

ELŐZETES VIZSGÁLATOK AUTOMATIKUS GESZTUSGENERÁLÁSHOZ

PRELIMINARY EXAMINATIONS FOR AUTOMATIC GESTURE GENERATION

CZAP LÁSZLÓ¹–KILIK ROLAND²

Az alábbi cikkben különböző alanyok fej által kifejezett gesztusai és hangbeli jellemzői időbeli kapcsolatának vizsgálatára kifejlesztett módszereink bemutatása, valamint korrelációs módszerekkel történő vizsgálataink eredményei olvashatók.

Kulcsszavak: gesztusok, intenzitás, hangmagasság, időbeli kapcsolat, korreláció

In the following paper first we present our methods for examining time-domain relationship between head gestures and vocal attributes as well as some significant results for different announcers.

Keywords: gestures, intensity, pitch, time-domain relationship, correlation

Bevezetés

Kutatásunk ezen részének célja, hogy a TÁMOP 4.2.2.C „Alap- és alkalmazott kutatások hallássérültek internetes beszédfejlesztésére és az előrehaladás objektív mérésére” című projekt keretében a virtuális beszélő fej mozgásának természetességét elősegítse a hangváltozások és mozgások kapcsolatának felderítésével, illetve később a pillanatnyi hangváltozások alapján mozgások generálásával.

Kiinduló feltételezésünk részben az volt, hogy kimutatható időbeli szabályszerűségek vannak különböző alanyok fej által kifejezett gesztusa és hangjuk különböző jellemzői között. Ehhez részint alapot szolgáltatott több kutatás, amelyek kapcsolatot mutattak ki különböző gesztusok (például kézmozdulatok) vagy a száj mozgása és hangbeli jellemzők változása között. Ezek azonban időbeli szabályszerűségeket nem vizsgáltak (Busso, Deng, Narayanan, Neumann 2007.; Yeiha, Vatikiotis-Bacon 1998.).

1. A vizsgálatokban felhasznált képanyagok jellemzői

Vizsgálatainkban összesen 36, többségében internetről letöltött videót használtunk fel, melyeket két kategóriába soroltunk be. Az első kategóriába kötött fejmozgású felvételek

¹ CZAP LÁSZLÓ
intézetigazgató, egyetemi docens
Miskolci Egyetem, Villamosmérnöki Intézet
Automatizálási és Infokommunikációs Intézeti Tanszék
3515 Miskolc-Egyetemváros
czap@uni-miskolc.hu

² KILIK ROLAND
tanársegéd
Miskolci Egyetem, Villamosmérnöki Intézet
Automatizálási és Infokommunikációs Intézeti Tanszék
3515 Miskolc-Egyetemváros

tartoztak, jellemzően híradók műsorvezetőinek szereplésével. A második kategóriába szabadabb fejmozgású videók csoportja sorolható, amelyek saját szavakkal történő közléseket tartalmaztak (mint például adott témák saját szavakkal történő felkonferálása). Utóbbi esetben a téma nem spontán voltát ellensúlyozta a saját szavakkal való közlés és az alanyok kamerához szokott, ezáltal nem feszélyezett viselkedése, így a fejmozgás szabadnak mondható.

A szabad fejmozgást tartalmazó anyagok a vizsgálatok szempontjából kedvezőbb tulajdonságokkal (természetesebb közelség álló, erőteljesebb fejmozgás, természetesebb hangsúlyozás) rendelkeztek, így a vizsgálatok többsége ezen 26 felvételre terjedt ki, illetve az eredmények ismertetése is ezek egy csoportját érinti. 22 felvétel műsortárból letöltés formájában került rögzítésre, 4 pedig egy digitális tananyag része. Az alanyok száma összesen négy, így mivel mindegyikükhöz több (4–9) videó tartozik, lehetőség volt a beszélőtől függő jelenségek vizsgálatára, a megfigyelt szabályszerűségek beszélőtől függésének tesztelésére is. A 22 videó közül 9 felvétel egyazon műsorvezető konferálását tartalmazza egy hetenként visszatérő műsorból rögzítve. Közülük a 2. és 3. videó egyazon adás két, illetve a 4–7. videó egy másik adás négy, 10 másodperces szelete. Mivel ezen 6 képanyag nem csupán egy beszélőt tartalmaz, de mindösszesen 2 adás egyforma hosszúságú szeletei, ezek vizsgálatával a többi esethez képest jobban kiszűrhetők voltak mind az esetleges hangulatváltozások, mind a hosszbeli különbségek lehetséges hatásai. További 7 videó egy másik, szintén heti rendszerességgel jelentkező műsor műsorvezetőjének előbbiekhöz hasonló jellegű hozzászólását tartalmazza, ezek közül azonban mindegyik külön adáshoz tartozik. Ugyanezen jellemzők tartoznak további 5 videóhoz. A felvételek hossza 8 és 12 másodperc közötti volt, a hangszótagszám pedig 40 és 70 között változott.

2. A hanganyagok feldolgozása

A videókhoz tartozó hanganyagokból a hangmagasság (pitch) információkat a megfelelő videó képkocka-sebességével megegyező lépésközökkel nyertük ki, majd ugyanezt a módszert alkalmaztuk az autokorrelációs és keresztkorrelációs pitch adatok kinyerésére is. A fenti módokon kinyert háromféle hangmagasságvektort rendre egy másik formában is alkalmaztuk, az adott vektor egymást követő értékeiből különbségvektort (delta pitch) képezve.

Az intenzitás számítása a szótagok időpillanatai által adott szegmenshatárok között az amplitúdók értékeiből különböző módszerek szerint történt, mint például abszolútérték-átlag, négyzetátlag, négyzetösszegek négyzetgyöke, négyzetek átlagának négyzetgyöke. Adott ablakba tartozó amplitúdó értékek W vektorára:

$$|\overline{W}|, \overline{W^2}, \sqrt{\sum W^2}, \sqrt{W^2}$$

A hangmagasság vektora mellett az intenzitásnak is felhasználtuk egy általunk leképzett variánsát. Azt feltételezve, hogy a nagyobb intenzitástartomány szélső értékeihez könnyebben érzékelhető mozgások tartozhatnak, az intenzitásvektor értékkészletének felső és alsó 10 százalékát tartottuk meg.

A létrejött intenzitásvektorok, illetve a pitch, autokorrelációs és keresztkorrelációs pitch (illetve ezek különbségvektorai) vektorok elemeit ezt követően az adott szegmentálási környezet (jelen esetben a szótagokon belüli intenzitásmaximumok) által adott határok között

szegmentáltuk, és a szegmensekben található értékeket átlagoltuk. Szegmentálási határként a szótaghatárokat is kipróbáltuk, ez a módszer azonban kevésbé adott megbízható eredményt, mint a szótagon belüli intenzitásmaximumok használata.

3. A képanyagok feldolgozása

A képi információk közül a pislogás vektora mellett a bólogatás, vízszintes mozgás, biccentés különböző változatai kerültek kimentésre.

Ezen információk kinyeréséhez egy, alapvetően a szemek pozícióját követő alkalmazás készült el. Az alkalmazás bemenő adatként fogadja a videóból szürkeárnyaltos képsorozattá konvertált szekvenciát, melynek első elemén a szemet kijelölve, ennek adott nagyságú környezetét vizsgálva, a legkisebb eltérést megkeresve követi az egymást követő képek között a szemet, és szolgáltatja ebből kiindulva a fenti mennyiségeket.

A pislogás detektálásában három különböző, általunk kialakított, világosságárány-változásokat vizsgáló szabályrendszer empirikus alapon beállított súlyozásos aggregált szabályát használtuk fel, némely esetben manuális korrekciót alkalmazva. Kezdetben a pislogás teljes időtartamát regisztráltuk, azonban pusztán a kezdetek pillanatainak figyelembevétele jobb eredményt szolgáltatott.

A bólogatás információja több különböző formában került kimentésre. Ilyenek voltak például a pusztán pixeles bólogatás, illetve ablakozott bólogatás, amely adott méretű ablakban adott pixelnyi mozgás megléte esetén teljesül, majd meghatározza annak 8 irányhoz legközelebbi irányát. Megkülönböztettünk nagy, illetve kis bólogatást is, a fej helyzetének kezdőpozícióhoz való viszonya alapján. Mindezen mennyiségek olyan változatát is képeztük, amely esetben csupán a bólogatások kezdete került regisztrálásra. Az alany általi biccentést hasonló módszerekkel érzékeltük.

A fentiekben felül megkülönböztettünk egy folyamatos mozgást is mindkét szem esetében, mind a függőleges, mind a vízszintes mozgást tekintve. Ezen vizsgálati mennyiséget úgy képeztük, hogy az adott szem pozíciójának adott irányú koordinátájából rendre kivontuk az adott vektornak a vizsgált videó egészére vett átlagát. Ezzel négy további mozgásvektort hoztunk létre. A későbbiekben a két szem függőleges és vízszintes irányú mozgásának vektora külön-külön átlagolásra is került, ez azonban nem vezetett jobb eredményre, mint a szemenként külön vizsgáltak.

A fent leírt eljárások szerint kinyert képi információs vektorokat az adott szegmentálási környezet (jelen esetben a szótagokon belüli intenzitásmaximumok) által adott határok között – a hanginformációk esetében alkalmazott módszer szerint – szegmentáltuk, és a szegmensekben található értékeket átlagoltuk.

4. A vizsgálatok menete

A fent ismertetett módszer szerint létrehozott képi, illetve hanginformációs mennyiségek szegmentálásra kerültek a szótagon belüli intenzitásmaximumok időpillanataiban (a hangsúlyt közelítve ezzel), majd párosítottuk őket, összesen 136 korrelációs vizsgálati párt képezve. Ilyenek voltak például a folytonos függőleges mozgás és intenzitás, folytonos függőleges mozgás és hangmagasság, pislogás és intenzitás, illetve az összes, fent említett mozgásfajta párosítása az intenzitás- és hangmagasságvektorok különböző variánsaival.

Mivel a vizsgált mennyiségek esetében azok kapcsolata időbeliségének vizsgálata volt a célunk, azokat rendre elcsúsztattuk egymáson, minden egyes eltolási értéken korrelációs módszerekkel vizsgálva a kapcsolat erősségét. Ez oly módon történt, hogy a képi információ vektorát fixnek tekintettük, a hanginformáció vektorából pedig annak elejéről is és

végéről is a csúsztatás méretét levágva, az így kialakult vektort rendre elcsúsztattuk a képi információ vektorán negatív és pozitív irányban is, minden elcsúszásnál a mozgó vektorral azonos hosszúságú részt vizsgálva a fix vektorból is. Ezáltal amennyiben negatív eltolásnál jelentkezett korrelációs maximum, annak jelentése az volt, hogy a fix vektor (a képi információ vektora) értékeinek változásai jellemzően adott eltolással előzik meg a hanginformációs vektor értékeinek változásait. A pozitív irányú eltolás pedig ennek ellenkezőjét jelenti.

Az elcsúsztatás mértéke jelen vizsgálatban 7 (azaz hét, szótagra számolt intenzitásmaximum) volt mindkét irányban.

Általánosságban elmondható, hogy a képi és hanginformációk eltolástól való függését öt korreláció (a Pearson-féle lineáris korreláció, kanonikus korreláció, a Kendall-féle korreláció, a Spearman-féle rangkorreláció, illetve távolságkorreláció) tekintetében vizsgáltuk. Két eltérő megközelítést is alkalmaztunk. Az egyik esetben videók adott csoportjára számított eredő görbéket vizsgáltunk, az így felmerülhető esetleges torzításokat (például 1-1 videó nagy korrelációs értékének torzító hatása az eredőben) különböző módszerek, mint például kölcsönös információ, P érték, szórások, átlagtól való eltérés stb. segítségével szűrve (Czap, Kilik 2014). A másik megközelítésben a korrelációk eltolásonkénti értékeit videónként jelenítettük meg az ez esetben kevésbé szükséges kiegészítő mennyiségek nélkül, azonban a korrelációs vizsgálatokat képező adatsorok tulajdonságainak vizsgálatára saját, egyéb eljárásokat alkalmazva.

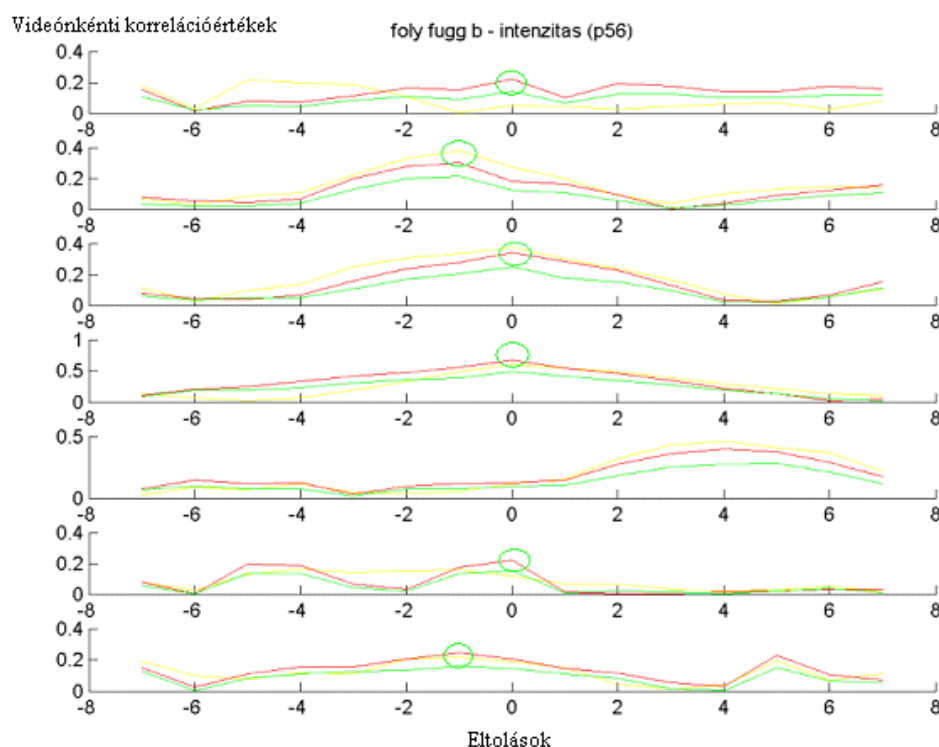
5. A videónkénti vizsgálatok eredményeinek ismertetése

Az alábbiakban néhány releváns példán keresztül mutatjuk be 1-1 alany szereplésével készült felvételek sorozatának korrelációs elemzéseit.

5.1. Folytonos függőleges mozgás és intenzitás

Az 1. ábrán az 56-os korrelációs pár (szegmentált folyamatos függőleges irányú bal szem mozgás és intenzitás) esetében láthatóak a korrelációs eredmények. A vizsgálat egy 7, azonos műsorsorozatból származó videót tartalmazó felvételecsoportra készült, az ábrán az ezen videókra számított értékek sorrendben láthatók. Az egyes görbék az egyes videókhoz tartozó korrelációs értékek, a vízszintes tengelyen pedig az eltolás mértéke látható. A korrelációs módszerek: **zöld** – Kendall-módszer, **piros** – Pearson-féle lineáris korreláció, **sárga** – rangkorreláció.

Az ábráról leolvasható, hogy az adott videók esetében többségében a 0-s és -1-es eltolási értéknél van korrelációs maximum, egyes esetekben a 0,4, 0,5-ös értéket meghaladó maximális korrelációs értékkel. A 0. eltolás maximuma esetében a két vizsgált mennyiség jellemzően együtt változott, míg a -1-es eltolás jelentése az, hogy a mozgás változása jellemzően egy szótaggal (szótagon belüli intenzitásmaximummal) előzte meg az intenzitás változását.



1. ábra. Folyamatos függőleges mozgás és intenzitás viszonya 7 videó esetében

Az adott korrelációs párok jellemzőit vizsgáló egyik módszerünk eredményeit mutatja a fenti videók esetében a 2. ábra, amely a többiekétől eltérő tulajdonságú 5. görbe különbségének okát magyarázza meg. Az ábra oszlopaiban láthatjuk az egyes videók adatait (az 1. ábra videóit tekintve), a sorok jelentése pedig a jobb oldalon olvasható.

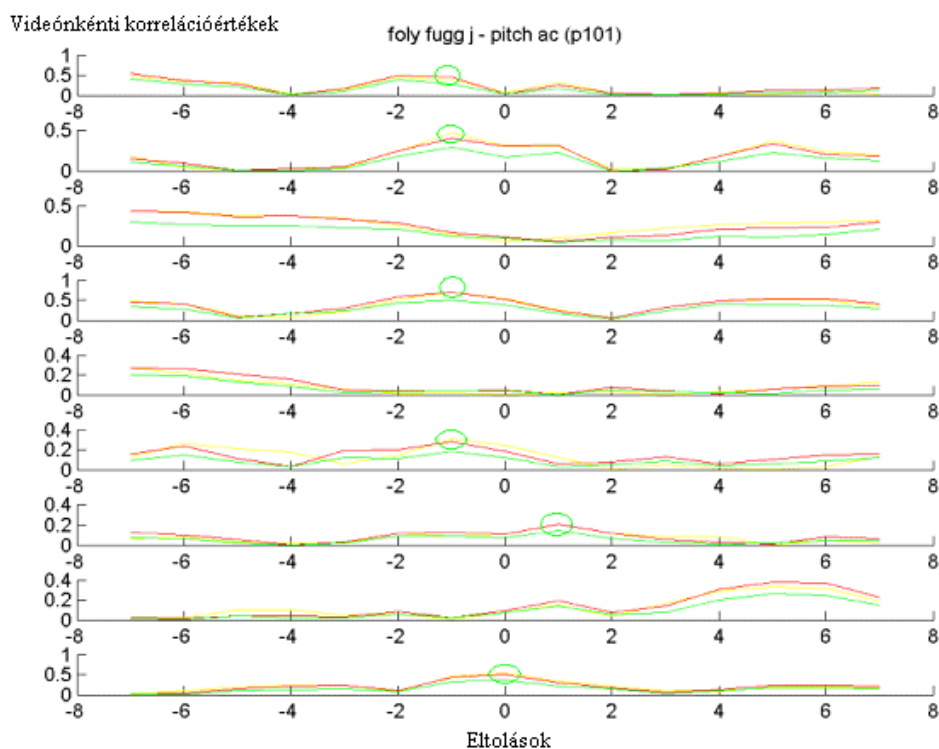
1. videó	2. videó	3. videó	4. videó	5. videó	6. videó	7. videó	
0.0011	0.0012	0.0015	0.0014	0.0007	0.0020	0.0010	intenzitás felső 10%-ok átlaga
272.1701	158.7489	192.9532	176.0341	351.7235	263.8913	140.9275	szemmozgások összege

2. ábra. Folytonos függőleges mozgás és intenzitás vektorainak jellemzői 7 videó esetében

A 1. ábrán látható volt, hogy az 5. videó korrelációs görbéje tér el nagymértékben az ugyanazon műsorvezetővel készült többi képanyag görbétől. Ugyanezen videó oszlopában megfigyelhető két, a többiekétől eltérő értéktartományú releváns adat, ami egyrészt a bal szem mozgásainak összege, illetve az intenzitásvektor értékeinek alsó és felső 10 százalékát (ezen kívül nullákat) tartalmazó vektor átlaga. Látható, hogy ezen videó esetében a bal szem mozgása 350-es értéket képvisel, míg a többi felvételt tekintve a 275-öt sem éri el, az intenzitások felső és alsó 10%-ából képzett vektor elemeinek átlaga ezzel szemben a legkisebb az 5. videónál. Megjegyzendő, hogy ez az összeg, illetve átlag azonban nem pont abból a két vektorból származik, amelyekből a korrelációs pár, azonban azokkal rokon, illetve abból származó mennyiségekből ered.

5.2. Folytonos függőleges mozgás és hangmagasság

A folytonos függőleges mozgás vizsgálata az előbbihez hasonlóan, azonban ez esetben az autokorrelációs pitch-csel párosítva történt meg, a 3. ábrán látható eredményekkel. Mindez egy másik műsorvezető szereplésével készült, a szintén 1. pontban említett tulajdonságokkal rendelkező további 9 videó korrelációs eredményeinek összegzését jelenti. Az ábra megjelenítési tulajdonságai az 1. ábra esetében leírtakéval megegyezők.



3. ábra. Folytonos függőleges mozgás és autokorrelációs pitch viszonya

Az ábráról leolvasható, hogy a videók többségében -1 -es eltoláson $0,5$ -öt megközelítő maximális korrelációs érték volt tapasztalható, illetve egy esetben a 0 -s eltoláson volt regisztrálható a többihez képest kiemelkedő és nagy értékű korrelációs maximum, egy további esetben pedig mindössze egy eltolással arrébb, az 1 -es eltoláson egy kisebb, $0,2$ -es korrelációs érték.

A korrelációs görbékben az is látható, hogy pusztán a 3., 5. és 8. videó esetében kapunk mind jellegében, mind a maximális korrelációhoz tartozó eltolásértékben nagyban különböző (a -1 , 1 tartományon kívül eső) görbéket. Az adatokat megvizsgálva (4. ábra) azonban láthatjuk ezen eltérések okait.

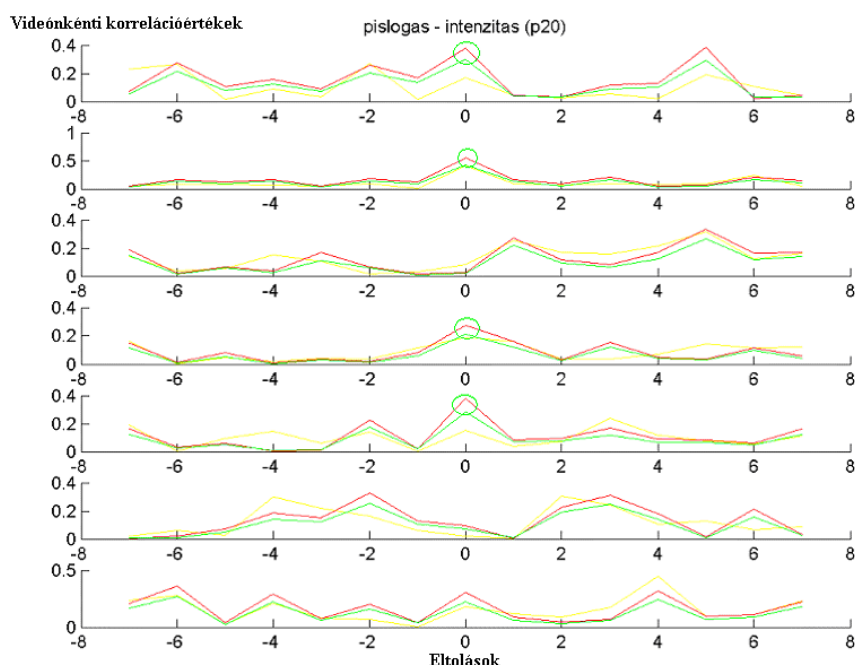
1. videó	2. videó	3. videó	4. videó	5. videó	6. videó	7. videó	8. videó	9. videó	
274.1082	316.5374	222.7985	574.5648	392.9824	365.0074	258.4099	343.0019	318.4712	hangmagasság maximum
0.1754	0.1000	0.2680	0.1280	0.2080	0.0600	0.0760	0.1196	0.0295	vízszintes mozgás átlaga
240.1537	235.5321	237.0172	242.2582	326.0556	264.8414	264.3274	387.1643	344.4409	bal szem mozgásának összege
256.9415	220.0779	243.2916	210.2337	326.4853	242.8973	258.5523	397.5516	343.0509	jobb szem mozgásának összege
3.4550	4.5480	2.9800	7.1240	7.0360	7.6000	5.0800	4.0000	5.5902	folyt. függ. mozgás átlaga
4.6981	5.7809	3.6250	8.9848	8.9660	8.0739	6.3806	5.1531	6.7192	folyt. függ. mozgás szórása

4. ábra. Folytonos függőleges mozgás és a hangmagasság vektorainak jellemzői 9 videó esetében

A három, a sorba nem illeszkedő videó közül kettőnél nagy a szemmozgások értéke, az előbbi esettel szemben pedig ezen felül az is megfigyelhető (8. sor), hogy a 3 videó közül kettőben az ablakolt vízszintes mozgások értéke az átlagosnál nagyobb (amely negatívan befolyásolhatja a tisztán függőleges mozgás vizsgálati eredményeit), egy esetében (3. oszlop) pedig attól kisebb. Az utóbbi videó esetében feltehetően az bír nagyobb jelentőséggel, hogy ez esetben a függőleges mozgás adataiban is a többi esetben tapasztaltnál kisebb értéket láthatunk. Ezen utóbbi felvétel esetében a hangmagasság maximum (első sor) is kisebb a többi esetben kapott értéknél.

5.3. Pislogás és intenzitás

Az előbbieken bemutatott módszer a következőkben egy további, jó eredményeket adó (20.) korrelációs pár esetében (pislogáskezdetek és intenzitás viszonya) egy videócsoport vizsgálati eredményeként látható (5. ábra). Az alkalmazott korrelációs párok és ezek megjelenítési színei a korábban leírtakkal egyezők. Az ábra ugyanazon 7 videó esetében mutatja az eredményeket, mint az 1. ábrán.



5. ábra. Pislogás kezdetek és intenzitás viszonya 7 videó esetében

A fenti ábrán jól látható, hogy a 0. eltolás a videók többségében kitüntetett szereppel bír, 7-ből 4 esetben maximumot, ebből 3 esetben pedig a többi eltoláshoz képest kiugróan magas értékű maximumot adva az adott videó korrelációs görbéjén, több esetben 0,4, ill. 0,5-öt közelítő korrelációs értékkel. A 7. ábrán 5-ből két esetben adja a korrelációs maximumot a 0. eltolás, egy esetben pedig ettől egy eltolásértékkal balra, a -1 . eltolásnál találjuk azt. Az 5. ábra görbéinek jellegét vizsgálva látható, hogy az adott műsorvezetővel készült összesen 7 videó esetében a 3., 6. és 7. felvétel görbéje tér el jellegében nagymértékben a többi esetben tapasztalttól.

A 6. ábra adatsorából azonban kiolvashatók az ezen videók esetében az átlagosnál nagyobb, illetve kisebb intenzitásértékek, az átlagosnál kevesebb pislogás, és mindezekon felül a függőleges mozgás mértékének átlagosnál nagyobb, illetve kisebb volta is.

1. videó	2. videó	3. videó	4. videó	5. videó	6. videó	7. videó	
0.2215	0.2312	0.2094	0.2455	0.2210	0.3585	0.1781	intenzitások maximuma
0.2215	0.2312	0.2094	0.2455	0.2210	0.3585	0.1781	intenzitás felső 10%-ok maximuma
16.0000	15.0000	15.0000	19.0000	20.0000	12.0000	9.0000	pislogások száma a videóban
0.1196	0.1410	0.0828	0.1017	0.1014	0.1111	0.0788	pislogások átlagos száma szótagonként
6.6678	3.7269	7.0448	3.7703	4.9507	7.8436	2.7143	függőleges mozgások átlaga
7.9669	4.4814	8.8939	4.6346	6.2059	9.9260	3.3302	függőleges mozgások szórása

6. ábra. Pislogás és intenzitás vizsgálati vektorainak jellemzői

6. Az eredmények összegzése, további feladatok

A folyamatos függőleges mozgás és intenzitás tekintetében egy adott műsorvezető szereplésével rögzített 7 videó alapján az a megállapítás tehető, hogy a -1 -es eltolás bír kitüntetett értékkel, amelynek jelentése, hogy a mozgás változása jellemzően egy szótaggal (szótagon belüli intenzitásmaximummal) előzte meg az intenzitás változását. A megfigyelésbe nem illeszkedő jellegű görbék esetében kimutatható a nagy szemmozgás mellett az intenzitásminimum-maximumok átlagának kisebb mértéke. Megjegyzendő, hogy az utóbbi ebben az esetben az intenzitásminimum-maximumokat kiemelő vektor tulajdonságait vizsgálva volt megfigyelhető. A vektorjellemzők vizsgálatával regisztrálható volt egy további megfigyelés is, miszerint az átlagostól nagyobb mozgás esetén az eddigiekhez hasonlóan kitüntetett maximummal rendelkező görbe volt regisztrálható, azonban az 5-ös eltolásérték maximumával, amelynek jelentése, hogy ez esetben az intenzitás változása 5 szótaggal (szótagon belüli intenzitásmaximummal) előzte meg jellemzően a mozgást.

Szintén a függőleges mozgást vizsgálva (azonban azt a pitch vektorával párosítva), egy másik műsorvezető esetében a -1 -es, illetve egy esetben a 0-s eltolás kitüntetett szerepe volt megfigyelhető, több esetben a 0,5-öt elérő korrelációs érték mellett. Ennek jelentése, hogy a mozgás változása jellemzően 1 szótaggal (szótagon belüli intenzitásmaximummal) előzte meg a hangmagasság változását, vagy egyidejű volt azzal. A többségtől jellegében eltérő görbék esetében az intenzitást tartalmazó pár vizsgálatánál tapasztaltakhoz hasonlóan kimutatható volt az átlagosnál nagyobb szemmozgás (illetve egy esetben az átlagosnál kisebb szemmozgás), továbbá a hangmagasság maximuma tekintetében volt kimutatható eltérés. Átlagosnál nagyobb szemmozgás esetében is kimutatható volt a többi eltoláshoz képest kiemelkedő korrelációs maximum, azonban ez esetben a 6-os, illetve -6 -os eltolásra eltolódva. Megjegyzendő továbbá, hogy az átlagosnál nagyobb szemmozgás ebben az esetben főként a vízszintes mozgást tekintve jelentkezett befolyásoló tényezőként, abban az esetben, amikor az ablakolt vízszintes mozgás átlaghoz viszonyított értéke alacsony volt, nem eredményezett eltérő jellegű görbét a nagy szemmozgás.

A pislogás és intenzitás viszonyának tekintetében kitüntetett szereppel a 0-s (illetve egy esetben a -1-es) eltolás bírt, több videó esetében meghaladva a 0,4-es korrelációs értéket. Ennek jelentése, hogy a pislogás jellemzően az adott szótagon belüli intenzitásmaximummal egy időben következett be (illetve egy videó esetében jellemzően 1 szótaggal hamarabb). Az adatokat megfigyelve tapasztalható volt, hogy az átlagtól kevesebb pislogás és ezzel együtt az intenzitásértékek átlagtól való nagymértékű eltérése eredményezett a többitől eltérő korrelációs görbét. További megfigyelés volt, hogy amennyiben a szemmozgások mértéke nem volt kiemelkedően nagy, a 0. eltolás lokális maximumai még az ilyen, a szabályrendszerbe nem illő esetekben is megmaradtak, azonban nem jelentettek globális maximumot.

A cikkben olvasható vizsgálatokkal sikerült igazolnunk bizonyos hangbeli jellemzők és fej által kifejezett gesztusok kapcsolatát, illetve megállapításokat tettünk arra, hogy egyes alanyok esetében milyen feltételek megléte esetén voltak nagyobb számban előforduló, egymáshoz hasonló minták, amely feltételek betartása elősegítheti a virtuális beszélő fej természetességét.

Ilyenek például az adott számú szótagra vagy időre eső pislogások minimális száma, függőleges irányú mozgások generálása esetén a vízszintes irányú mozgások alacsonyabb értéke, a függőleges mozgások adott értéktől nagyobb mértéke stb. Emellett az időbeli kapcsolatok, eltolódások jellemző mértékének meghatározása segítséget ad a kutatás részét képező automatikus gesztusgenerálás paramétereinek megadásához (felhasználandó ablakméretek, korábbi időpillanatok közül a kitüntetett szerepű meghatározása stb.).

A virtuális beszélő fej természetességének javulása tehát azáltal érhető el, hogy az ismertett eredmények által megadhatóvá válik, hogy a fej mozgása vagy pislogásainak száma adott időintervallumban minimálisan mekkora legyen, s az adott mértékű intenzitás-, illetve hangmagasság-változásokat mennyivel kövesse vagy előzze meg jellemzően nagyobb mozgás vagy pislogás. Az ismertett eredmények mindemellett az automatikus mozgásgenerálás paraméterezését is segítik. Célunk pedig mindezek felhasználásával az, hogy adott hanganyag megléte esetén, annak alapján a hallássérültek beszédfejlesztésében fontos szerepet játszó virtuális beszélő fej felhasználó számára természetes mozgása automatikusan generálható legyen.

A kutatómunka a Miskolci Egyetem stratégiai kutatási területén működő Mechatronikai és Logisztikai Kiválósági Központ keretében és a TÁMOP-4.2.2.C-11/1/KONV-2012-0002 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

Irodalom

- CZAP László–KILIK Roland 2014. *Gesztusok és hangsúlyok kapcsolatának változása az időbeli eltolás függvényében*. MicroCAD 2014, Miskolc.
- BUSO, Carlos–DENG, Zhihang– NARAYANAN, Shrikanth– NEUMANN, Ulrich 2007. Learning: expressive human-like head motion sequences from speech. *Data-Driven 3D Facial animation*. 113–131.
- RUBIN, P.– YEIHA, H.– VATIKIOTIS-BACON, E. 1998. Quantitative associaton of vocal-tract and facial behavior. *Speech Communication*. 26. 23–43.