

Interaktív formánsérték-módosító fejlesztése

Abari Kálmán¹, Olasz Gábor²

¹ Debreceni Egyetem, Pszichológia Intézet
abari.kalman@arts.unideb.hu

² BME Távközlési és Médiainformatikai Tanszék
olasz@tmit.bme.hu

Kivonat: A cikkben bemutatjuk egy webalapú interaktív formánsérték-módosító program felépítését és használatát. Az alkalmazás kötött szerkezetben várja a kiinduló formánsértékeket, melyeket egy Flash-ben készült program segítségével tudunk kényelmesen módosítani, azaz hozzáigazítani a hangszínekhez. A kiindulási és módosított értékeket is MySQL adatbázisban tároljuk, melyek fel- és letöltésről külön funkció gondoskodik. A formánsmódosítás során használt hangszínek megjelenítéséhez a WAV formátumú hangfájlok feltöltése is szükséges. A fejlesztést a magyar formánsadatbázis készítése és továbbfejlesztése ihlette.

1 Bevezetés

Az elmúlt 2 évben már bemutattuk az első magyar formáns adatbázist, amely a BME Távközlési és Médiainformatikai tanszékén kezdeményezett félautomatikus formáns-elemző eljárás alapján [1,2,3]. A formánsmeghatározáshoz használt szóadatbázis a következő adatokat tartalmazza minden szóra: ortografikus szöveg, fonetikai átírat, a szó hullámformája (férfi és női ejtésben), hanghatár-jelölések a hullámformában és a mért formánsok. Az adatbázis szabadon hozzáférhető, webalapú keresőfelülettel rendelkezik (<http://magyarbeszed.tmit.bme.hu/formans>). A teljes formánsadatbázisban közel 3000 szó és összesen 10 391 magánhangzó szerepel. Egy magánhangzón belül 3 mérési pontot jelöltünk ki: a teljes hang időtartamának 25, 50 és 75%-os pontját. Kivételt képeztek a kezdő és befejező magánhangzók, ahol csak két mérési pontot vettünk fel: kezdőhöz 50% és 75%, befejezőhöz 25% és 50%.

A formánsadatbázis létrehozása során hozzátvetőleg a magánhangzók negyedében volt szükség a formánsértékek kézi korrekciójára. Már ekkor felmerült, hogy szükség lenne egy interaktív formánsérték-módosító eszközre, amely a grafikus felhasználói felület előnyeit kihasználva, kényelmes formánsérték-leolvasást tesz lehetővé a szó színe alapján, és így az esetleges korrekciók is rugalmasabban megoldhatók. Jelen cikkben ennek az eszköznek egy továbbgondolásáról számolunk be, amely megnyitja az utat további formánsadatbázisok készítése előtt azzal, hogy lehetővé teszi tetszőleges beszédatadatbázisból származó – többnyire automatikus módszerekkel meghatározott – formánsértékek egyszerű, vizuális alapú kézi javítását.

2 Az interaktív formánsérték-módosító felépítése

Az interaktív formánsérték-módosító eszköz egy szabadon hozzáférhető webes alkalmazás, melynek fő komponensei a MySQL adatbázis, a PHP/HTML forráskódú állományok és a Flash-ben készült „animáció”. Egyelőre az alkalmazás béta verziója készült el, várhatóan az év végére az alkalmazás minden funkciója elérhető lesz a <http://magyarbeszed.tmit.bme.hu/ifem> címen.

A használat szempontjából az alkalmazás 3 fő részt tartalmaz: (1) a formánsadatbázis-feltöltőt, (2) a formánsértékeket módosító Flash alkalmazást és (3) a javított beszédatadabázist eltároló modult. A következőben ezeket tekintjük át részletesebben.

2.1 A formánsadatbázis feltöltése

A formánsmódosító programunk a saját adatbázisába feltöltött formánsfrekvencia értékeket ajánlja fel korrigálásra. Ezt a beszédatadabázist nevezzük a továbbiakban *formánsadatbázisnak*, mely alapvetően címkézési adatokat és hangfelvételeket tartalmaz. A formánsadatbázisba feltöltendő adatok forrása egy ún. *nyers formánsadatbázis*, mely legtöbbször valamilyen automatikus formánsmeghatározó algoritmus segítségével áll elő. A nyers formánsadatbázisból kell előállítanunk a feltöltéshez szükséges két állományt: (1) egy kötött szerkezetű, tabulátorral tagolt szöveges állományt és (2) a bemondásokat tartalmazó WAV fájlok (ajánlott 22 kHz, 16 bit, de nem követelmény) tömörített állományát.

A nyers formánsadatbázisban a bemondások alapegysége lehet szó, de a szónál kisebb (akár egy magánhangzó) vagy szónál nagyobb nyelvi egység is, erre nézve nincs megkötés a feltöltés szempontjából. A címkézéssel kapcsolatos adatokkal szemben azonban elvárás, hogy álljon rendelkezésre minden bemondáshoz (1) az ortografikus szöveg, (2) a fonetikus átírat, (3) a hanghatárok és (4) valamilyen formánsmeghatározó algoritmussal megmért idő- és formánsfrekvencia-érték párok halmaza.

2.1.1 A tagolt szöveges állomány előkészítése

Feltöltés előtt a rendelkezésre álló – tetszőleges nyers formánsadatbázisból származó – adatainkat konvertálni kell egy tabulátorral elválasztott szöveges állományba. A tagolt szöveges állomány minden sora egy-egy beszédhangra vonatkozó információt tartalmaz. Ezek tipikusan magánhangzók vagy zöngés mássalhangzók lesznek, de formailag erre semmilyen megkötés nincs.

A tagolt szöveges állomány kötelezően tartalmaz fejléctet, azaz az első sor az oszlopnevek tabulátorral elválasztott listája lesz. Javasoljuk a következő oszlopnevek használatát:

WAV	HANGSORSZAM	FORMANS	HANGHATAROK	BETUSOR	HANGSOR	BESZELO	ID_2
-----	-------------	---------	-------------	---------	---------	---------	------

A második sortól kezdve azoknak a beszédhangoknak az adatai következnek az oszlopnevek fenti sorrendjében, amelyeknek a formánsértékeit szeretnénk vizuálisan

ellenőrizni, esetleg manuálisan módosítani. A WAV mezőbe a hangfájl neve kerül, elérési út nélkül. Feltételezzük, hogy két azonos nevű hangfájl nem fordul elő a szóveges állományban. A HANGSORSZAM mező a hangállományban tárolt bemondás szegmentáltságát tételezi fel, és annak a hangnak a sorszámát tartalmazza, amelynek a formánsait vizsgáljuk, illetve módosítani akarjuk. Egy hangfájlhoz (azaz WAV állományhoz) természetesen több hangsorszám mező is tartozhat, ebben az esetben ez a szóveges állományban új sorként fog megjelenni. Új sorban meg kell ismételni a WAV fájl nevét és a HANGSORSZAM mezőbe a következő, formánsmódosításban részt vevő hang sorszámát kell megadnunk. A FORMANS mezőbe a beszédhang iniciális formánsértékei kerülnek, pl.:

0.103:746;0.122:788;0.1538:810;0.1856:759;0.2047:647@0.103:1359; 0.122:1382;
0.1538:1380; 0.1856:1398; 0.2047:1364@0.103:2698; 0.122:2670; 0.1538:2700; 0.1856:2532;
0.2047:2464@0.103:2900; 0.122:3363; 0.1538:3299; 0.1856:3455; 0.2047:3455

A fent felsorolt összes idő- és formánsérték egy adott beszédhangra vonatkozik (tehát sortörés nélkül egy sorba kellett volna őket írunk). Először az F1, majd az F2, F3, F4 értékei következnek. Az egyes formánsokat kukac (@) karakterrel választjuk szét, a formánsion belüli, időben elkülönülő méréseket pontosvessző tagolja. Egy méréshez két adatra van szükség egy időkoordinátára (s, másodperc) és egy frekvenciaértékre (Hz). Ezt a két koordinátát kettőspont (:) választja el. A fenti példában minden formánsra 5 mérési pontot adtunk meg és minden formáns esetén azonos időkoordinátákat használtunk (0.103 s, 0.122s, 0.1538 s, 0.1856 s és 0.2047 s). Most ezek a beszédhang hosszának 10, 25, 50, 75 és 90%-os pontjait jelentik, így most rendelkezésre áll 5 olyan mérési hely, amely a teljes hangot lefedi. Mivel abszolút időértékeket kell megadnunk a FORMANS mezőben, a mérési helyek rendszerének kialakítása tetszőleges lehet. A formánsérték módosító alkalmazásunkban az összes itt tárolt frekvenciaértéket meg tudjuk változtatni.

A következő négy mező a WAV állományban tárolt bemondást jellemzi és nem a sort meghatározó beszédhangot. Ennek megfelelően értékük minden olyan sorban azonos lesz, ahol a WAV mezőben is azonos érték található. Ez redundáns tárolást jelent, de így egyszerűbb, könnyebben kezelhető szerkezetet kapunk. A HANGHATAROK mező pontosvesszővel elválasztva tartalmazza a teljes bemondás másodpercben (s) meghatározott hanghatár-jelölőinek időkoordinátáját. A BETUSOR mezőbe a bemondás ortografikus szövege kerül, tagolás nélkül. A HANGSOR mező pontosvesszővel elválasztva a fonetikai átíratot tárolja. Itt tetszőleges jelölést használhatunk, bármit, amit az ASCII szóveges állomány tárolni enged (pl. TMIT, SAMPA). A BESZELO mezőbe a hangfelvételt adó személyről adhatunk egy leírást (pl. azonosító, neme stb).

Az ID_2 mezőbe egy tetszőleges karaktersorozat szerepeltethetünk, amely az adatok visszatöltését segíti a saját adatbázisunkba a formánsmódosítás után. Ez a mező tipikusan a forrás adatbázis valamilyen azonosítóját tartalmazza, mely vonatkozhat bemondásra vagy akár beszédhangra is. Szerepe egyértelműen a formánsmódosítás eredményének egyszerű visszavezetése a korábban használt adatbázisunkba.

Amennyiben előállítottuk a tabulátorral tagolt szóveges állományt, érdemes néhány ellenőrzést elvégezni. (1) A WAV és HANGSORSZAM mezők együtt egyértelműen azonosítják a szóveges állomány sorait (azaz elsődleges kulcsok). (2) Azok-

ban a sorokban, ahol a WAV értéke megegyezik, ott a HANGHATAROK, a BETUSOR, a HANGSOR és a BESZELO mezők értéke is megegyezik. (3) A HANGHATAROK mezőben a pontosvesszővel elválasztott értékek száma mindig eggyel több, mint az ugyanazon sor HANGSOR mezőben lévő pontosvesszővel elválasztott elemeinek száma. (4) A HANGSORSZAM mező a HANGSOR pontosvesszővel elválasztott elemeinek valamelyikének sorszámát tartalmazza (1-től induló sorszámmal).

A fenti ellenőrzéseket, néhány további kíséretében, maga a program is elvégzi, miközben a szöveges állományt eltárolja az adatbázisban. A weboldalon ez egy egyszerű állománykiválasztást követően automatikusan végbemegy, az esetleges hibák, illetve a feltöltött sorok statisztikája szintén a weboldalon követhető nyomon.

2.1.2 A WAV állományok feltöltése

A formánsértékek kézi módosításnak alapja a hangszínekép. Ezek létrehozásához a bemondásokra is szükség van. Az összes WAV állományt gyűjtjük össze egy könyvtárba, majd csomagoljuk őket össze ZIP tömörítővel. Az összecsomagolt állományt a weboldal megfelelő funkciójának kiválasztásával tölthetjük fel a szerverre. A hangszínekép létrehozása után a WAV állományok a szerverről automatikusan törlődnek, a továbbiakban nincs szerepük.

A formánsadatbázis két komponensének (tagolt szöveges állomány, ZIP fájl) feltöltése után a weboldalon tájékoztatást kapunk a MySQL adatbázisba felmásolt formánsérték adatokról (különböző bemondások száma, a formánsaiban módosítható beszédhangok száma, beszédhangonként a formánsok száma, illetve a mérési pontok száma formánsenként). Az elkészült hangszíneképekről is kapunk egy statisztikát, ellenőrizhetjük, hogy az adatbázisunk minden bemondásához elkészült-e a hangszínekép.

2.2 A formánsértékek módosítása

Az adatbázisba kiindulásképpen feltöltött formánsértékek módosítását egy Flash alkalmazás végzi el. Megmutatja a bemondás hangszíneképét és pontokkal jelzi az adott mérési helyeken a nyers formánsadatbázisból származó, korábban meghatározott formánsértékeket. Az azonos formánsához tartozó, de különböző mérési pozíciókban megjelenő pontokat egyenes vonal köti össze. Az 1. ábra a *lábmelegítő* szó (mint bemondási egység) elejének formánsmódosítását szemlélteti. A kép tetején lévő szűrke görgetősáv tájékoztat minket, hogy a képernyőn nem látjuk a teljes bemondást. A görgetősáv alatt TMIT hangjelöléssel a bemondás fonetikai átíratát láthatjuk, mely a HANGSOR mezőből származik. Az éppen formánsmódosítás alatt álló beszédhang szimbólumát halványkék háttérszínnel jelzi a program.

A Flash alkalmazás legnagyobb részét a hangszínekép teszi ki. A hangszíneképeket az R statisztikai program [4] *seewave* [5] csomagjával készítettük, és a HANGHATAROK adatbázismező segítségével rajzoltuk meg a hanghatár jelölő függőleges vonalakat. A frekvencia tengelyt 0-5000 Hz-ig jelenítjük meg. A hangszínekép mint képállomány magasságát figyelembe véve (347 pixel), a formánsértékek módosításának pontossága 14 Hz, azaz egyetlen pixelnyi mozgás az y tengely men-

tén kb. 14 Hz-et jelent a frekvenciatengelyen. Az időtengely mentén egyszerre kb. 0,8 másodpercet láthatunk a bemondásból 540 pixel széles területen. Egy rövid, 50 ms-os magánhangzóra ekkor kb. 34 pixelnyi széles terület jut. A formánsméréseket reprezentáló piros pontok szélessége 6 pixel, így maximum 5 mérési helyhez tartozó pontot tudunk egyszerre úgy megjeleníteni, hogy az a kézi módosítás során ne legyen zavaró. A most felsorolt, megjelenítésből adódó korlátozásokat a program használata előtt vegyük figyelembe, a mérési pontok számát és a módosított adatokból levonható következtetéseket ez alapján határozzuk meg!

A hangszínekép alatt szöveges mezőket láthatunk, amelyekben az éppen módosított formánsfrekvencia érték jellemzőit olvashatjuk: formánsorszám (pl. F4), időkoordináta és frekvenciakoordináta. A bemondó személyéről is kapunk tájékoztatást a BEMONDO adatbázismező alapján.

A formánsfrekvenciák módosítását billentyűzet segítségével végezhetjük el. Egy szokásos munkamenet a weboldalon megjelenő Flash alkalmazással a következő lehet:

1. Egérrel kattintunk Flash alkalmazás területén, pl. a hangszíneképen. Ezzel aktiváljuk a programot, amely most már fogadja billentyűparancsainkat.
2. Eldöntjük, hogy a szó mely beszédhangját szeretnénk vizsgálni, módosítani. A hangok közötti választást a Ctrl+JOBBRA NYÍL és a Ctrl+BALRA NYÍL segíti. A hangok közötti mozgás a hangszínekép görgetését is maga után vonhatja, amit a felső görgetősávon követhetünk nyomon. A hangok közötti váltásnál a program biztosítja, hogy a vizsgált hang környezetét is láthassuk.
3. A magánhangzón belül a módosítandó formáns kiválasztására a kurzormozgató nyilakat használhatjuk (LE NYÍL, FEL NYÍL, BALRA NYÍL, JOBBRA NYÍL). Az aktuális pontot eltérő színezés különbözteti meg a többi ponttól. A pontok közötti mozgás hatása az alsó információs mezőkben is nyomon követhető.
4. Az aktuális pont – és így a formánsérték – mozgására a Q és A billentyűket használhatjuk. A Q-val növeljük az A-val csökkentjük a formánsértéket. Az információs mezőben ezt is követhetjük.
5. A módosítások mentésére az ENTER billentyűt használjuk. Ez azonnal az adatbázisba rögzíti a módosításokat.

Összefoglalva a Flash alkalmazásban használatos billentyűparancsok:

Ctrl+JOBBRA NYÍL és a Ctrl+BALRA NYÍL: az aktuális beszédhang kiválasztása, a hangszínekép vízszintes görgetése

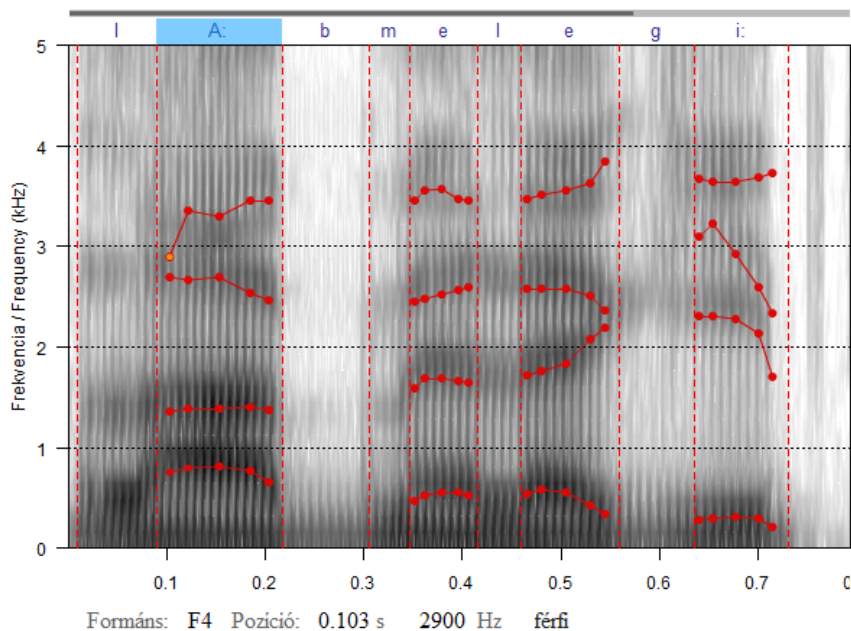
LE NYÍL, FEL NYÍL, BALRA NYÍL, JOBBRA NYÍL: az aktuális pont kiválasztása az aktuális hangon belül

Q és A billentyűk: a pont mozgása fel és le

ENTER: a változtatások mentése.

2.3 A javított beszédatadabázis mentése

A formánsmódosítás elvégzése után a javított adatokat tartalmazó tabulátorral tagolt szöveges állomány mentése következik. A mentés során letöltött adatok mindenben megegyeznek a feltöltés során használt adatbázissal, kivéve, hogy az kiegészül a korrigált formánsértékeket tartalmazó FORMANS_JAV oszloppal. Ez hasonló szerkezetben tárolja az idő- és formánsfrekvencia-értékeket, de természetesen már a korrigált adatokat tartalmazza.



1. ábra. Az interaktív formánsmódosító Flash alkalmazás képe a vizuális megfigyeléshez. A *lábmelegítő* szó első magánhangzójában az F4 első (0.103 s) pontban mutatott értékét módosíthatjuk.

3 Összefoglalás

Jelen cikkben egy webalapú formánsérték-módosító program felépítését mutattuk be. Az automatikusan meghatározott formánsfrekvencia értékek kézi módosítása a beszéd hangszíne alapján történik, amelyet szintén az alkalmazás állít elő. Erre a hangszínekre vetíti rá a program az automatikus mérésből származtatott Hz értékeket (kis pontok formájában). Ez adja a vizuális ítékezés alapját. Amennyiben az automatikusan meghatározott formánsérték kiugróan eltér a hangszíneken leolvashatótól, akkor a mért értéket a hangspektrum alapján módosítjuk, és ezt eltároljuk a

formánsadatbázisunkban. A Flash alkalmazásban billentyűparancsokkal határozhatjuk meg a módosítás helyét (hang), a hangon belül a formánst és végül a formáns függőleges pozícióját, azaz a formánsfrekvencia értéket. Tetszőleges hosszúságú bemondás (hang, szó, mondat) formánsait tudjuk kezelni, a hangszínekép vízszintesen görgethető. A megjelenítés ennek ellenére rendelkezik korlátokkal, ezeket a mérések megkezdése előtt figyelembe kell venni. Az alkalmazás segíti a nagy pontosságú formánsadatbázisok létrehozását.

Bibliográfia

1. Abari K., Olasz G.: Magyar formánsadatbázis az interneten. In: Gósy, M. (szerk.): Beszédkutatás. MTA Nyelvtudományi Intézet, Budapest (2011) 73–82
2. Olasz G., Rác Zs. Zs., Bartalis M.: Formánsmérések automatizálása, formánsadatbázisok létrehozása. In: Gósy M. (szerk.): Beszédkutatás 2009. MTA Nyelvtudományi Intézet, Budapest (2009) 134–147
3. Rác Zs., Abari K., Olasz G.: A formant trajectory database of Hungarian vowels. In: Németh G., Olasz G. (eds.) *The Phonetician* 97 (2011) 6–13 (<http://www.isphs.org>)
4. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
5. Sueur, J., Aubin, T., Simonis, C. Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics* Vol. 18 (2008) 213–226