

# Érzelmi tartalmak észlelése a magyar beszédben

Bartók Márton

ELTE BTK Nyelvtudományi Doktori Iskola,  
MTA-ELTE Lendület Lingvális Artikuláció Kutatócsoport  
bartokmarton@gmail.com

**Kivonat:** Érzelmi állapotaink kifejezésében nagy szerepe van a szupraszegmentumoknak. Az itt bemutatott kísérlet az érzelmi tartalmak ilyen módon történő, csak a beszéd nemnyelvi elemeit hasznosító közvetítésének észlelését vizsgálja. A kísérlet során eljátszott érzelmi állapotokat kifejező beszédhangokat játszottunk le 57 adatközlőnek, akik azokat az érzelmeket jellemző három, sematikus illusztrációkkal megjelenített, az érzelmi állapotok jellemzésében gyakran használt skálán (aktiváció, valencia, dominancia) értékelték. Az eredmények azt mutatják, hogy az adatközlők képesek a különböző aktivációjú, valenciájú és dominanciájú érzelmi tartalmak elkülönítésére kizárólag az adatközlő beszéde alapján az alkalmazott vizuális skála segítségével, azonban ennek mértékét több tényező is befolyásolja: mind a vizsgált érzelmi dimenzió, mind a stimulust produkáló beszélő személye befolyásolta az érzelmi állapotok elkülönítésének mértékét.

## 1 Bevezetés

Az emberi beszéd a nyelvi információn kívül számos nyelven kívüli, azaz paranyelvi tartalmat is hordoz. Ilyen nemnyelvi információ többek között a beszélő által kifejezett érzelem, melyet jellemzően a beszéd időzítési mintázatai, az intonáció és hangszín, azon belül is a hangszalagrezgés egyéb jellegzetességeit magába tömörítő fonációtípus milyensége segítségével vagyunk képesek kifejezni (Murray–Arnott 1993). Azzal, hogy ezek az érzelmi tartalmak milyen artikulációs-akusztikai változásokkal jelenítődnek meg, számos tanulmány foglalkozik, az itt bemutatott kutatás célja azonban a beszédben megjelenő érzelmi tartalmak észlelésének vizsgálata – az itt bemutatott kísérlet azt a kérdést járja körbe, hogy a beszédben kifejeződő érzelmi tartalmak különböző komponenseit hogyan dolgozzuk fel, és hogyan befolyásolják ezt a feldolgozási folyamatot különböző tényezők.

Az emberi érzelmek kategorizációját leíró tudományos modelleket kétfelé oszthatjuk: az alapérzelem-modellek (pl. Ekman 1992) kötött számú diszkrét, univerzális alapérzelmet feltételeznek (pl. düh, félelem, szomorúság, boldogság, undor és félelem), melyek egymással kombinálódva alkothatnak más érzelmeket. A dimenzionális érzelemelméleti modellek ezzel szemben az érzelmek kategorizációját és elrendeződését bizonyos dimenziók mentén gondolják el: azt feltételezik, hogy az emberi érzelmi

---

Bartók Márton: Érzelmi tartalmak észlelése a magyar beszédben. In Várad Tamás (szerkesztő), Ludányi Zsófia, Grácsi Tekla Etelka (szerkesztő): *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2019. XIII. Alkalmazott Nyelvészeti Doktoranduszkonferencia*. Budapest: MTA Nyelvtudományi Intézet. 2019. 34–44. DOI: 10.18135/Alknyelvdok.2019.13.3

állapotok sokféleségét két, három, vagy esetleg több kontinuum segítségével elégségesen jellemezhetjük. Russell (1980) szerint az érzelmek az érzelmi állapotban jellemző aktiváció és a valencia, tehát az érzelem pozitív-negatív jellege alapján jellemezhetők. Magas aktivációval jellemezhető például a félelem vagy a meglepődés, alacsony aktivációval az unalom és a nyugalom, ezeket az érzelempárokat pedig megkülönbözteti egymástól a valencia dimenziójában elfoglalt helyük (ti. az unalom negatív, a nyugalom pozitív valenciájú). Emellett több modell is tartalmazza az egyén adott állapotban érzett dominanciájának mértékét (pl. Mehrabian–Russell 1974), tehát lényegében azt az érzetet, hogy az egyén mennyire érzi képesnek magát környezete kontrollálására. Az itt bemutatott kutatásban azt vizsgálom, hogy az érzelmi állapotok beszédalapú kategorizációja hogyan működik ebben a három dimenzióban (aktiváció, valencia, dominancia). Azt vizsgálom tehát, hogy az érzelmi tartalmak észlelése során, amennyiben csak a beszédjel áll rendelkezésünkre, hogyan vagyunk képesek megkülönböztetni egymástól a különböző aktivációval, valenciával vagy éppen dominanciával jellemezhető érzelmi állapotokat. Az egyes érzelmeknek az ebben a háromdimenziós térben elfoglalt pozícióját Fontaine és mtsai. (2007) eredményei alapján határozom meg: ez a kutatás az érzelmi állapotok hat komponensét (az érzelmet kiváltó események értékelése, pszichofiziológiai változások, motoros kifejezések, cselekvési tendenciák, szubjektív érzések, érzelemszabályozás) megragadó 144 Likert-skálás kérdés varianciaanalízise alapján, közel 500 adatközlő részvételével helyezte el a különböző érzelmi állapotokat ebben a térben.

Az érzelmek észlelését vizsgáló kutatások általában eljátszott érzelmek észlelését vizsgálják, és arra fókuszálnak, hogy milyen arányban azonosítják helyesen az adatközlők a felvételen hallott érzelmi állapotokat – tehát kötött számú, a stimulusok között valóban megjelenő érzelmi állapot közül milyen arányban választják ki az adott stimuluson valóban kifejezni szándékozott érzelmi állapotot. Ezek alapján a kutatások alapján a megjelenített érzelmi tartalmat egészséges beszélők esetén a véletlen szintje fölött képesek vagyunk beazonosítani. Scherer (1989) harminc, jellemzően alapérzelmek észlelését vizsgáló tanulmány eredményei alapján átlagosan 60 százalékosra teszi az érzelmi állapotok helyes azonosításának arányát, ami valamivel kisebb pontosságra utal, mint amilyen arányban az érzelmeket közvetítő másik fontos tényező, az arckifejezések alapján vagyunk képesek azok azonosítására. Bár a legtöbb kutatásban profi színészek bemondásai alapján vizsgálják az érzelmi tartalmak észlelését, Jürgens és mtsai (2015) eredményei alapján az azonosítás pontosságára nem volt hatása annak, hogy színészek vagy laikus beszélők hangját használták stimulusként.

Bár a jelen kutatás során is eljátszott érzelmek beszédalapú észlelését vizsgáltam, a fentiekkel ellentétben azt kíséreltem meg leírni, hogy az érzelmi állapotokat jellemző dimenziók mentén hogyan jellemzik az adatközlők az észlelt érzelmi állapotot. Vizsgáltam egyrészt, hogy a hallott stimulusokat valóban úgy jellemzik-e az adatközlők a három vizsgált érzelmi dimenzióban, mint ahogy az adott stimulus produkciójakor eljátszott érzelem Fontaine és mtsai. (2007) eredményei alapján jellemezhető. A stimulusként használt beszédanyagon végzett akusztikai mérések eredményeit figyelembe véve azt is vizsgáltam továbbá, hogy mutat-e eltérést az érzelmi állapotok észlelése annak függvényében, hogy a beszélő mekkora változatossággal valószínűsíti meg az érzelmi kifejezést. Azt feltételeztem, hogy amennyiben a beszélő hangmintáin mért, a szakirodalom alapján az érzelmek kifejezésében relevánsnak tekinthető akusztikai paraméterek nagy változatosságot mutatnak a különböző érzelmek között, abban az esetben az egyén érzelmi állapotait nagyobb mértékben képesek vagyunk elkülöníteni, mint kisebb akusztikai változatossággal jellemezhető beszélő hangmintái esetén.

## 2 Módszertan

A kutatás tehát egy olyan percepciókísérletre épült, mely során az adatközlők különböző érzelmi töltetű beszédanyagokat minősítettek a kutatás során alkalmazott érzelmelméleti modell dimenziói (aktiváció, valencia, dominancia) szerint.

Mivel a kísérlet célja az érzelmi tartalmak észlelésének vizsgálata volt, így stimulusként olyan hanganyagokat használtam, melyek kizárólag a kifejezett érzellem tekintetében különböznek, a nyelvi forma minden esetben állandó volt. Ehhez 17 magyar anyanyelvű, 18–34 év közötti (Geifman et al. 2013 orvostudományilag elhatárolt korcsoportjai alapján meghatározott intervallum) nőt kértem arra, hogy produkáljanak kitarított [a:] beszédhangokat, különböző érzelmeket kifejezve. Az eljátszott érzelmeket a GRID projekt (Fontaine et al. 2013) 34 nyelven elérhető, visszafordítással validált érzelemlistájából, a Fontaine és mtsai. (2007) által feltárt dimenziótér minél teljesebb lefedésével állítottam össze. Az érzelmelek a következők voltak: lenézés, undor, düh, ingerültség, stressz, meglepődés, érdeklődés, boldogság, szeretet, csalódás, együttérzés, büntudat, szégyen, szomorúság, illetve egy érzelmileg semleges állapot (1. táblázat). A beszélők öt ismétlést produkáltak minden érzelemből, olyan módon, hogy az öt ismétlés egymást követte, azonban az érzelmelek sorrendje randomizálva volt. A vizsgált érzelmelek megoszlását a Fontaine és mtsai. (2007) által kialakított dimenziótérben az alábbi táblázat szemlélteti.

érzelem	aktiváció	valencia	dominancia
boldogság	magas	pozitív	domináns
csalódás	alacsony	negatív	nem domináns
düh	magas	negatív	domináns
együttérzés	alacsony	pozitív	domináns
érdeklődés	alacsony	pozitív	domináns
ingerültség	magas	negatív	domináns
lenézés	alacsony	negatív	domináns
meglepődés	magas	pozitív	nem domináns
stressz	magas	negatív	nem domináns
szégyen	magas	negatív	nem domináns
szeretet	magas	pozitív	nem domináns
szomorúság	alacsony	negatív	nem domináns
undor	alacsony	negatív	domináns

1. táblázat. Az aktiváció, valencia és dominancia értékei a vizsgált érzelmekekre Fontaine és mtsai. (2007) eredményei alapján

Ebből a 17 beszélőtől származó anyagból választottam ki a kísérletben használt stimulusnak három beszélőt arra törekedve, hogy ez a három beszélő eltérő akusztikai változatossággal valósítsa meg a különböző érzelmi tartalmakat: az volt tehát a cél, hogy az egyik beszélő csak kis mértékben, a második közepesen, a harmadik pedig nagyon változatosan fejezze ki a különböző érzelmekeket. A megfelelő beszélők kiválasztása érdekében az egyénre jellemző szisztematikus, érzelmemfüggő változatosságot az alábbi módon jellemeztem. Minden beszédhangon megmértem 12 akusztikai paramétert, melyek a hangszalagok működésének akusztikai következményét, a zöngét jellemzik, és amelyek nagymértékben különböznek egyes érzelmelek kifejezésekor (Patel et al. 2011). Ezek a paraméterek a következők voltak: jitter, shimmer, HNR,  $f_0$ , alpha

ratio,  $L_{eq}$ , az automatikus inverz szűréssel kinyert forrásjel amplitúdója, MFDR, NAQ, CQ, valamint H1-H2 az akusztikai és a forrásjel alapján. Az akusztikus mérések a Praat szoftver (Boersma–Weenink 2018), illetve a COVAREP szkriptcsomag (Degottex et al. 2014) használatával történtek. A beszédhangok teljes idején mért átlagértékekből az érzelmenként öt ismétléshez tartozó öt adatpontot beszélőnként átlagoltam, és az adott beszélő semleges állapotban produkált ejtésein mért értékek átlagához képesti százalékos eltérés formájában normalizáltam. Végül ennek a beszélőnként 168 értéknek (12 paraméter  $\times$  14 érzelem) vettem az átlagát, ami tehát egy, a beszélőre jellemző, a semlegestől való átlagos relatív eltérésként értelmezhető mérőszámot eredményez – ez a szám tehát az adott beszélőre jellemző általános érzelemfüggő akusztikai változatosságot számszerűsíti. A beszélőket sorba rendeztem e szerint a mérőszám szerint, és az akusztikai változatosság alsó kvartilisénél, mediánjánál és felső kvartilisénél található beszélőktől használtam beszédmintákat a percepciók kísérletben. Beszélőnként egy érzelemhez egyetlen megvalósulást, az ismétlések sorrendjében a középsőt választottam stimulusnak. Így végeredményben a semleges beszédmintákat is felhasználva 45 hangmintát használtam ((14 + 1 érzelem)  $\times$  3 beszélő). Az adatközlők minden stimulust csupán egy alkalommal hallottak, mivel az érzelmi tartalmak észlelése általában nagy igénybevételt jelent az adatközlők számára. Az ismétlések számának növelésével a jövőben fokozható lenne az eredmények megbízhatósága, azonban ezt feltehetőleg a felismerendő érzelmek számának csökkentésével kellene ellensúlyozni.

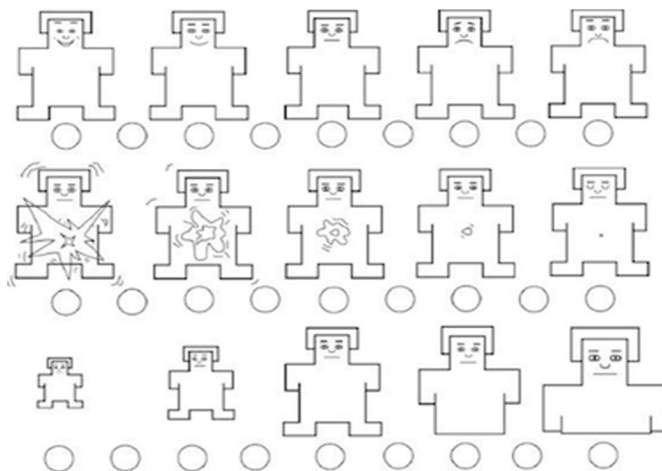
Mivel a kutatás során az érzelmi állapotok különböző dimenziók mentén történő kategorizációját vizsgáltam, olyan percepciók teszt használata volt szükséges, melyben az adatközlők valamilyen módon a vizsgált dimenziók mentén minősítik a hallott stimulus érzelmi tartalmát. Jelen kutatásban az úgynevezett Self-Assessment Manikin (Bradley–Lang 1994, a továbbiakban SAM) módszert alkalmaztam. A SAM egy három skálából felépülő vizuális értékelési rendszer (vö. 1. ábra), mely az általam vizsgált három dimenzió pontjait sematikus emberfigurákkal jeleníti meg. A valencia értékelése egy olyan skálán történik, melynek egyik végpontján a sematikus figura arckifejezése negatív érzelmeket tükröz (lefelé görbülő száj), míg a másik végponton pozitív érzelem jelenik meg (mosoly), a skála közbeeső pontjaihoz pedig a két arckifejezés közötti átmenetet megjelenítő illusztrációk tartoznak. Az aktivációhoz tartozó skálát egy relaxált (inaktív) és egy dinamikus, aktív figura közötti átmenet alkotja, míg a dominanciát mérő skálán a figura mérete változik: a nagyobb méret az egyén dominanciáját jeleníti meg a kisebb figurákkal szemben. A SAM előnye, hogy nem kötődik kultúraspecifikus verbális kategóriákhoz, ezért széles körben használható az érzelmi állapotok egységes jellemzésére. Adatközlőimet arra kértem, hogy a hallott ingereket 9 fokú SAM-skálán jellemezzék.

A percepciók tesztet online bonyolítottam le, a jsPsych könyvtár (Leeuw 2015) testreszabásával. Bár ezzel a módszertani megoldással kapcsolatban gyakran felmerül, hogy a hallgatás körülményei nem kontrollálhatók, Horton és mtsai. (2011), valamint Germine és mtsai. (2012) eredményei alapján az, hogy egy auditív percepciók kísérlet laboratóriumi vagy online lebonyolítással történik, nincs szignifikáns hatással a teszt eredményeire. A kísérlet elején egy szélsőségesen halk és egy szélsőségesen hangos megnyilatkozás alapján az adatközlőket arra kértem, hogy állítsák be készülékük hangerejét olyan módon, hogy kényelmesen hallják mindkét hangot, és ezen a beállításon a kísérlet során ne módosítsanak. Ezután néhány személyes adatra kérdeztem rá.

Ezt követően a SAM alsóskálájának bemutatása következett, a képi skálák mellett minimális verbális leírással magyarázva azok használatát (Bradley–Lang 1994 alapján). Az adatközlők a következő leírást kapták: „A következőkben arra foglak kérni, hogy

a hangminták alapján három képi skála segítségével jellemezze a felvételen hallható ember lelkiállapotát. Az első skála arra kérdez rá, mennyire pozitív (balra) vagy negatív (jobbra) a kifejezett érzelem. A második skálán azt jelöld, hogy mennyire érzed aktívnak (balra) vagy passzívnak (jobbra) a személyt a hangja alapján. Végül a harmadik skálán azt kell megjelölni, mennyire tűnik dominánsnak, irányítónak (jobbra) vagy éppen irányítottnak (balra) az épp hallott ember”. Ezután következett a 45 hangminta értékelése, randomizált sorrendben. Az adatközlők egy hangmintát többször is meghallgathattak, mielőtt jellemezték volna azokat. A kísérletben összesen 57 adatközlő vett részt, 49 nő és 8 férfi. Átlagéletkoruk 29 év volt.

A SAM alskáláin kapott értékeket adatközlőnként z-transzformáltam, hogy az egyéni válaszadási tendenciák hatását minimalizáljam. A Fontaine és mtsai. (2007) által meghatározott aktiváció, valencia és dominancia nominálissá alakított változók és a beszélőre jellemző akusztikai variabilitás szintje (szintén nominális, háromszintű változó) közötti, az adott dimenzióhoz tartozó SAM-alskála z-transzformált értékeire kifejtett interakciós hatások tesztelésére lineáris kevert modelleket alkalmaztam (Kuznetsova et al. 2015) az R szoftver segítségével (R 2018). A modellek az interakciós hatások mellett adatközlőnként random interceptet tartalmaztak.



1. ábra. 9 fokú SAM-skála. Az alskálák sorrendje felülről lefelé: valencia, aktiváció, dominancia

### 3 Eredmények

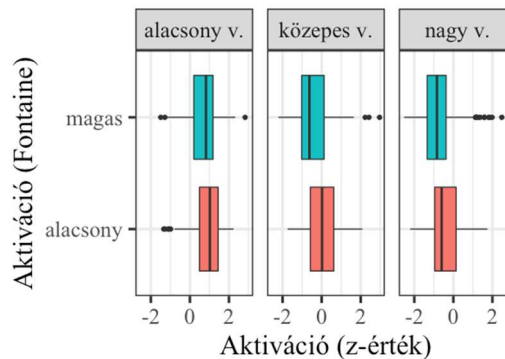
Az alábbiakban az egyes alskálákra vonatkozó eredményeket egyesével mutatom be. A SAM-skálák (z-transzformált) értékeinek értelmezésekor fontos figyelembe venni, hogy azok a vízszintesen tájolt vizuális skálák elrendezéséből kiindulva értelmezendők, melyeken az értékek mindig balról jobbra növekednek (1-től 9-ig), azonban az adott skálán megjelenített dimenzió tartalmilag ennek ellentmondhat – így például a magas aktivációt reprezentáló figura éppen a skála bal oldalán helyezkedik el. A könnyebb értelmezhetőség érdekében az eredményeket a vizuális skálákkal megegyező tájolású

dobozdiagrammokon ábrázoltam, minden esetben részletezve, hogy a skála mely végpontja milyen sematikus emberfigurával volt jelölve.

### 3.1 Aktiváció

Szignifikáns interakciót találtam a beszélőre jellemző akusztikai variabilitás és a fontaine-i aktiváció között ( $F[2, 2217] = 8,08, p < 0,001$ ) az aktiváció alskála z-transzformált értékeire (2. ábra). Az aktiváció vizuális SAM-alskáláján a magas aktivációs szintű, dinamikusabbnak ábrázolt figura a bal oldalon helyezkedik el, ennek megfelelően azt vártam, hogy ezekre az érzelmekre az adatközlők alacsonyabb értékeket adnak a Likert-skálán – ez így is történt. Ez a különbség továbbá várakozásaimnak megfelelően változik az adatközlőkre jellemző akusztikai variabilitás függvényében – a legkisebb eltérést a kis variabilitású beszélő esetén találtam, ugyanakkor a közepes és a nagy variabilitású beszélő stimulusain közel hasonló mértékben különböztették meg az alacsony és magas aktivációjú érzelmeket az adatközlők.

Ezen túl az is látszik, hogy nem csak a különböző aktivációjú érzelmek elkülöníthetősége változik beszélőnként, hanem az is, hogy általánosságban milyenek minősítik az adott beszélőt ebben a dimenzióban. Az alacsony variabilitású beszélő esetén az értékek az érzelmek alacsony vagy magas aktivációjától függetlenül általánosan magasabbak, tehát alacsonyabb aktivációjúnak (átlagosan 0,79,  $sd = 0,74$ ) vannak minősítve, mint a közepes (átlagosan  $-0,2$ ,  $sd = 0,86$ ), valamint a kifejezett érzelmi tartalomtól függetlenül általánosan is nagyobb aktivációval (kisebb értékekkel) jellemzett nagy variabilitású beszélő (átlagosan  $-0,58$ ,  $sd = 0,81$ ).

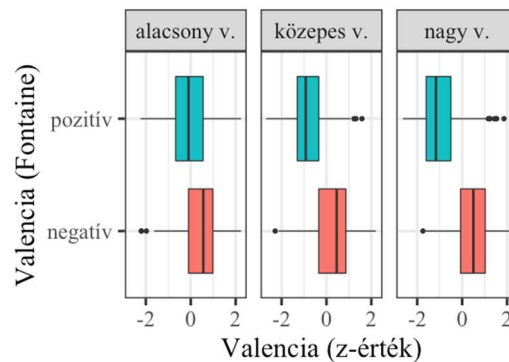


2. ábra. Az aktiváció alskála értékei az érzelmek aktivációjának Fontaine és mtsai. (2007) szerinti alacsony/magas kategorizációja és a beszélőre jellemző akusztikai variabilitás mértéke szerint. Az alacsonyabb értékek a vizuális skálán bal oldalon reprezentált magas aktivációjú állapotot jelentik.

### 3.2 Valencia

Szintén szignifikáns interakciós hatást találtam a beszélőre jellemző variabilitás és a fontaine-i valencia között ( $F[2, 2217] = 57,17, p < 0,001$ ) a valencia SAM-alskálájának z-transzformált értékeire (3. ábra). A vizuális skálán itt balra található a

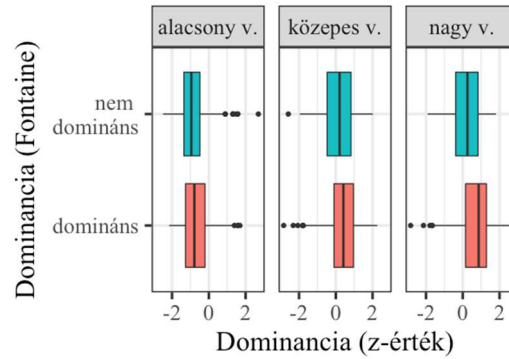
pozitív valenciát reprezentáló, mosolygós figura, tehát a pozitív érzelmek esetén alacsonyabb értékeket vártunk – és az adatok valóban ezt a tendenciát mutatják. Az eltérés a pozitív és a negatív érzelmek között fokozatosan nő a beszélőkre jellemző akusztikai variabilitás növekedésével. Az adatközlőre általánosan jellemző valencia nem mutat olyan fokozatos változást, mint amit az aktiváció esetében tapasztalhattunk, inkább az alacsony variabilitású beszélőre jellemző, inkább negatív valencia (nagyobb értékek, átlagosan 0,22,  $sd = 0,86$ ) áll szemben a közepes (átlagosan  $-0,12$ ,  $sd = 1$ ) és a nagy variabilitású beszélőre jellemző pozitívabb valenciájú minősítéseivel (átlagosan  $-0,1$ ,  $sd = 1,05$ ), azonban ezek a különbségek közel sem akkorák, mint az aktiváció esetén.



**3. ábra.** A valencia alskála értékei az érzelmek valenciájának Fontaine és mtsai. (2007) szerinti pozitív/negatív kategorizációja és a beszélőre jellemző akusztikai variabilitás mértéke szerint. Az alacsonyabb értékek a vizuális skálán bal oldalon reprezentált pozitív valenciájú állapotot jelentik.

### 3.3 Dominancia

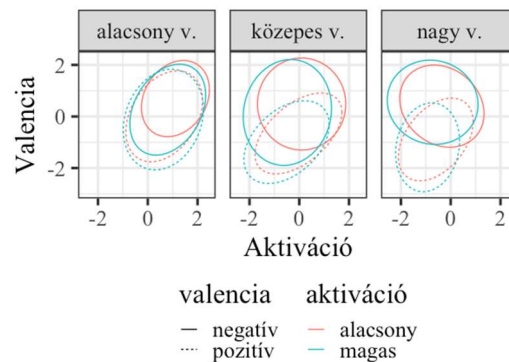
Szignifikáns interakciós hatást találtam a beszélőre jellemző akusztikai variabilitás és a dominancia mértéke között a dominancia alskálájának  $z$ -transzformált értékeire ( $F[2, 2217] = 9,48, p < 0,001$ ) (4. ábra). Ezen az alskálán a kisméretű figurával reprezentált nem domináns végpont balra helyezkedett el, így az ebbe a kategóriába tartozó érzelmek esetén várunk kisebb értékeket – és valóban, a Fontaine és mtsai. (2007) által nem dominánsnak kategorizált érzelmeket jellemzően kevésbé dominánsnak értékelték az adatközlők, mint a domináns kategóriába tartozókat. Ez az eltérés a kis akusztikai variabilitással jellemzett beszélő esetén még minimális, és fokozatosan nő a nagyobb variabilitású beszélő felé. Az aktiváció alskálájához hasonlóan az interakciós hatás nem csak annak eredménye, hogy az érzelmek két csoportja (domináns/nem domináns) közötti különbség növekszik a beszélőre jellemző variabilitás hatására, hanem a kifejezett érzelmetől függetlenül az adatközlők szerint a három beszélőre általánosan jellemző általános dominancia mértékében is találunk különbséget, mégpedig ismét a két kategória közötti eltérés nagyságával párhuzamosan: az alacsony variabilitású beszélő esetén a dominancia alacsony, tehát kevésbé dominánsnak ítélik a beszélőt (átlagosan  $-0,77$ ,  $sd = 0,76$ ), a közepes beszélőt ennél magasabb dominanciájúnak találják (átlagosan  $0,29$ ,  $sd = 0,85$ ), a legdominánsabbnak pedig a kifejezett érzelmetől függetlenül a legnagyobb akusztikai variabilitású beszélőt jellemezték (átlagosan  $0,48$ ,  $sd = 0,85$ ).



**4. ábra.** A dominancia alskála értékei az érzelmek dominanciájának Fontaine és mtsai. (2007) szerinti domináns/nem domináns kategorizációja és a beszélőre jellemző akusztikai variabilitás mértéke szerint. Az alacsonyabb értékek a vizuális skálán bal oldalon reprezentált nem domináns állapotot jelentik.

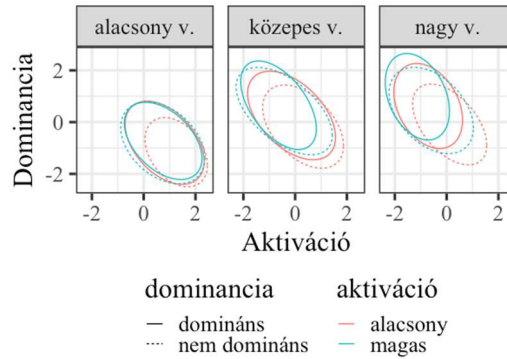
### 3.4 Összefoglalás

Eredményeink összességében azt mutatják, hogy adatközlőink a SAM mindhárom skáláján képesek voltak elkülöníteni a Fontaine és mtsai. (2007) alapján különböző aktivációjú, valenciájú és dominanciájú érzelmeket. Az alskálák együttes ábrázolásával jól látszik, hogy a legjobban az adatközlők a valencia alskálája mentén képesek differenciálni a különböző kategóriákba tartozó érzelmi állapotokat (5. és 7. ábra). Ezzel szemben a dominancia és az aktiváció alskáláin jóval kisebb az értékek eltérése a fontaine-i kategorizáció alapján elkülönített érzelmek csoportjai között, különösen a kis akusztikai variabilitással jellemzett beszélő esetén (6. ábra).

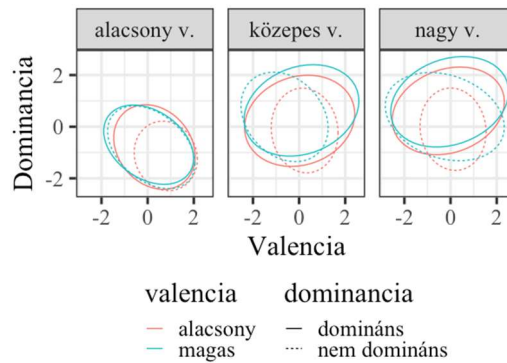


**5. ábra.** A valencia és aktiváció alskálák értékei 95%-os konfidenciaellipszisekkel ábrázolva, Fontaine és mtsai. (2007) alapján csoportosítva az érzelmeket





**6. ábra.** Az aktiváció és dominancia alszkálák értékei 95%-os konfidenciaellipszisekkel ábrázolva, Fontaine és mtsai. (2007) alapján csoportosítva az érzelmeket



**7. ábra.** A valencia és dominancia alszkálák értékei 95%-os konfidenciaellipszisekkel ábrázolva, Fontaine és mtsai. (2007) alapján csoportosítva az érzelmeket

#### 4 Következtetések

Kutatásom célja az érzelmi tartalmak beszédalapú észlelésének vizsgálata volt. Kutatásomban arra fókuszáltam, hogy három olyan dimenzió mentén, mely segítségével az érzelmeket gyakran jellemzi a szakirodalom (aktiváció, valencia, dominancia), mennyire vagyunk képesek elkülöníteni egymástól az e tekintetben különböző érzelmi állapotokat akkor, ha kizárólag a beszélő hangjával találkozunk, míg a nyelvi forma állandó. Emellett a kérdéssel együtt arra is kerestem a választ, hogy mennyiben befolyásolja az érzelmi tartalmak észlelését azoknak az akusztikai paramétereknek az érzelemfüggő változatossága, melyek a szakirodalom alapján akusztikai szempontból megbízhatóan különítik el egymástól a különböző érzelmi állapotokat – jelenti ez vajon azt is, hogy az észlelésben is nagyobb különbséget teszünk ezen hangminták között? 57 adatközlő vett részt abban a kísérletben, mely során 14 különböző érzelmi állapot eljátszását rögzítő beszédhangokat játszottam le, hogy azt a vizsgált három dimenzió-

ban jellemezzék. Minden érzelmet három beszélő megvalósításában hallottak az adatközlők. Ezt a három beszélőt egy nagyobb, összesen tizenhét beszélőt tartalmazó anyagból választottam ki úgy, hogy az akusztikai variabilitás eltérő szintjeit mutassák (Patel et al. 2011) alapján. Az adatközlők a stimulusokat a Self-Assessment Manikin (SAM, Bradley–Lang 1994) vizuális skáláin értékelték a három fent említett érzelmi dimenzió szempontjából.

Az eredmények azt mutatják, hogy az adatközlők képesek a különböző aktivációjú, valenciájú és dominanciájú érzelmi tartalmak elkülönítésére kizárólag az adatközlő beszéde alapján a SAM vizuális skáláinak segítségével, azonban ennek mértékét több tényező is befolyásolja. Egyrészt nagy különbség van az egyes alskálák között: leginkább az érzelmek valenciáját, tehát pozitív/negatív jellegét voltak képesek megkülönböztetni az adatközlők, míg az érzelem aktivációja és dominanciája jóval kevésbé különült el a válaszok alapján.

Nagyban befolyásolta az eredményeket az is, hogy milyen mértékű akusztikai variabilitást mutat az adatközlő – várakozásaimat alátámasztva a nagyobb akusztikai variabilitás esetén mindhárom alskálán jobban elkülönülnek egymástól a különböző kategóriákba sorolható érzelmek, azonban az akusztikai változatosság csökkenésével ezek a szembenállások egyre redukálódtak. A kismértékű akusztikai változatosságot mutató beszélő esetén már főként csak az érzelmi állapotok valenciája mentén voltak képesek azokat szétválasztani az adatközlők.

A beszélőre jellemző akusztikai variabilitás mértékével nem csak az változik, hogy milyen mértékben vagyunk képesek elkülöníteni egymástól a különböző érzelmi tartalmakat, hanem az is, hogy általánosságban (a kifejezett érzelemtől függetlenül) hogyan minősítjük az adott beszélő érzelmi állapotát a vizsgált dimenziókban. Az eredményekből az a tendencia rajzolódik ki, hogy a magasabb akusztikai variabilitást mutató adatközlőt általánosan aktívabbnak és nagyobb dominanciájúként jellemezték, mint a stimulusokat kisebb akusztikai variabilitással produkáló beszélőket. Ez az összefüggés arra utalhat, hogy az érzelmi tartalmak változatosabb kifejezése a beszédben a nagyobb (érzelmi) aktivitás, az extrovertáltság percepcióját is eredményezi, ami a társas viszonyrendszerekben esetleg nagyobb dominanciához is vezethet.

## Irodalom

- Boersma, P. – Weenink, D. 2018. PRAAT: Doing phonetics by computer. v 6.1. Elérhető: <http://www.fon.hum.uva.nl/praat/> (letöltés ideje: 2019. 02. 15.)
- Bradley, M. M. – Lang, P. J. 1994. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25(1): 49–59.
- Degottex, G. – Kane, J. – Drugman, T. – Raitio, T. – Scherer, S. 2014. COVAREP – A collaborative voice analysis repository for speech technologies. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 960–964.
- Ekman, P. 1992. An argument for basic emotions. *Cognition and Emotion* 6(3-4): 169–200.
- Fontaine, J. J. R. – Scherer, K. R. – Soriano, C. (szerk.) 2013. *Components of emotional meaning: A sourcebook*. New York, NY, US: Oxford University Press.
- Fontaine, J. J. R. – Scherer, K. R. – Roesch, E. B. – Ellsworth, P. C. 2007. The world of emotions is not two-dimensional. *Psychological Science* 18(12): 1050–1057.
- Geifman, N. – Cohen, R. – Rubin, E. 2013. Redefining meaningful age groups in the context of disease. *Age* 35(6): 2357–2366.
- Germine, L. – Nakayama, K. – Duchaine, B. C. – Chabris, C. F. – Chatterjee, G. – Wilmer, J. B. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review* 19(5): 847–857.

- Horton, J. J. – Rand, D. G. – Zeckhauser, R. J. 2011. The online laboratory: conducting experiments in a real labor market. *Experimental Economics* 14: 399–425.
- Jürgens, R. – Grass, A. – Drolet, M. – Fischer, J. 2015. Effect of Acting Experience on Emotion Expression and Recognition in Voice: Non-Actors Provide Better Stimuli than Expected. *Journal of Nonverbal Behavior* 39(3): 195–214.
- Kuznetsova, A. – Brockhoff, P. – H. B. Christensen, R. 2015. LmerTest: Tests in linear mixed effects models. *R Package Version 2*.
- Leeuw, J. R. de 2015. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* 47(1): 1–12.
- Mehrabian, A. – Russell, J. A. 1974. *An approach to environmental psychology*. Cambridge, MA, US: The MIT Press.
- Murray, I. R. – Arnott, J. L. 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America* 93(2): 1097–1108.
- Patel, S. – Scherer, K. R. – Björkner, E. – Sundberg, J. 2011. Mapping emotions into acoustic space: the role of voice production. *Biological Psychology* 87(1): 93–98.
- R, Core Team 2018. R: A language and environment for statistical computing. v. 3.5.3. Elérhető: <https://www.r-project.org>.
- Russell, J. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39(6): 1161–1178.
- Scherer, K. R. 1989. Vocal correlates of emotional arousal and affective disturbance. In: H. Wagner – A. Manstead (szerk.) *Handbook of social psychophysiology*. Wiley handbooks of psychophysiology. Oxford, England: John Wiley & Sons. 165–197.