

FEJMOZGÁS MÉRTÉKEK HANGANYAGBÓL TÖRTÉNŐ MEGÁLLAPÍTÁSÁNAK KEZDETI EREDMÉNYEI

Kilik Roland
tanszéki mérnök
Miskolci Egyetem
Geofizikai Tanszék

ABSZTRAKT

Az alábbi cikkben különböző alanyok spontán beszédében regisztrált függőleges fejmozgásai mértékének hangból történő megállapításának kezdeti eredményei olvashatók. Az eredmények bemutatását megelőzően a megállapításokhoz megalkotott mérőszámok és az osztályozási kategóriák is ismertetésre kerülnek, melyeket követően az egyes megfigyelések a figyelembe vett paraméterek hangolásával és azok optimális beállításával adódó jóságukkal egyidejűleg kerülnek bemutatásra.

BEVEZETÉS

A fent olvasható vizsgálatok kiindulási alapját egyrészt az szolgáltatta, hogy a témában folyó kutatások igazolták, hogy a fejmozgások és a hangból kinyerhető jellemzők között kapcsolat mutatható ki [1] [2] [3]. Mindemellett, neurális hálóval történő automatikus gesztusgenerálás [3] során elért eredmények kezdeti felhasználói minősítése során olyan megfigyelések születtek, miszerint például a hang amplitúdójából adódó szorzók a generált mozgás amplitúdójára alkalmazva pozitív hatással lehetnek a természetességre.

Közvetlenül a neurális hálók által előállított mozgásokon történő alkalmazás helyett azonban jobb, és szélesebb körű felhasználást is eredményezhető módszernek kínálkozott az eredeti mozgások mértékének hanginformációk alapján történő becslése. Ezen becslések különböző hangjellemzők, azok egymáshoz képesti viszonyai és határértékei által képzett általános, és specifikusabb szabályrendszerek összességéből épülnek fel. Mivel a kevés feltétellel rendelkező szabályok esetében a kutatás jelen fázisában is több igazoló minta érhető el, így ebben a cikkben ezek a szabályok kerülnek bemutatásra.

1. KAPCSOLÓDÓ KUTATÁSOK

Más, a témához kapcsolódó kutatások (például [1] és annak előzményei) az angol és japán nyelv, 1-2 alany, és nem spontán beszéd esetében tértek ki a fejmozgás mértékének becslésére. Emellett, a mozgás mértékét illetően nem állapítottak meg pontos becslést (a bemutatott mintában is adott esetben 2-10-szeres szorzó állapítható meg a becsült és tényleges érték között), és leginkább az arc egyes részei mozgásának megállapításában értek el jó – azonban ezek esetében is az adott hangtól függő – eredményeket. Ezzel szemben jelen kutatás magyar nyelv esetében, számos alanyra

spontán beszéd felhasználásával megalkotott tartalomfüggetlen szabályok okán általános érvényű megállapítások megalkotására törekszik.

2. A MINTÁK ELŐÁLLÍTÁSA

A szabályrendszerek megalkotásához 180 darab, különböző forrásokból (internetes videómegosztókon ill. műsortárakban elérhető tartalmak, és saját rögzítések) származó videóanyag került felhasználásra, melyek mindegyike egy mondatot tartalmazott. A mondatok jellemzően spontán megnyilatkozások, témájuk, hosszuk, alanyuk, illetve az alanyok neme és életkora változó volt. Egy alanytól maximálisan négy, jellemzően 2-3 mondat került felhasználásra.

A képanyagokból képkockánkénti mentést követően a maximális amplitúdó került rögzítésre, a felvételekhez tartozó hanganyagok pedig részben a Praat (mely a pitch, autokorrelációs, ill. keresztkorrelációs pitch mennyiségeket szolgáltat), részben pedig MATLAB szoftverrel kerültek feldolgozásra (az intenzitást, mozgóátlagolt amplitúdót, illetve minden ezekből és a pitch-ekből származtatott mennyiséget előállítva). Az intenzitás vektorának előállítására a következő formula szolgált:

$$\text{intensity}(i) = \sqrt{\sum(\text{ablak}^2)} \quad (1)$$

Tehát az intenzitásvektor minden eleme az ablak elemeinek négyzete szummájának négyzetgyökeként adódott, ahol az ablak rendre az amplitúdóvektor képkocka méretű szelete. Emellett előállításra került egy átlagolt amplitúdókat tartalmazó (amp) vektor is, melynek elemei az eredeti amplitúdóvektor minden 0.2 másodperces szeletén belül egy 11-es méretű csúszóablakkal végzett átlagolásával álltak elő:

$$\text{amp}_k = \frac{\sum_{k-r \leq j \leq k+r} |\text{wave}_j|}{2r+1}, \text{ ahol } r=11 \quad (2)$$

A kiinduló szabályok megalkotásakor a konstans 0.2 másodperces határok helyett a szótaghatárok voltak alkalmazottak, azonban az automatizálhatóság okán az ehhez leközelebbi eredményt adó konstans szegmentálás került bevezetésre (és az addigi eredmények eszerint átszámításra kerültek). A pitch vektorok előállításának a Praat szoftverben megadott időablaka minden esetben a képkocka időtartam volt.

3. A HANGANYAGBÓL ELŐÁLLÍTOTT MENNYISÉGEK

A korábbiakban ismertetett intenzitás, amplitúdó, illetve különböző pitch vektorok nem önmagukban és nem vektorként kerültek felhasználásra, hanem az azokból nyerhető különböző mennyiségekként. Ezen mennyiségek képzésének kiinduló alapját az adta, hogy a megnyilatkozások egyes csoportjainak mozgásmérete és a hanganyagaik intenzitásának varianciájának nagysága között összefüggés volt megfigyelhető.

A cikkben ismertetett kutatási fázisban az alábbi mennyiségek kerültek felhasználásra:

$$P1 = \overline{autokorrelációs_pitch} \quad (3)$$

$$P2 = var(FFT(intenzitás)) \quad (4)$$

$$P3 = \frac{\sum_{i=2}^N |amp_i - amp_{i-1}|}{képkockaszám} \quad (5)$$

$$P4 = \frac{\sum_{i=2}^N |intenzitás_i - intenzitás_{i-1}|}{képkockaszám} \quad (6)$$

$$P5 = \frac{\sum_{i=2}^N |pitch_i - pitch_{i-1}|}{képkockaszám} \quad (7)$$

$$P6 = \overline{keresztkorrelációs_pitch} \quad (8)$$

$$P7 = P3 * 50000 \quad (9)$$

$$P8 = var(FFT(amp)) \quad (10)$$

$$P9 = \frac{P3 * 50000}{P4^2 * 200} \quad (11)$$

$$P10 = var(amp) \quad (12)$$

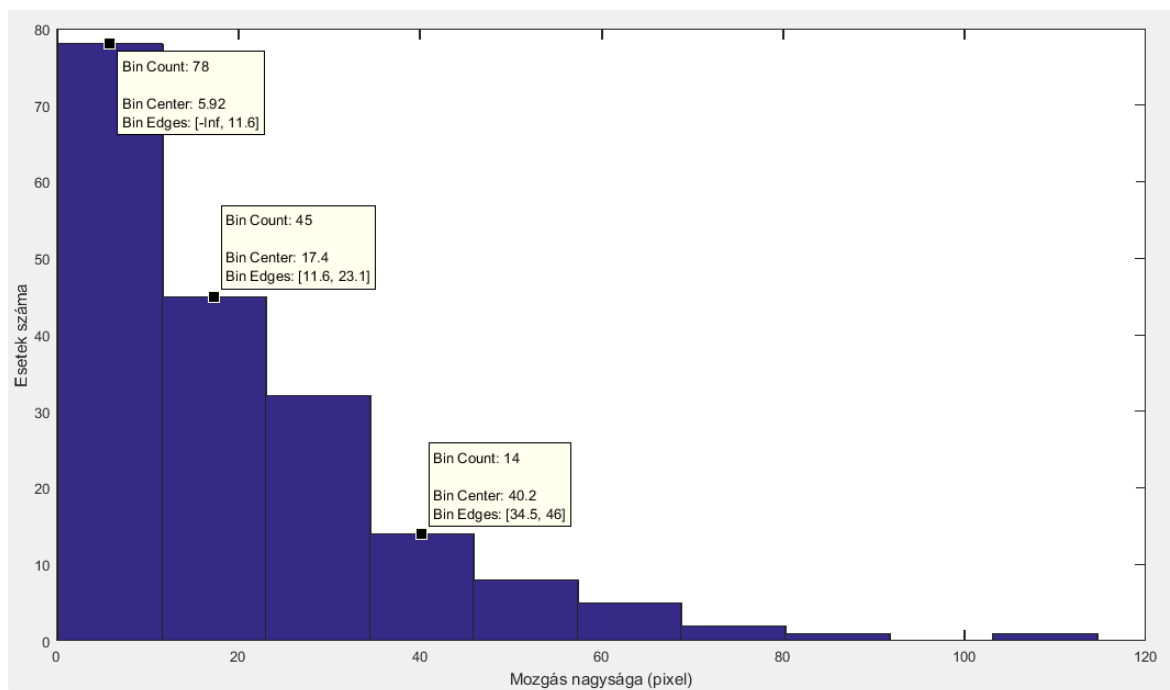
$$P11 = var(intenzitás) \quad (13)$$

$$P12 = szórás(intenzitás) \quad (14)$$

A 9. paraméter esetében elsőként a $P3*50000$, ill. a $P4*200$ részmennyiségek kerültek definiálásra. A két konstans azért került alkalmazásra, hogy a részmennyiségek több más (például $P1$, $P2$) mennyiséggel azonos értéktartományba kerüljenek. Mivel a két részmennyiség aránya később $P4$ -gyel osztásra került, így az eredő képletben négyzet szerepel. A 7. paraméter esetében a konstans érték megválasztásának oka a 9. paraméter esetében leírttal azonos.

4. A MOZÁS MÉRTÉKÉT BECSLÓ MEGFIGYELÉSEK

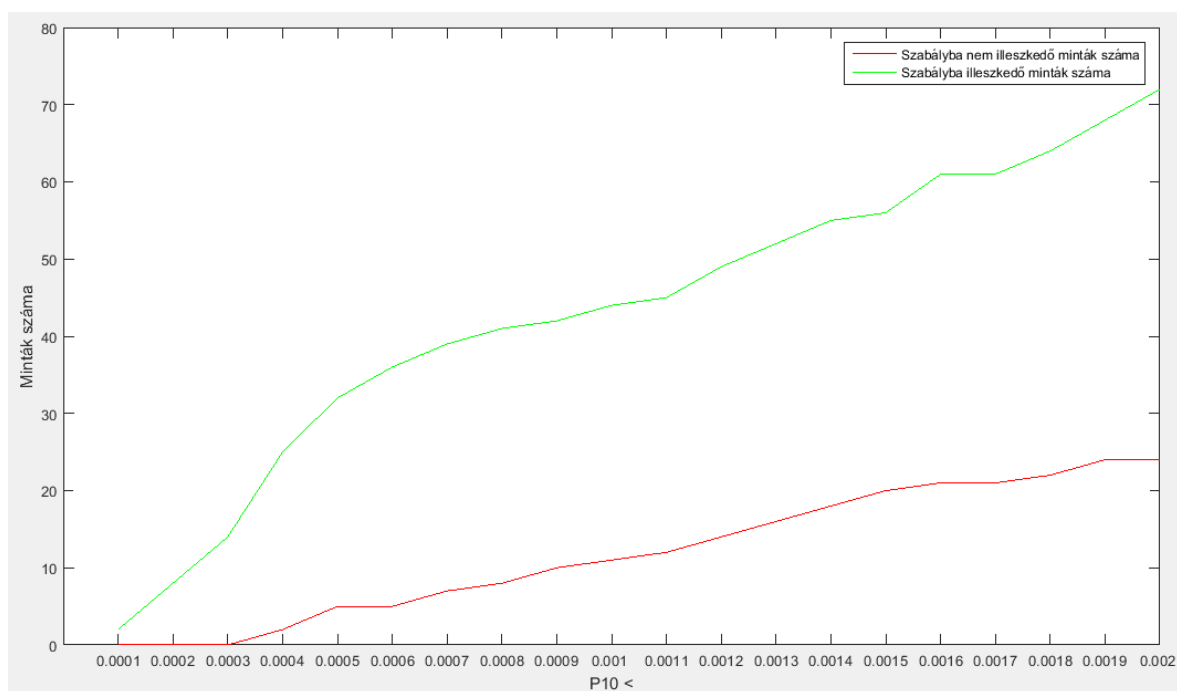
A felhasznált 180 minta mozgásainak mértékét jellemző hisztogram (1. ábra) indokolta azt, hogy a mozgásméret becslésekor – kerekítéssel – 10 pixel alatti, 25 pixel alatti, 25 pixel feletti, ill. 50 pixel feletti mozgásméretekre lettek felállítva a jelen publikációban ismertetett szabályok. A mozgások mértéke minden esetben 45 pixeles szemtávolságú fejméretre lett standardizálva.



1. ábra: Esetek darabszáma a mozgásnagyság függvényében

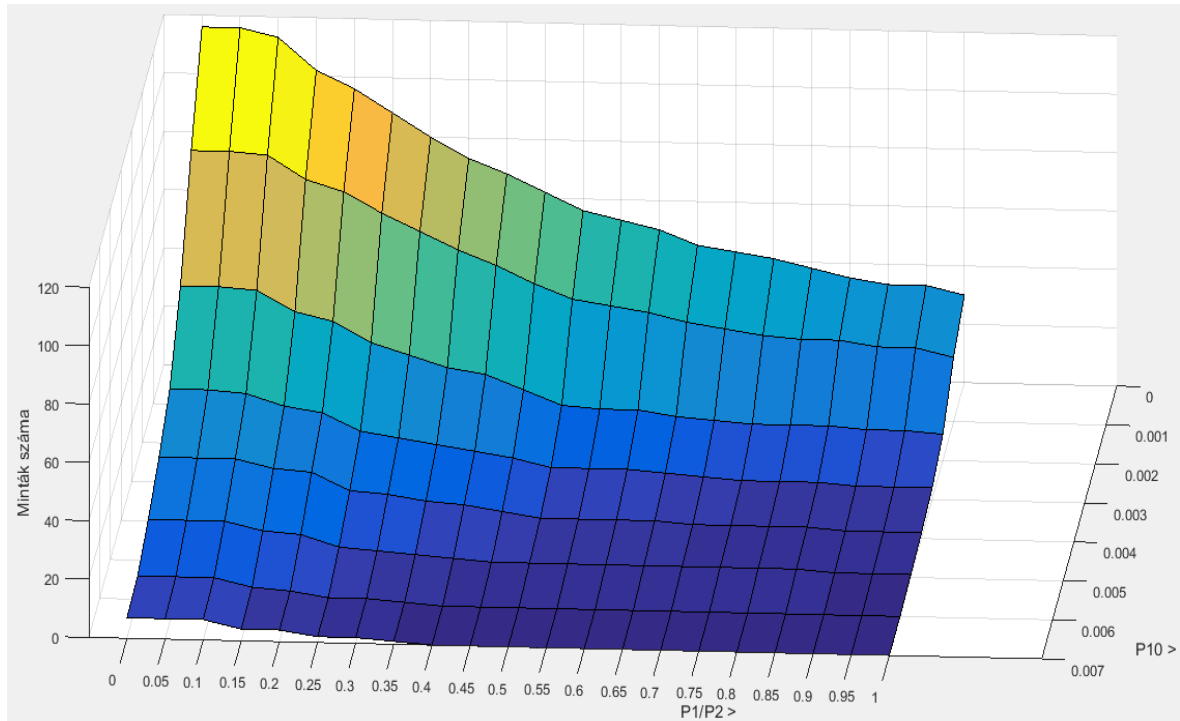
Az első megállapítás szerint, amennyiben a 10. paraméter értéke egy bizonyos határérték alatti, minden, a szabályba ilyen módon illeszkedő mondat mozgása 25 pixel alatti.

A második ábrából leolvasható, hogy a P10 0,0003 értéke alatt már nem volt a mintahalmazban a szabályba nem illeszkedő mozgású videó, illetve az is látható, hogy ezen – nem megfelelő mozgású esetet nem tartalmazó – tartományon belül ugyanezen küszöbérték alkalmazása esetén találhatók kedvező esetek a legnagyobb számban.



2. ábra: Az első szabály esetszámainak változása

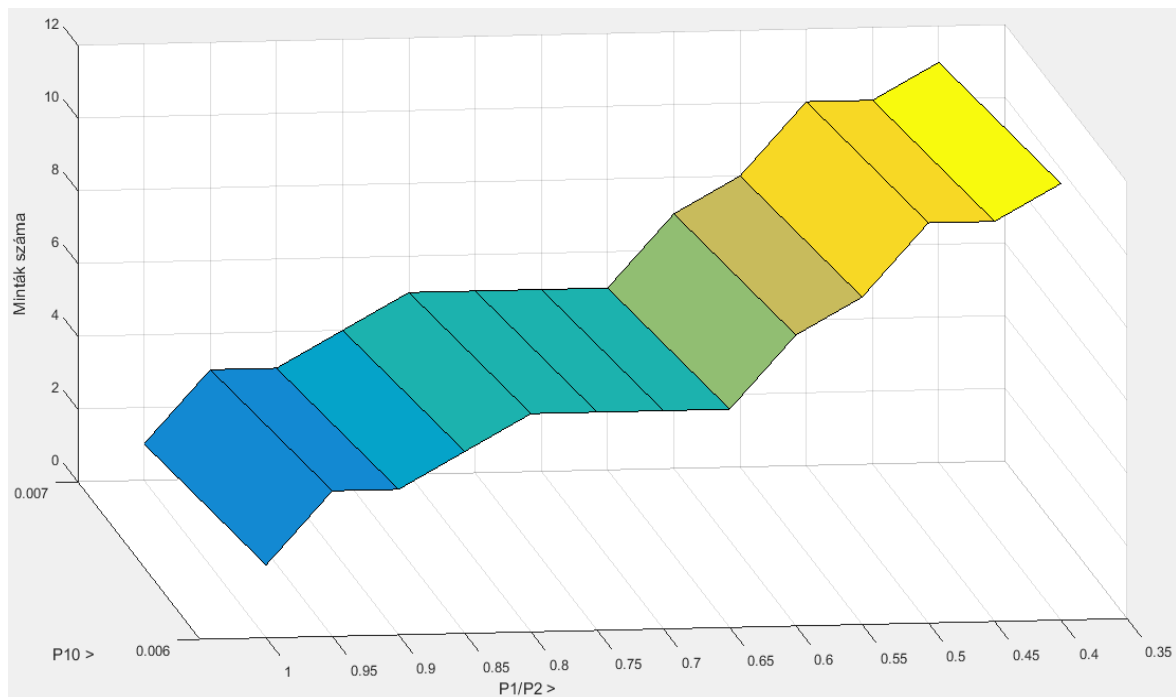
Az második szabály azt mondja ki, hogy a P1, P2 és P10 paraméterek megfelelő értékei mellett a mondatban regisztrálható mozgás mértéke biztos, hogy nem éri el a 10 pixelt.



3. ábra: A 2. szabálynak ellentmondó esetek számának változása

A harmadik ábrán a második szabályban figyelembe vett paraméterek különböző értékeinek függvényében a nem megfelelő mozgású esetek száma látható. Az ábrából megfigyelhető, hogy a P1/P2 paraméterarány 0,4 vagy nagyobb értéke esetén, amennyiben emellett a P10 paraméter értéke nagyobb, mint 0,006, nem regisztrálható ilyen eset.

Az alábbi (4.) ábrán a nem megfelelő mozgásmértékű esetek nulla számát eredményező paraméterbeállítások alkotta tartományban a kedvező esetek számának alakulása figyelhető meg. Az ábrából leolvasható, hogy a P10 paraméter nagyobb, mint 0,006 értéke mellett, a P1/P2 paraméterek 0,4-es aránya esetében kapjuk a legtöbb, 11 mintát, amely 10 pixeltől kisebb mozgással bír (0 ettől nagyobb mozgású minta mellett).



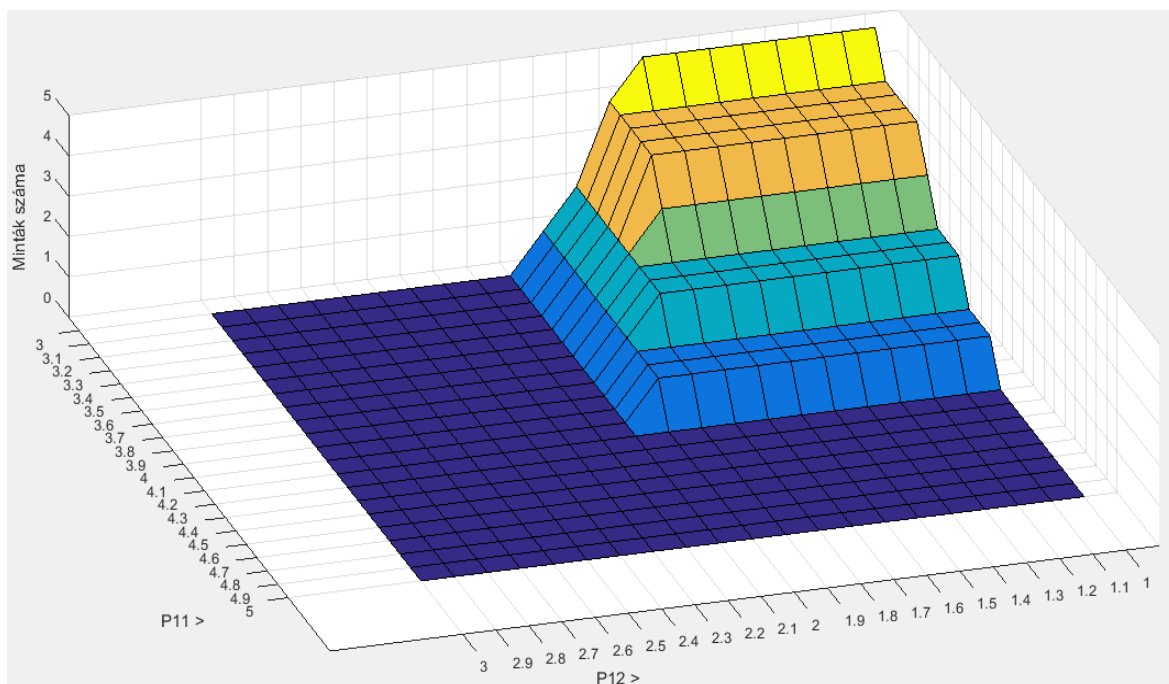
4. ábra: A 2. szabályba illeszkedő esetek számának alakulása

A harmadik szabály esetében, azt az első szabály esetében ismertetett módszer szerint (tekintve hogy az első szabályhoz hasonlóan, egyparaméteres esetről van szó) vizsgálva, az a megállapítás volt megfigyelhető, amely szerint az 5. paraméter nagyobb, mint 100 értéke esetén csak 25 pixeltől kisebb mozgás volt jelen a mintahalmazban. Ez a paraméterbeállítás egyben a maximális számú megfelelő mozgású esetet is eredményezte (7 minta).

A negyedik – szintén egyparaméteres – szabály esetében az előbbihez hasonló módon, azonban a 6. paraméter értéke esetében tehető meg ugyanilyen mozgásszabályra vonatkozó megfigyelés. Amennyiben a 7. paraméter értéke nagyobb, mint 180, mindössze három 25 pixel alatti mozgásnagyságú esetet figyelhetünk meg a mintahalmazban. Amennyiben a feltételen lazítva, a 7. paraméter értékeként csupán nagyobb, mint 170-et követelünk meg, az esetben 7 darab 25 pixeles mozgás alatti eset adódik, azonban ez esetben egy darab 25-től nagyobb mozgású is. Ez az eset azonban kizárható, ha nem engedjük meg, hogy a második paraméter értéke kisebb legyen, mint 200 (ez ugyanis több esetben eredményez nagy mozgást – ezen megfigyelés pontosítása még folyamatban van).

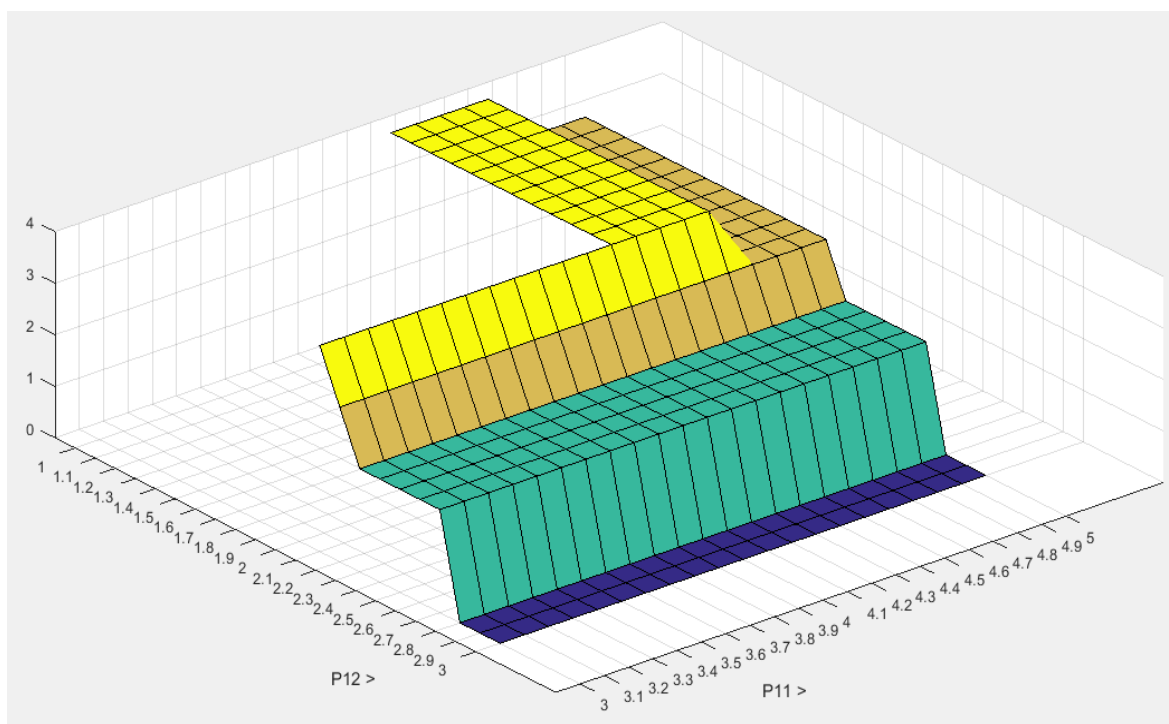
Az ötödik megállapítás szerint, a P11 és P12 paraméterek adott határértéket meghaladó értékei mellett a mondatban regisztrálható mozgás mértéke biztos, hogy nem éri el a 10 pixelt.

Az ötödik ábrán a megállapításban figyelembe vett paraméterek különböző értékeinek függvényében a nem megfelelő mozgásmértékű esetek száma látható. Az ábrából megfigyelhető, hogy amennyiben a P11 paraméter értéke nagyobb, mint 4,2, a P12-es paraméter 1-től nagyobb értéke mellett már nem regisztrálható ilyen mozgás, illetve hogy a P11 paraméter nagyobb, mint 3-as, és ezzel egyidejűleg a P12-es paraméter nagyobb, mint 2,1-es értéke mellett szintén nem volt 10 pixelt elérő mozgású eset a vizsgált mintahalmazban.



5. ábra: Az 5. szabálynak ellentmondó esetek számának alakulása

A következő (6.) ábráról a nem megfelelő mozgású esetek nulla számát eredményező paraméterbeállítások alkotta tartományban a kedvező esetek számának alakulása olvasható le.

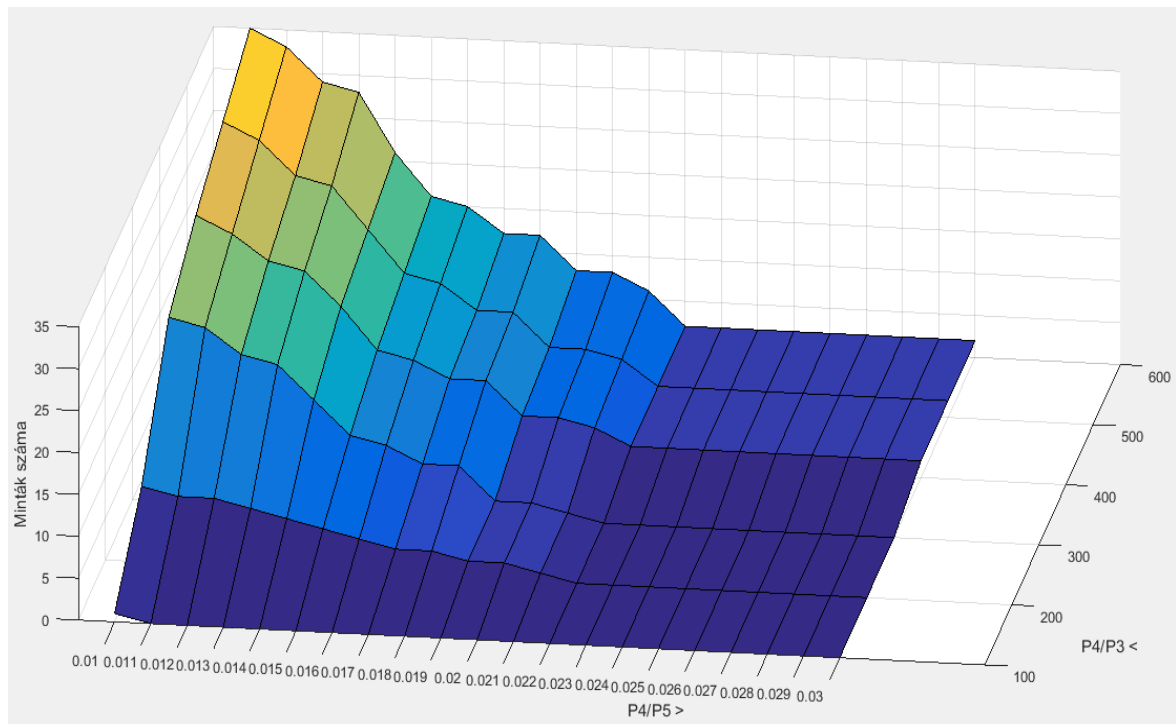


6. ábra: Az 5. szabályba illeszkedő esetek számának alakulása

Látható, hogy a korábban ismertetett, nulla nem megfelelő esetet eredményező határértékekre való beállítással egyben a maximális számú jó esetet eredményező paraméterbeállítást is kapjuk.

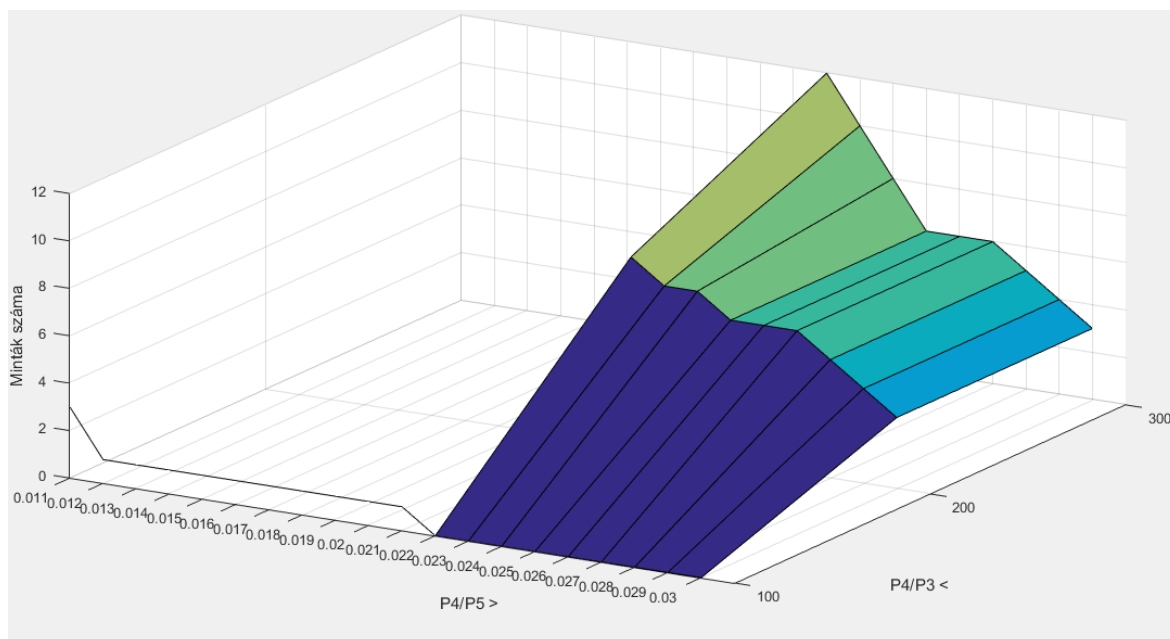
A hatodik megállapítás abban nyilvánul meg, hogy amennyiben a P4 és P3 paraméterek aránya kisebb egy adott értéknél, és ezzel egyidejűleg a P4 és P5 paraméterek aránya nagyobb egy határértéknél, csak 25 pixelnél kisebb mozgású mintát figyelhetünk meg.

A következő (7.) ábrán a szabályban figyelembe vett paraméterek különböző értékeinek függvényében a szabálytól eltérő mozgásmértékű esetek száma látható. Az ábrából megfigyelhető, hogy a P4/P5 paraméterarány 0,0022 feletti, és a P4/P3 paraméterarány kisebb, mint 200-as értékének együttes fennállása nem eredményezett ilyen esetet.



7. ábra: A 6. szabálynak ellentmondó esetek számának alakulása

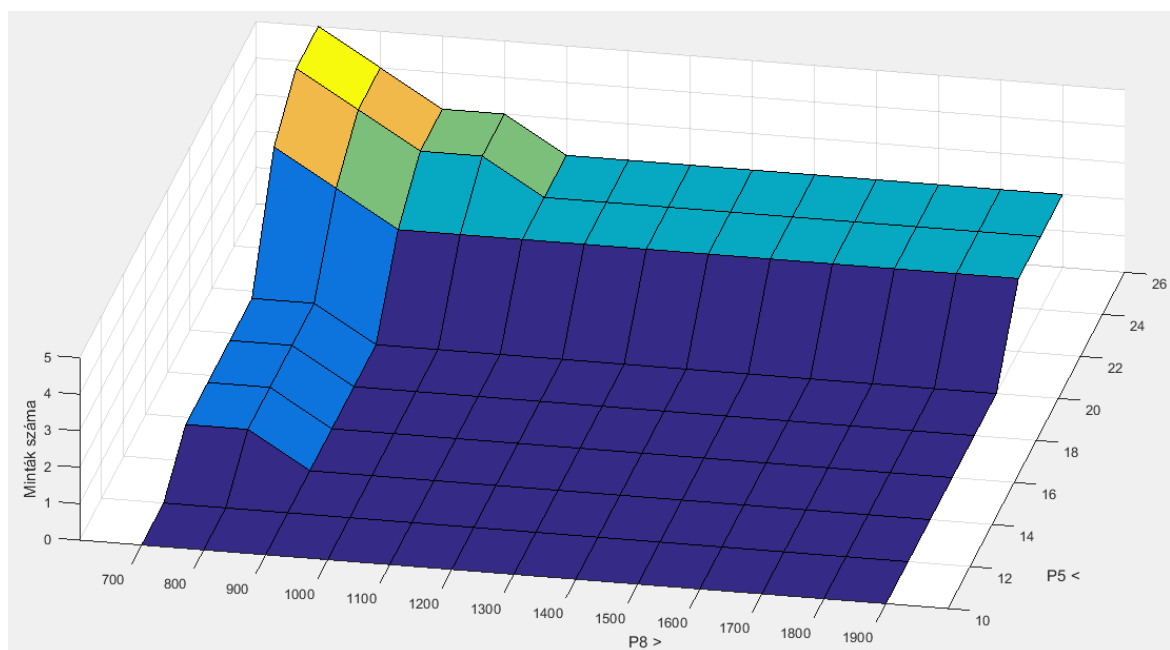
A következő (8.) ábráról a nem megfelelő mozgásmértékű esetek nulla számát eredményező paraméterbeállítások alkotta tartományban a kedvező esetek számának alakulása olvasható le.



8. ábra: A 6. szabályba illeszkedő esetek számának alakulása

Látható, hogy a szabályt alátámasztó maximális esetszám akkor volt regisztrálható (a nulla ennek ellentmondó esetet tartalmazó tartományban), amennyiben a 4. és 3. paraméter aránya 300-tól volt kisebb, és a 4. és 5. paraméter aránya 0,022-től nagyobb volt (12 eset). Emellett, amennyiben a 4. és 3. paraméter aránya 100-tól kisebb volt, amellyel egyidőben a 4. és 5. paraméter aránya nagyobb volt mint 0,011, további 3 kedvező mintát eredményezett.

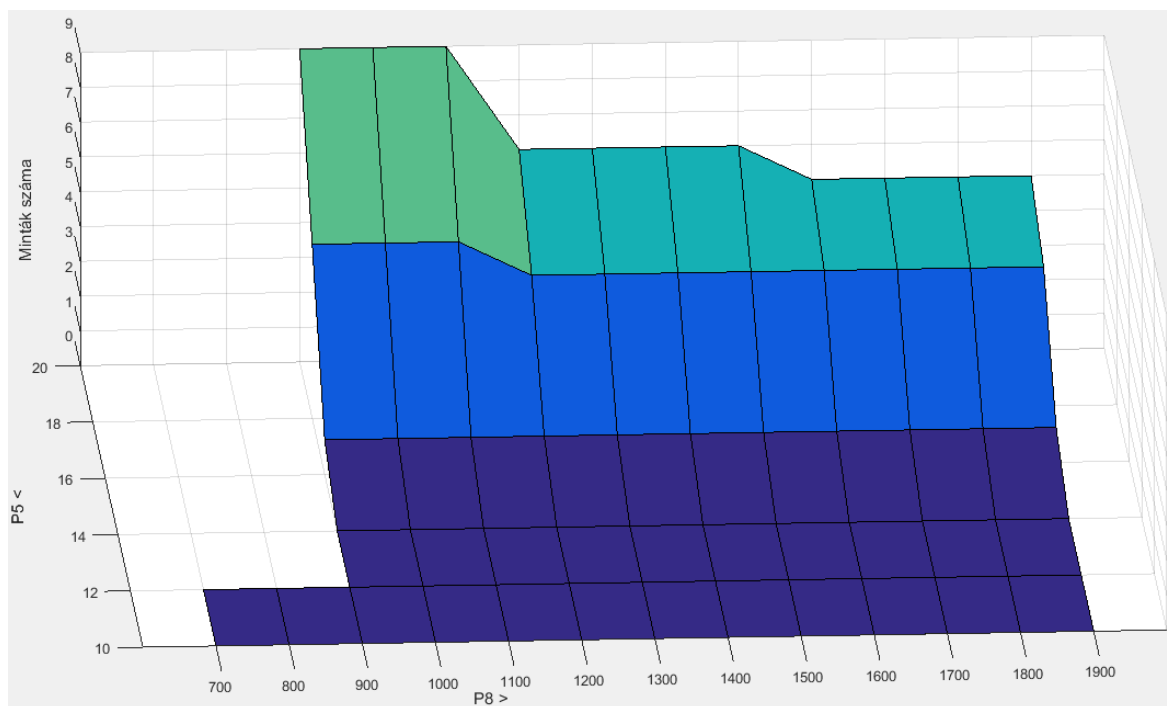
A hetedik megállapítás szerint, amennyiben a P5 paraméter értéke bizonyos határérték alatti, miközben a 8. paraméter értéke adott érték feletti, a mozgás értéke 25 pixel alatti minden mintára.



9. ábra: A 7. szabálynak ellentmondó esetek számának alakulása

A fenti (9.) ábráról leolvasható, hogy az 5. paraméter kisebb, mint 20-as értéke mellett, amennyiben a 8. paraméter értéke nagyobb, mint 900, vagy amennyiben az 5. paraméter értéke 12 alatti, és a 8 paraméter értéke nagyobb, mint 800, nem megfelelő mozgású eset nem regisztrálható a vizsgált mintahalmazban.

Az egyetlen rossz esetet sem tartalmazó tartományban a kedvező esetek számát bemutató alábbi (10.) ábrából pedig azt figyelhetjük meg, hogy a $P5 < 20$, és $P8 > 900$ paraméterkombinációnál található a legtöbb ilyen minta.



10. ábra: A 7. szabályba illeszkedő esetek számának alakulása

A 8. megállapítás szerint, amennyiben a P1 mennyiség értéke kisebb, mint 45, az esetben csak 25 pixel alatti mozgást tapasztalhatunk. Ez az állítás 9 mintapéldára teljesült, nulla ennek ellentmondó eset mellett.

A 9. megállapítás szerint, ha a P6 és P7 mennyiség különbségének abszolútértéke kisebb, mint 2, és P6 mennyiség értéke nagyobb, mint 100, vagy nagyobb, mint P7 értéke, 50 pixel feletti mozgást tapasztalhatunk. Ez az adatbázisban található videók közül négy esetében teljesült, nulla a megfigyelésnek ellentmondó eset mellett.

A 10. szabály a P9 mennyiséggel kapcsolatos, és azt a megállapítást tartalmazza, hogy amennyiben annak értéke nagyobb, mint 15, az esetben a mozgás mértéke legalább 50 pixel. Ezt a feltételt kielégítő minta azonban csak két esetben fordult elő az adatbázisban – bár 2 különböző alany szereplésével –, így ezen megállapítás későbbi esetleges megcáfolódásának esélye nagyobb, mint a többi esetben.

Az ismertetett megfigyelések az illeszkedő felvételeket tekintve nem diszjunkt halmazokat eredményeznek, azonban ha egy adott anyag több – egymásnak természetesen nem ellentmondó mozgásnagyságú – megállapítás feltételeit is kielégíti, az csak még biztosabbá teszi, hogy a videó ténylegesen az adott mozgáskategóriába tartozik. Az alábbi (1.) táblázatban a jelen publikációban ismertetett 10 megfigyelés legalább egyikébe illeszkedő videók esetében figyelhető meg, hogy a videó mely szabályokba tartozott bele. Az adott oszlopba tartozó

mintánál az adott megfigyelés sorában 1-es érték található ha az a videó annak esetei közé tartozik, és 0, ha nem. Az is megfigyelhető a táblázatan, hogy az ismertetett 10 szabály összesen 61 minta mozgásnagyságának megállapítására volt elegendő a rendelkezésre álló 180 közül.

		Minta sorszám																																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
Szabály (1-10)		0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0			
		1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1				
		0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0				
		1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0				
		0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0				
		0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0			
		0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0			
		0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Szabály (1-10)		32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61					
		0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1				
		0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0				
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0				
		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0			
		0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0			
		1	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0			
		0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				

1. táblázat: Szabályokba beletartozás/nem beletartozás jelzése legalább egy szabályba beletartozó videók esetében

ÖSSZEFOGLALÁS

A cikkben különböző, a beszédinformációból kinyerhető mennyiségeknek a beszéd közben végbemenő fejmozgások mértékének becslésére való alkalmazásának lehetősége került bemutatásra. Az egyes módszerek ismertetése példákon, és azokon keresztül a hatékonyság számszerűsítésével történt meg. A hatékonyságot tekintve összegzésként megállapítható, hogy az ismertetett megfigyelések a 180 mintából álló adatbázis 61 mintáját sorolták be a definiált osztályok valamelyikébe, a paraméterek ismertetett hangolását követően nulla rossz döntés mellett. A kutatás további céljai közé tartozik a döntésképeség arányának javítása további megfigyelések által, illetve a tesztminták számának növelése a megfigyelések pontosítására, illetve igazolására.

KÖSZÖNETNYILVÁNÍTÁS

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.

HIVATKOZÁSOK

- [1] YEHA, H; KURATATE, T.; VATIKIOTIS-BATESON, E.: Linking facial animation, head motion and speech acoustics, Journal of Phonetics, doi:10.1006/jpho.2002.0165, 2002., 555-568. old.
- [2] KURATATE, T.; YEHA, H; VATIKIOTIS-BATESON, E.: Kinematics-based synthesis of realistic talking faces, International Conference on Auditory-Visual Speech Processing (AVSP '98), 1998., 185-190. old.

- [3] CZAP, L.; KILIK, R. : Előzetes vizsgálatok automatikus gesztusgeneráláshoz, Alkalmazott Nyelvészeti Közlemények, Miskolci Egyetemi Kiadó, ISSN 1788-9979, 2015., 105-113. old.
- [4] CZAP, L.; KILIK, R. : Automatic Gesture Generation, Production Systems and Information Engineering ISSN 1785-1270, 7. kötet, 2015., 5-14. old.