Interaktív fonetikai eszköz az artikulációs csatorna keresztmetszet-függvényének meghatározására

Jani Mátyás[1], Björn Lindblom[2], Sten Ternström[3]

[1] Pázmány Péter Katolikus Egyetem, ITK,
Budapest, Práter utca 50/A, e-mail: janma@digitus.itk.ppke.hu
[2] Department of Linguistics, Stockholm University
106 91 Stockholm, Sweden
[3] Department of Speech, Music and Hearing, School of Computer Science and Communication, Kungliga Tekniska Högskolan (Royal Institute of Technology)
100 44 Stockholm

Kivonat A projekt célja annak az eldöntése volt, hogy a SuperCollider programozási környezet mennyire alkalmas egy interaktív artikulációs modell implementálására. Az elkészült szoftver az APEX nevű, kétdimenziós modellt használja, amit az artikulációs csatorna alakja és a formánsok közötti összefüggés vizsgálatára hoztak létre.

Kulcsszavak: artikulációs modell, supercollider, beszédszintézis

1. Bevezetés

Manapság a konkatenatív beszédszintetizálásra használt módszer a legelterjedtebb, annak ellenére, hogy az összefűzéssel készített beszédhang minősége elmarad az artikulációs módszer által elméletileg előállítható beszédhang minőségétől. Emiatt újabban egyre nagyobb figyelmet kap az artikulációs beszédszintetizálás és egyre több artikulációs modell jön létre [1]. Ezen modellek feladata nem mindig a beszédszintetizálás, használhatók kutató és pedagógiai eszközöknek is. Segítségükkel többek között meg lehet figyelni a formáns frekvenciák és az artikulációs csatorna alakja közötti összefüggést. Jelen munka fő célkitűzése egy meglévő kétdimenziós artikulációs modell implementálása, valamint a SuperCollider környezet ilyen jellegű feladatra való használhatóságának kiderítése.

2. APEX modell

Az eredeti APEX program célja adott artikulációból formáns adatok (frekvencia, sávszélesség) kinyerése volt [2]. A modell egy virtuális kétdimenziós artikulációs csatornát használ, ennek geometriáját tesztalanyról készített röntgenképekből nyerték ki. A formáns adatok előállításához több lépésre van szükség. Először

az ajkak, a nyelvcsúcs és nyelv törzs állapotaiból, az állkapocs és a gégefő helyzetéből egy artikulációs profil készül egy mesterséges középvonallal, ami az artikulációs csatorna első és hátsó oldala között félúton helyezkedik el. Ezután le lehet mérni a középvonal mentén tetszőleges pontokban az artikulációs csatorna keresztmetszetét. A keresztmetszetek hosszát egy adott szabály felhasználásával keresztmetszeti területekké kell konvertálni, ez már lényegében az artikulációs csatorna csőmodelljének felel meg. Hangszintézis megvalósításának egyik módja a formánsszintézis, ehhez a csőmodellből ki kell nyerni a formánsparamétereket. Az APEX modell az orrüreget nem modellezi, így a nazális hangokat nem tudja megfelelően szintetizálni.

2.1. Adatok kinyerése

A körvonalak és egyéb geometriai adatok kinyeréséhez röntgenfelvételekre volt szükség [3]. A röntgenfelvételek fő problémája, hogy a tesztalanyokat sugárzás éri és a biztonság érdekében bizonyos biztonsági előírások korlátozzák a felvételek hosszát és az elszenvedett sugárzási mennyiséget. A hangképzőszervek körvonalai 0,5 - 1 mm pontossággal határozhatók meg.

A keresztmetszetek számításához szükséges együtthatók meghatározásához keresztmetszeti MR (mágneses rezonancia) képeket készítettek az artikulációs csatorna mentén több helyen [4]. A felvétel alatt használt szöveganyag svéd magánhangzókat tartalmazott, és az MR képek mellett videó- és hangrögzítés is történt.

2.2. Keresztmetszetek területekké alakítása

A kétdimenziós módszerek közvetlenül csak az artikulációs csatorna oldalnézeti keresztmetszetét tudják felhasználni. A valódi alakzatok nem állnak rendelkezésre, így az artikulációs csatorna irányára merőleges szeletek területét az oldalnézeti keresztmetszethosszakból kell kiszámolni.

Többféleképpen is lehet becsülni ezeket a területeket [5], általában mérésekből adódó együtthatókat felhasználva. A leggyakrabban Heinz és Stevens (1964, 1965) által publikált hatványfüggvényt használják:

$$A = K \cdot d^{\alpha}$$

ahol A az artikulációs csatorna irányára merőleges metszet területe, d a mért hossz, K és α pedig együtthatók, melyek értéke függ a tesztalanyon és a vizsgált metszet pozícióján.

2.3. A nyelv alakjának meghatározása

A nyelv alakjának paramétereit főkomponens-analízis segítségével határozták meg. Körülbelül négyszáz nyelvkörvonalat nyertek ki röntgenképekből, majd

ezeket a körvonalakat 25 pontban mintavételezve tárolták [6]. A főkomponens-analízis eredménye néhány bázisfüggvény súlyozott lineáris kombinációja:

$$V(x) = N(x) + c_1(v) \cdot PC_1(x) + c_1(v) \cdot PC_1(x) + ...$$

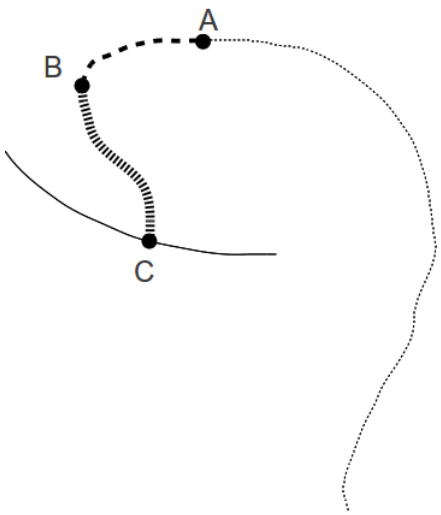
ahol x a kontúr mintavételezett pontjának indexe, $V(x)$ a kiszámolt nyelvalakzat, $N(x)$ egy semleges nyelvkontúr (a megfigyelt körvonalak átlaga) és $PC_i(x)$ az i. bázisfüggvény. Az egyes c_i együtthatók a bázisfüggvények súlyai. c_i egy kétdimenziós vektor, értéke a megszólaltatott magánhangzótól függ, amit bemeneti paraméterként használ a modell.

Pontosság: egyetlen PC bázisfüggvénnyel 85,7% pontosságot lehetett elérni, két bázisfüggvénnyel már 96,3%-ot [6].

2.4. Artikuláció

A modellben használt artikuláció egyszerűsített változata a tényleges artikulációnak. Csak a programban megvalósított részeket mutatjuk be. A hangképző szervek közül néhányat rögzített alakzatként kezeltünk, ilyen például az artikulációs csatorna hátulsó fala és a szájpadlás. A mozgatható alakzatok közé tartozik a gége a hangszalagokkal, a nyelv és az egész alsó állkapocs.

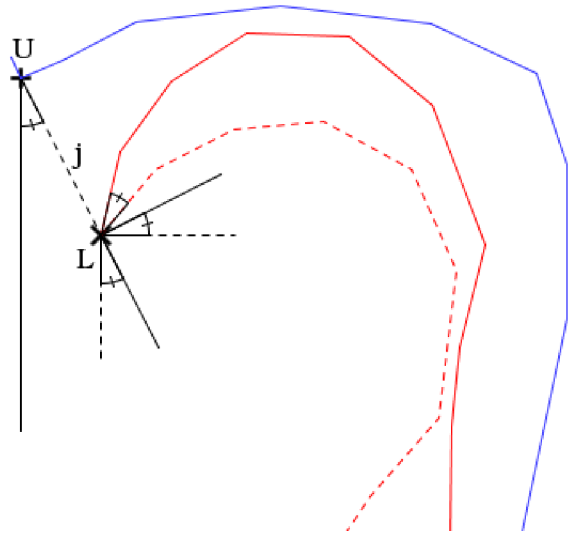
A gége fix kontúrral rendelkezik, azonban függőleges irányban mozgatható, ezzel lehet rövidíteni, illetve hosszabbítani az artikulációs csatornát.



1. ábra. A nyelv alakja három részből tevődik össze.

A nyelv alakja 3 részből áll (1. ábra). A hátulsó részének formáját a főkomponens-analízissel nyert egyenlettel számoljuk ki. A nyelv csúcsának helyzete (B pont) külön állítható, a csúcspontot Hermite interpolációval készített görbe

köti össze a hátsó nyelvformával. Ahhoz, hogy a kapcsolódás törésmentes legyen, az első derivált használatára is szükség volt a kapcsolódási pontban (A pont). A nyelv csúcspontja a szájüregben a száj alsó részén egy rögzített ponthoz (C pont) csatlakozik. Ennek a harmadik görbének az alakjához megfigyelt adatokat használtunk fel.



2. ábra. Az alsó állkapocs mozgatása.

Alsó állkapocs mozgása az alsó állkapocs koordináta rendszerének eltolását és forgatását foglalja magába. Ezzel együtt mozog az alsó fogsor, a szájüreg alsó fele és a nyelv. Az elforgatás szögét az alábbi egyenlettel számoljuk:

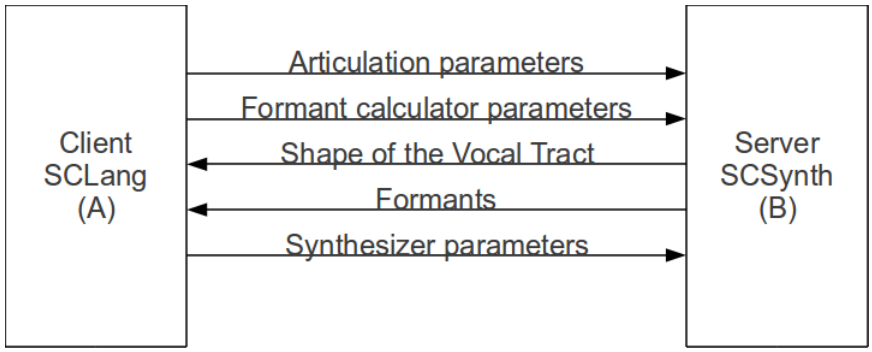
$$\alpha_{deg} = \frac{j}{2} + 7$$

ahol α_{deg} a szög fokban, j pedig az állkapocs nyitottsága (a távolság az alsó és felső metszőfogak között, mm-ben). A 2. ábrán a kék görbe az artikulációs csatorna hátulsó fele, az U pont a felső állkapocs koodináta rendszerének origója. Ha a nyitottság j-re van állítva, akkor U és L közötti távolság j. Az ábrán jelölt összes szög α. A belső szaggatott piros vonal a j-vel eltolt nyelv, a folytonos piros vonal az eltolt, majd elforgatott nyelv.

3. Megvalósítás

A modellt a SuperCollider környezetben implementáltuk. A SuperCollider egy programozási környezet algoritmikus zeneszerzésre és hangfeldolgozásra. Kliens-

szerver architektúrájú a felépítése, a kliensben található interpretált, objektum-orientált small-talk-szerű programozási nyelv felel a szerver vezérléséért. A szerver feladata a gyors jelfeldolgozás, valamint a hang be- és kimenet kezelése, natív bővítmények segítségével [7].



3. ábra. Kommunikáció a SuperCollider szerver és a kliensalkalmazás között.

A megvalósítandó program első verziója csak a kliens oldalon helyezkedett el, a szerver részt csak a hangszintetizáláshoz használta. A sok geometriai művelet sajnos nem volt elég hatékony az interpretált nyelvben, így később a számításigényes részek átkerültek a szerverre. A kliens-szerver közti aszimmetrikus kommunikáció szinkronizálása sok nehézséget okozott (3. ábra).

4. Eredmények

Az APEX modellnek létezik egy korábbi implementációja is, de annak fejlesztése félbemaradt, és a program elavult. Az új program még további fejlesztésre szorul, mivel hiányzik a szájüregi rész helyes kezelése (ajkak, fogak, nyelv alatti terület). Ezt leszámítva a modell megvalósítása sikeresnek mondható. Előrelépés a korábbi változathoz képest, hogy a használt környezetnek köszönhetően könnyebb a programot átírni más platformokra (Linux rendszeren készült, Macen is sikerült futtatni).

A hangszintézis az elkészült új verzióban interaktív, a bemenetet változtatva azonnal hallható a változás eredménye. A bemenő paraméterekből listát készítve lehet több hangot is összefűzni. A többi artikulációs modellhez hasonlóan az APEX-ben is megfigyelhetőek a hangok közötti átmenetek, a koartikuláció. Az artikulációs modell alkalmas a hangátmenetek beszédszervek tényleges fizikai jellemzőin alapuló interpolációjára.

5. Továbblépési lehetőségek

Több irányban is tovább lehet folytatni a fejlesztést. A hiányzó rész elkészítésével a teljes modell meg lenne valósítva. A teljes modell leprogramozása után a modell által kiszámolt formánsfrekvenciákat össze lehetne vetni valóságos mérésekkel.

A program jelenlegi felépítése a szerver-kliens közötti kommunikáció miatt nem ideális. Ennek egyik kiküszöbölési módja, hogyha a SuperCollider kliens helyett saját, natív klienst készítenénk. Ekkor nem lennénk korlátozva az interpretált nyelv sebességével, másrészt a SuperCollider szerver csak a hang kiadásáért lenne felelős, és csak a formánsadatokat kellene továbbítani.

A számítások sebességet tovább lehetne gyorsítani SIMD (Single Instruction Multiple Data) utasításkészlettel, mivel a keresztmetszetfüggvény kiszámításánál például minden keresztmetszeti szeleten ugyanazt az algoritmust kell végrehajtani.

A munka Erasmus ösztöndíj keretében, MSc diplomaterv formájában lett elfogadva a Kungliga Tekniska Högskolan Stockholm Speech, Music and Hearing tanszékén.

Hivatkozások

1. Shadle, C.H., Damper, R.I.: Prospects for articulatory synthesis: A position paper. In: 4th ISCA workshop, Pitlochry, Scotland. (2001)
2. Stark, J., Ericsdotter, C., Branderud, P., Sundberg, J., Lundberg, H.J., Lander, J.: The apex model as a tool in the specification of speaker-specific articulatory behavior. In: Proc XIVth Int'l Congr Phonetic Sci (ICPhS 99), San Francisco. (1999)
3. Branderud, P., Lundberg, H.J., Lander, J., Djamshidpey, H., Wäneland, I., Krull, D., Lindblom, B.: X-ray analyses of speech: Methodological aspects. In: FONETIK 98. (1998)
4. Ericsdotter, C.: Articulatory-Acoustic Relationships in Swedish Vowel Sounds. PhD thesis, Stockholm University (2005)
5. Soquet, A., Lecuit, V., Metens, T., Demolin, D.: Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with mri. Speech Communication **36**(3-4) (2002) 169–180
6. Lindblom, B.: A numerical model of coarticulation based on a principal components analysis of tongue shapes. In: 15th Int'l Congr Phonetic Sci, Barcelona. (2003)
7. Wilson, S., Cottle, D., Collins, N.: The SuperCollider Book. The MIT Press (2011)