sok más szerzőhöz hasonlóan, a zárhangok képzésének három egymás utáni mozzanatát különbözteti meg:

- a) a záralkotást (implosio): kezdő mozzanat
- b) a zár tartamát (occlusio): középső mozzanat
- c) a zár megszüntetését (explosio): befejező mozzanat

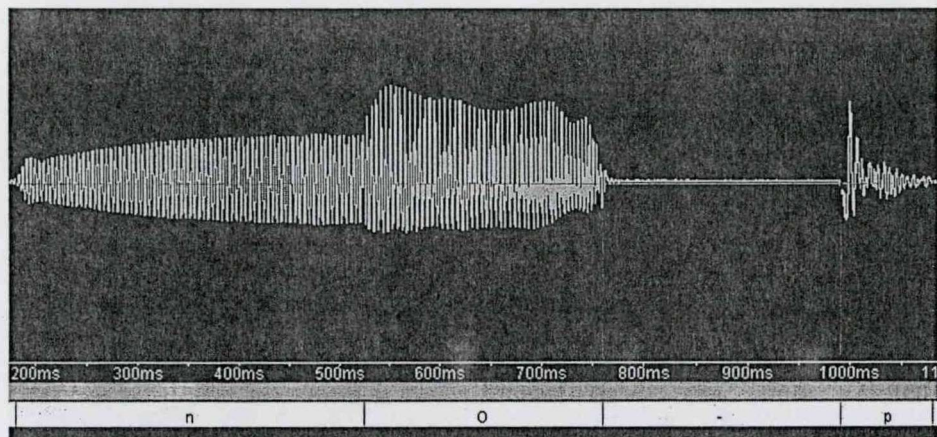
A zöngétlen zárhangok középső mozzanata teljesen néma, a zöngés zárhangok középső mozzanatában pedig csak gyenge zöngé hallatszik, amelyből nem lehet kideríteni, hogy a zárhangot hol képezzük. Ha a beszédfelismerő program a zárhangoknak is levágná a széleit, és csak a közepüket tartaná meg, akkor zöngétlen zárhang esetén egy néma szakaszból kellene kitalálnia, hogy a beszélő [p]-t, [t]-t vagy [k]-t mondott, vagy egyszerűen csak szünetet tartott, zöngés zárhang esetén pedig egy

olyan szakaszból kellene képzési helyre is vonatkozó információt kinyernie, amelyben ilyen információ nincs. Nem tudná tehát megkülönböztetni a két ajakkal képzett [b]-t, a fogmedernél képzett [d]-től vagy a szápadlás közepén képzett [g]-től.

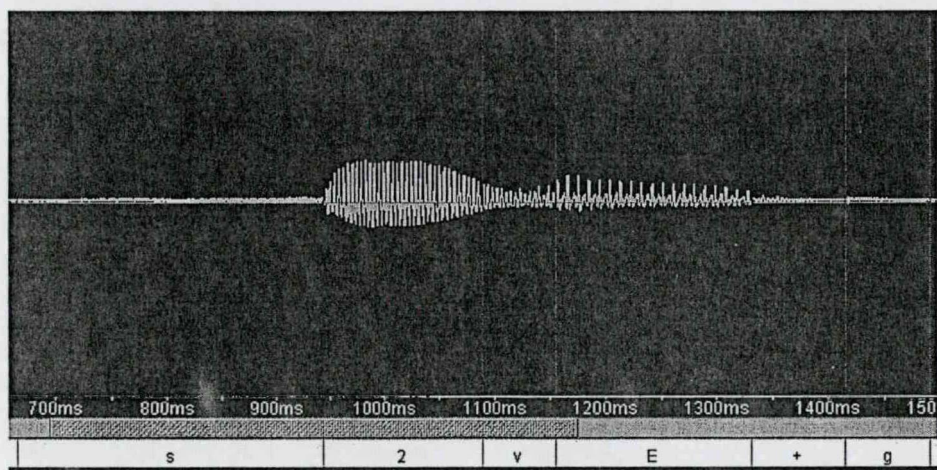
A zárhangok képzési helyét legkönnyebben a zárt követő felpattanó zörejből azonosíthatjuk, de ha ez például szó végén vagy szótag végén elmarad, akkor a záralkotásból, vagyis a megelőző magánhangzó és a zár közötti átmeneti szakaszból is kinyerhető ez az információ. Az egyetemi tankönyvek ez utóbbi lehetőséget nem említik. Papp István (1966: 86) szerint „...a kezdő és a középső mozzanat rendesen néma, tehát hangjelenség nincs jelen, a befejező mozzanat ellenben hangjelenséggel jár együtt.” R. Molnár Emma és Kassai Ilona is ugyanezt tanítja: „A kiáramló hangképző levegőáram útját a szájüregben képzett zár állítja el, a beszédhang csak a zár felpattanásakor hallható pillanatnyi ideig...” (R. Molnár 1989: 43) – „A zárfeloldás nyomában keletkező hangjelenség a zárhang.” (Kassai 1994: 617, 1998: 112, 2003: 531).

A záralkotás figyelmen kívül hagyása a magyarban ritkán okoz gondot, mert a magyar zárhangok az esetek túlnyomó többségében felpattannak. De nem mindig. Zárhang + azonos képzési helyű orrhang találkozásakor, például *népmese, kötni, sa[ty]nya*, csak akkor pattan fel a zárhang zárja, ha a zárhang és a rá következő orrhang között szünetet tartunk. Ritkábban ugyan, de elmaradhat a zár felpattanása szó végén vagy egyéb zárhang + mássalhangzó kapcsolatban is. Az, hogy a magyarban is vannak fel nem pattanó zárhangok, csak a közkézen forgó egyetemi tankönyvekből maradt ki, Vértes O. András ezt többször is említi: „Bizonyos felfogás szerint a zárhangok appercipiálásakor a hallgató a legjellemzőbb mozzanatnak a hirtelen megnyitás következtében támadó exploziós zörejt tartja. Ez a megállapítás azonban semmiképp sem általánosítható, hiszen a zárhangok egy részének zára nem nyílik fel: e zárhangok jellemző mozzanata tehát a záralkotás.” (1952: 32) – „A zár felpattanása néha elmarad, például [b]+[m] kapcsolatban: *rab madár*, vagyis nem [b]-t, hanem [b-]-t ejtünk.” (1982: 158).

Ha megelégszünk egy olyan beszédfelismerő programmal, amely az esetek túlnyomó többségében jól felismeri a zárhangokat, akkor választhatjuk a könnyebbik megoldást, és figyelmen kívül hagyhatjuk a fel nem pattanó zárhangokat. Százalékosan így is nagyon jó eredményt fogunk kapni. Nem kell a folyamatos képzésű beszédhangok esetében jól bevált módszeren sem változtatnunk, ha a zárhangoknak a zár részét különválasztjuk a felpattanástól (l. Tóth László – Kocsor András, megjelenés előtt). Ez a szegmentálási mód a záralkotást nem tekinti a zárhang részének, hanem a megelőző magánhangzó lecsengéseként elemzi. A zöngétlen zárhangok zár részét mínusszal, a zöngéseket pedig plusszal jelöli. A beszédfelismerő program számára ezek a szakaszok csak az utánuk következő felpattanás előrejelzésére, valamint a zárhang időtartamának a megállapítására szolgálnak. Magát a zárhangot a program csak a felpattanásból azonosítja. Az első ábra a *nap* szó szegmentálását mutatja be. A hullámforma alatti átírat a SAMPA átírását követi (SAMPA 2003).

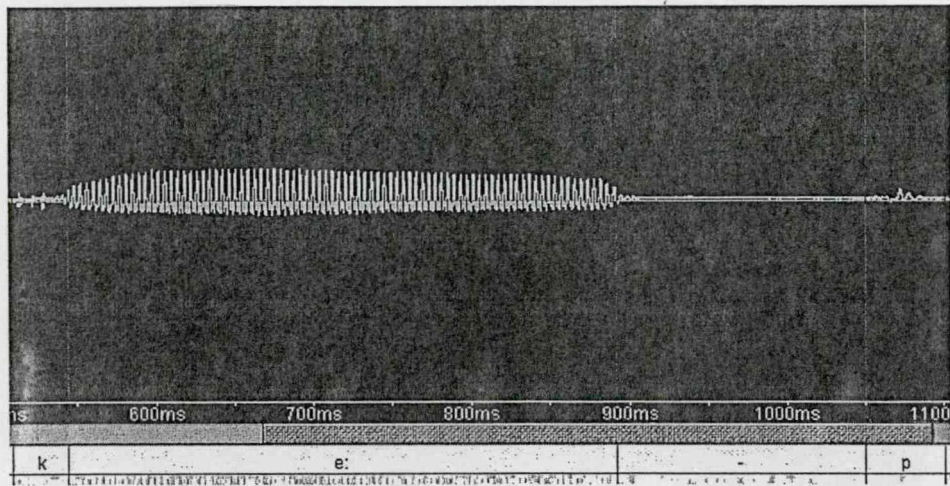


Az ábrán jól megfigyelhető a mínusszal jelölt zárszakasz utáni, amplitúdónövekedéssel együtt járó felpattanás. A szó végi zöngés zárhangok felpattanása nem ilyen erőteljes:



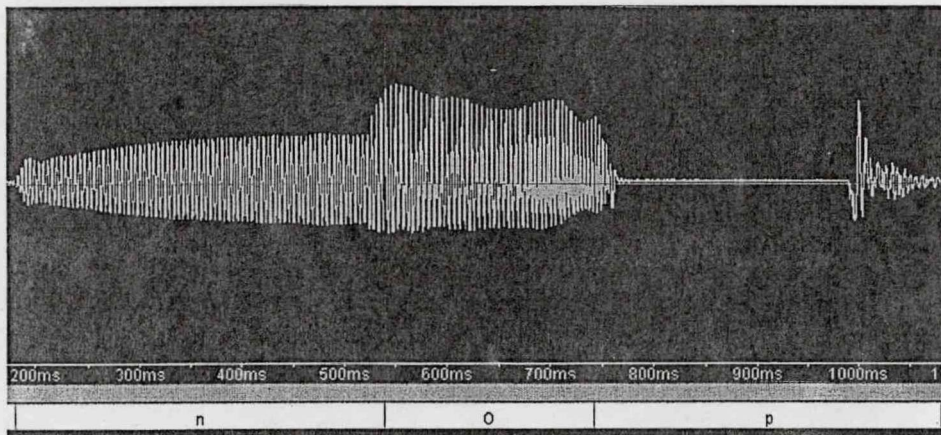
Ha lejátszuk a *nap* [a]-ját vagy a *szöveg* [e]-jét, emberi füllel mindkét magánhangzó lecsengéséből azonosítani tudjuk az utánuk következő zárhangot, de ezt a kulcsot a beszédfelismerő program nem veszi figyelembe, mivel a magánhangzók kezdetét és végét levágja. Ezekben a példákban nincs is szükség erre az információra, hiszen a [p] és a [g] jól azonosítható a zárrész utáni, amplitúdónövekedéssel együtt járó felpattanásból.

Az alábbi ábra egy hibás szegmentálást mutat be. A fel nem pattanó [p]-vel ejtett *kép* szó hullámképében a néma szakaszt követő amplitúdónövekedés nem a [p] felpattanó zöreje, hanem valamilyen egyéb zaj:

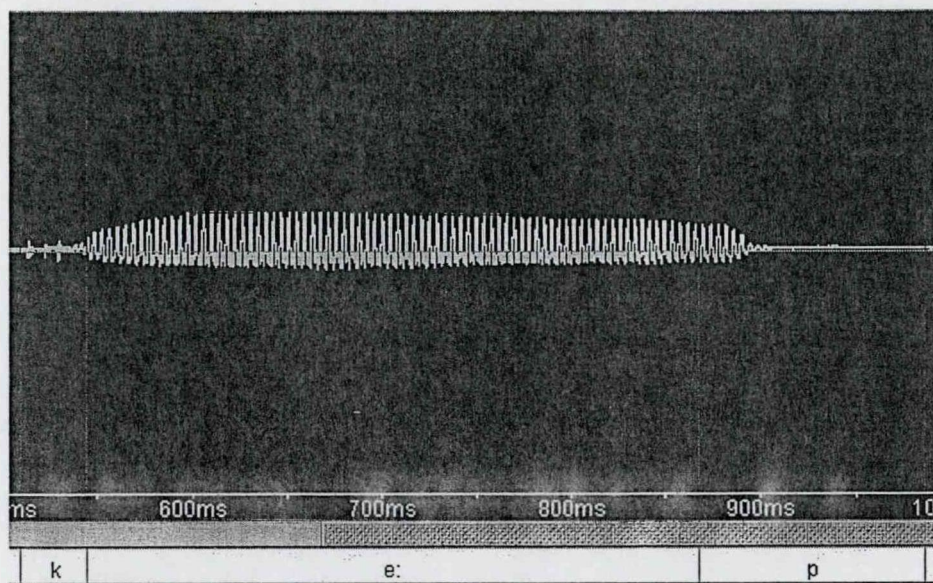


Ezt hiába húzza be a szegmentáló [p]-ként, a beszédfelismerő program nem fogja tudni belőle megtanulni a [p]-t azonosítani, mert nincs benne semmi olyan, ami a [p]-re emlékeztetne. Az anyanyelvi beszélők nem is erről ismerik fel a [p]-t, hanem az [é] végén hallható zörejről.

Ha nem ragaszkodunk ahhoz, hogy a beszédfelismerő program minden szakasznak a közepét figyelje, ahhoz viszont igen, hogy az emberekhez hasonlóan a gép is felismerje a fel nem pattanó zárhangokat, akkor a zárhangok szakaszát a záralkotástól a zár felpattanásáig kell kijelölnük:



A zárhangok felismerőmechanizmusa így valamivel bonyolultabb lesz, mert ezeknek így, a többi beszédhangtól eltérően, nem a közepét, hanem a széleit kell ellenőrizni. Ez a módszer megnehezíti ugyan a programozó munkáját, de csak az övét. A szegmentálóról nagy terhet vesz le azzal, hogy nem kell a zárhangok zár részét a elkülöníteniük a felpattanástól. Ráadásul a felismerés hatékonyságát is javítja azzal, hogy nemcsak a felpattanó zárhangok felismerését teszi lehetővé, hanem a fel nem pattanókét is:



Irodalom

- Kassai Iлона. 1994. A fonetikai háttér, in Kiefer Ferenc szerk. *Strukturális magyar nyelvtan 2: Fonológia*, 581–665. Budapest: Akadémiai Kiadó.
- Kassai Iлона. 1998. *Fonetika*. Budapest: Nemzeti Tankönyvkiadó.
- Kassai Iлона. 2003. Fonetika, in Kiefer Ferenc – Siptár Péter szerk. *A magyar nyelv kézikönyve*, 507–548. Budapest: Akadémiai Kiadó.
- Papp István. 1966. *Leíró magyar hangtan*. Budapest: Tankönyvkiadó.
- R. Molnár Emma. 1989. *Leíró magyar hangtan*. Budapest: Tankönyvkiadó.
- SAMPA. 2003. <http://www.phon.ucl.ac.uk/home/sampa/home.htm>
- Tóth László – Kocsor András. Az MTBA Magyar telefonbeszéd-adatbázis kézi feldolgozásának tapasztalatai. Megjelenés alatt a *Beszédkutatás 2003* kötetben.
- Vértes O. András. 1952. *Bevezetés a fonetikába*. Második, bővített kiadás. Budapest: Gyógypedagógiai Tanárképző Főiskola.
- Vértes O. András. 1982. Az artikuláció akusztikus vetülete, in Bolla Kálmán szerk. *Fejezetek a magyar leíró hangtanból*, 155–164. Budapest: Akadémiai Kiadó.