

**A prozódia szerepe a több-beszélős szituációban történő
beszédfeldolgozásban**

Kovács Petra

Számítógépes és kognitív idegtudomány MSc

Témavezető:

Dr. Tóth Brigitta

Természettudományi Kutatóközpont, Kognitív Pszichológiai Intézet

Hang- és beszédészlelés kutatócsoport

2021 november

Tartalomjegyzék

1. Bevezetés	2
2. Irodalmi áttekintés	2
3. Módszerek	
3.1. Résztvevők	5
3.2. Ingeranyag	6
3.3. Eljárás	7
3.4. Adatelemzés	
3.4.1. <i>Viselkedéses adatok</i>	9
3.4.2. <i>EEG felvétel és előfeldolgozás</i>	9
3.4.3. <i>ERP elemzés</i>	10
3.4.4. <i>EEG forráslokalizáció és funkcionális konnektivitás</i>	10
3.5. Statisztikai elemzés	11
4. Eredmények	
4.1. Viselkedéses válaszok	12
4.2. ERP eredmények	13
4.3. EEG funkcionális konnektivitás	14
5. Diskusszió	16
5.1. A prozódia megváltozásának hatása a figyelmi szelekcióra és a hallási láncra bontásra	
5.1.1. <i>Viselkedéses hatások</i>	17
5.1.2. <i>ERP hatások</i>	17
5.1.3. <i>FC hatások</i>	18
5.2. Különböző prozódiai jegyek különböző hatással vannak a több-beszélős szituációban történő beszédfeldolgozásra	19
6. Összefoglalás	20
Hivatkozott irodalom	21
Függelék	25

1. Bevezetés

Az ember sok esetben látszólagos könnyedséggel veszi azt az akadályt, amikor számos beszélő egyidejű közlései közül kell az egyiket figyelemmel kísérnie. Pedig ha erre a jelenségre csupán az akusztika szintjéről tekintünk, beláthatjuk, hogy a különböző hangforrásokból érkező auditoros jelek nem külön-külön érkeznek a hallgató hallórendszerébe, hanem összeadódnak. A hallókészülékek és a videokonferencia-szoftverek például nem is tudnak egészen megbirkózni a feladattal: ezen technológiáknak nehézséget okoz feloldani több megszólaló “együttbeszélésének” problémáját. Az emberi hallgatók viszont képesek egymástól elkülönülő beszédflowamokat észlelni, ezért az inger és az észlelet között, a központi idegrendszer szintjén le kell zajlania olyan folyamatoknak, melyek ezt lehetővé teszik. A kérdés tehát, hogy melyek ezek a folyamatok.

A Hang- és beszédészlelés kutatócsoport korábban már számos olyan vizsgálatot végzett, melyben a több-beszélős szituációban történő beszédfeldolgozás került górcső alá. A beszéd tulajdonságai közül a feldolgozást befolyásoló tényezőként azonosították például a beszédtempót (Szalárdy et al., 2020), a neurális folyamatok oldaláról pedig felderítették, hogyan változnak az agyi funkcionális hálózatok a feladat (Tóth et al., 2019) vagy a legnagyobb feldolgozandó beszédegység (Tóth et al., 2020) függvényében.

A beszédben fontos információkat hordoz a “hanglejtés” vagy más néven prozódia (Myers et al., 2019). Ezért a jelen kutatásban azt vizsgáltuk, hogy a beszéd prozódiaja elősegíti-e egy beszédflowam feldolgozását egy másik beszédflowam jelenlétében, és hogy a prozódiai különbségek hatással vannak-e az agyi funkcionális hálózatokra a feldolgozás során. Az én feladatom az adatok feldolgozása és az eredmények értelmezésében való részvétel volt. Az adatfeldolgozás keretében az EEG-felvételek elektromos műtermékektől való tisztítását és forráslokalizációját, a funkcionális konnektivitás elemzését végeztem, valamint a viselkedési eredmények statisztikai elemzéséből és az adatvizualizációból vettem ki a részemet.

2. Irodalmi áttekintés

Több-beszélős környezetben történő beszédfeldolgozás során a hallgató előtt alapvetően két feladat áll. A Bregman (1994) által hallási színtérelmézésnek nevezett lépés során el kell különítenie, hogy mely hangok tartoznak ugyanazon hallási láncba, majd pedig szelektíven figyelnie kell ezek közül az egyikre (Broadbent, 1952). Ezek a folyamatok magasabbrendű kognitív funkciókat vesznek igénybe, ezáltal számos agyterület összehangolt működése

szükséges a végrehajtásukhoz (Hill & Miller, 2009; Saur et al., 2010). Emellett a beszéd akusztikai jellemzői is befolyásolják a láncra bontást és a szelektív figyelmet. Wöstmann és munkatársai (2017) hangsúlyozzák, hogy a beszéd spektrális szerkezetének torzítása a figyelmi szelekciót is nehezíti, míg a természetes spektrális tulajdonságokkal rendelkező beszédingerekre könnyebb figyelmet irányítani – akár akaratlanul is. Jelen tanulmányban a beszédingerek spektrális szerkezetét nem manipuláljuk, hanem az alapfrekvencia és beszédritmus változatosságát távolítjuk el, így a prozódia hatását vizsgáljuk a felvázolt folyamatokban.

Myers és munkatársai (2019) alátámasztják, hogy a prozódia lényeges szerepet tölt be a beszédfeldolgozás szempontjából: a hangsúlyozás, hangmagasság, időtartam és szünetezés mind elősegítik a hallott beszéd szavakra és frázisokra való bontását. Frazier és munkatársai (2006) kiemelik, hogy a prozódia feldolgozása automatikus, és tudatos kontroll nélkül végbemegy. Hipotézisük szerint azonban a beszédprodukciónban a prozódiai mintázatok használatának keretei kevésbé kötöttek, mint például a szintaktikai szabályszerűségek, így beszélők között előfordul némi variancia az információ prozódiai csoportosítását illetően. Alapvetően viszont a prozódia segít kijelölni a közlések szerkezetét. Számos kísérlet bizonyítja, hogy a tipikus prozódijú beszédet könnyebb megérteni, mint az atipikus prozódijút (Carlson, 2009; Kjelgaard & Speer, 1999; LaCroix et al., 2020; Roncaglia-Denissen et al., 2013; Sheppard et al., 2017). Ezért a tipikus prozódia valószínűleg elősegíti a kognitív kapacitások hatékony allokációját, a munkamemória működését, és a figyelem orientációját a beszéd lényeges részleteire.

A prozódia a beszéd több szupraszegmentális vonásából tevődik össze, melyek mind hozzájárulhatnak a figyelem hatékony irányításához. Az alapfrekvencia (F0) változásai és a szünetezés kijelölhetik a beszéd egységei közötti határokat. Emellett a beszédtempó hatással lehet a beszédértésre és az emlékezeti folyamatokra. Szalárdy és munkatársai (2020) kimutatták, hogy a gyors beszédtempó hatására romlik a célinger-detekció, az emlékezeti kódolás, vagy mindkettő. Összességében a prozódia valószínűleg hatással van a figyelmi szelekcióra több-beszélős szituációban, mert segít kijelölni a különböző feldolgozási egységeket.

A beszédfeldolgozást kísérő neurális aktivitást számos kutatásban vizsgálták elektroencefalogram/magnetoencefalogram (EEG/MEG) segítségével. Luo és Poeppel (2007) hívja fel a figyelmet a neurális nyomkövetés (*neural tracking*) jelenségére, mely során az alacsony frekvenciájú delta/theta oszcillációk (<8 Hz) fázisa folyamatosan leköveti a beszéd akusztikájának változásait. Több beszélő esetén pedig az alacsony frekvenciájú neurális

aktivitás minden egyes hallási láncot külön lekövet, ám a figyelt szöveget kiemelten reprezentálja az auditoros kéreg posterior része (Ding & Simon, 2012). Bár beszédfeldolgozás során számos fonetikai jellegzetesség egyidejű követése zajlik, Teoh, Capelloni és Lalor (2019) kimutatta, hogy az EEG-jelben elkülöníthető a prozódia egy jegyének a követése: a delta (0.2–4 Hz) frekvenciatartományba eső neurális oszcillációk fázisa követi a hangmagasság változásait (intonációt). Ez a hatás a beszéd spektrális szerkezetének torzítása esetén eltűnik. Ezzel a megállapítással összhangban vannak Rimmele és munkatársainak (2015) eredményei is, ahol a figyelt szöveg neurális nyomkövetése torzított akusztika esetén romlott, míg az ignorált szöveg követését nem befolyásolta, hogy torzított volt-e az akusztikája. Összességében tehát a neurális nyomkövetést fokozza a szelektív figyelem, de ez a beszédinger akusztikai minőségének a függvénye.

Az alacsony frekvenciájú neurális oszcillációk mellett az alfa (8–12 Hz) frekvenciatartományú aktivitás is fontos szerepet játszik a több-beszélős környezetben történő beszédfeldolgozásban. Az alfa oszcillációknak nem a fázisa, hanem a teljesítménye releváns ebből a szempontból. Ahogy Strauß, Wöstmann és Obleser (2014) is kiemeli, az alfa frekvenciatartományú hullámok teljesítményének növekedése az ignorálandó hallási lánc inhibíciójához köthető. Ez a teljesítménynövekedés top-down moduláció eredménye, amit egyrészt az bizonyít, hogy időben később megy végbe, mint a lassú hullámoknál megfigyelhető neurális nyomkövetés, másrészt pedig magasabbrendű folyamatokban szerepet játszó, frontoparietális agyterületekről ered (Fiedler et al., 2019; Wöstmann et al., 2016).

Ahogy korábban kiemeltem, a hallási színtérelmezés és a szelektív figyelem számos agyterület összehangolt aktivitásának az eredménye. Az egyik lehetőség ennek az összehangoltságnak a karakterizálására a funkcionális konnektivitás (*functional connectivity, FC*) megállapítása. Ezzel a módszerrel Obleser és munkatársai (2007) például kimutatták, hogy az akusztikailag torzított, de szemantikai szempontból bejósolható beszédingerek feldolgozása során nő a funkcionális konnektivitás bizonyos, az auditoros kérgen kívüli területek között, így a prefrontális, inferior frontális és posterior cinguláris kérgekben. Emellett megemlítendő, hogy a prozódiai információ feldolgozásának feladata már önmagában is számos agyterület között oszlik meg. Sammler és munkatársai (2015) egy kétutas modellt valószínűsítene relatív jobb féltekei lateralizációval: a dorzális pálya a premotoros kérget és az inferior frontális gyrust foglalná magába, a ventrális pálya pedig a superior temporális sulcus posterior és anterior területeit.

Arra azonban, hogy több-beszélős szituációban a figyelt és ignorált beszédingerek prozódiai jellegzetességei elősegítik-e a hallási láncra bontást, vagy inkább a figyelmi

szelekciót támogatják, még nem történtek vizsgálatok. Ezért a jelen tanulmányban az ingeranyag prozódiaját manipuláltuk háromféleképpen (F0 mesterséges kiátlagolása, természetes prozódia-visszafojtás, normál prozódia). A résztvevők egyik csoportjában csak a figyelt szöveget manipuláltuk, míg a másik csoportban csak az ignorált szöveget. A hallgatóknak minden esetben számneveket kellett detektálniuk a figyelt szövegben. Hipotéziseink szerint 1) a prozódiai jegyek lerontása hátrányos hatással lesz a figyelmi szelekcióra, míg a hallási láncra bontást kevésbé érinti, és 2) a természetes prozódia-manipuláció kevésbé lesz hatásos, mint a mesterséges manipuláció, mivel előbbi esetben valamivel több varianca marad meg az intonációs kontúrban.

Az első hipotézist az támasztaná alá, ha a viselkedéses és neurofiziológiai hatások nagyobbak lennének a figyelt szöveg manipulált csoportban, mint az ignorált szöveg manipulálása esetén. Funkcionális konnektivitás szempontjából ez azt jelentené, hogy a figyelmi szelekcióban résztvevő területeket a lassú (delta, theta; 0.5-8 Hz) oszcillációk erősebben kapcsolnák össze akkor, amikor a figyelt szöveg prozódiaja leromlik, hiszen ez arra utalna, hogy a szelekció megnehezült, több erőforrást igényel. Ezzel szemben ha az alfa oszcillációk kerülnek előtérbe, az a hallási láncra bontás nehézségét jelezné. Ha a hallási láncra bontásban jelentős szerepet játszik a prozódia, akkor nem számítana, hogy a figyelt vagy az ignorált szöveget manipuláljuk.

A második hipotézist az támasztaná, ha a hatások a mesterségesen manipulált prozódia kondícióban lennének nagyobbak, mint természetes prozódia-manipuláció esetén. Mivel az alfa teljesítmény növelésére annak függvényében van szükség, hogy mennyire vonzza magára a figyelmet az ignorálandó szöveg a figyelt szöveghez képest, ezért arra számítunk, hogy az alfa tartományú funkcionális hálózatok eltérő mértékű aktivitást fognak mutatni akkor, amikor egy normál prozódiajú szöveget egy mesterségesen manipulált illetve egy természetesen manipulált prozódiajú szöveg javára kell ignorálni.

3. Módszerek

3.1. Résztvevők

A kísérletben 50 magyar anyanyelvű fiatal felnőtt vett részt, akiket két kísérleti csoportba osztottunk: az első csoportban a figyelt szöveget manipuláltuk ($N = 25$; 18 nő, 7 férfi, $M = 21.2$ év, $SD = 1.4$ év, 24 jobbkezes), a második csoportnál az ignorált szöveget manipuláltuk ($N = 25$; 19 nő, 6 férfi, $M = 21.9$ év, $SD = 2.5$ év, 21 jobbkezes). A nem manipulált oldalon minden esetben normál prozódiajú beszéd hangzott el.

A résztvevők ép hallással rendelkeztek, és egyikük sem számolt be pszichiátriai vagy neurológiai kórelőzményről. A részvétel önkéntes volt, a kísérletet az Egyesített Pszichológiai Kutatási Etikai Bizottság (EPKEB) hagyta jóvá.

3.2. Ingeranyag

A résztvevők minden blokkban kettő, egyszerre bemutatott beszédanyagot hallottak, melyek körülbelül 6 perc hosszúságúak voltak. Az elhangzott szövegek magyar nyelvű híroldalokról származtak, és érzelmileg semleges, kevésbé ismert tényeket tartalmaztak. Az ingeranyag felvétele során a szövegeket két különböző magyar anyanyelvű férfi színész olvasta föl. A hangfelvételek ugyanabban a laborban készültek, ahol később maga a kísérlet folyt, és a beszélők minden felvételnél ugyanazon a helyen álltak, ahol később az adott felvételt lejátszó hangszóró helyezkedett el. Ezáltal a hangfelvételek a beszélő elhelyezkedését illetően valószínű hangzással rendelkeztek.

Az ingeranyag prosódiajának manipulálásával három kondíciót hoztunk létre:

1. **Mesterségesen lapított prosódia:** A Keitel et al. (2013) által alkalmazott módszert követtük, miszerint a Praat szoftver segítségével (Boersma & Weenink, 2005) eltávolítottuk a hangfelvételek intonációs kontúrját, és helyettesítettük azt a felvétel átlagos alaphangfrekvenciájával.
2. **Természetesen visszafojtott prosódia:** A színészeket a felvétel során arra kértük, hogy monoton intonációval olvassák föl a szövegeket, hasonlóan ahhoz, mintha csupán egy szólistát olvasnának.
3. **Normál prosódia:** Nem alkalmaztunk manipulációt, a színészek a felvétel során természetes intonációval beszéltek.

A kondíciók intonációs kontúrjai közti különbséget az 1B ábra is szemlélteti. Ahhoz, hogy jellemezzük a mesterséges és természetes manipuláció közti különbséget, két-két minta-hangfelvételt akusztikai elemzéseknek vetettünk alá a Praat szoftver segítségével (Boersma & Weenink, 2005). Az alaphangfrekvenciát és a beszédtempót karakterizáltuk. A leíró adatokat az 1. táblázat szemlélteti. Alapvetően az mondható el, hogy az F0 varianciája a mesterségesen lapított kondícióban jóval kisebb, mint a természetesen prosódia-visszafojtott kondícióban, míg utóbbi esetben a beszédtempó változott meg leginkább. Megjegyzendő, hogy a beszédtempó gyorsulását nemcsak a gyorsabb artikuláció okozhatja, hanem a rövidebb vagy teljesen hiányzó beszédsszünetek.

		Mesterségesen lapított prozódia	Természetesen visszafojtott prozódia	Normál prozódia
F0 (Hz)	Md	147.41	117.075	131.185
	SD	3.876	14.625	35.6
	IQR	0.675	21.16	61.87
	Terjedelem	180.45	198.25	224
Beszédtempó (szótag/s)	M	4.81	5.26	4.81
	SD	0.45	0.36	0.45

1. táblázat. Az ingerek prozodiához köthető akusztikai paraméterei kondícióként.

Minden szöveg tartalmazott számneveket, amelyek a figyelt szövegekben célingerként, az ignorált szövegekben pedig disztraktorként szolgáltak. Emellett az ignorált oldalon elhangzó szövegek (és csak ezek) szintaktikai sértéseket is tartalmaztak abból a célból, hogy megállapítsuk, hogy az ignorálandó információk bizonyos kondíciókban tudatos feldolgozásra kerülnek-e. A tudatosan feldolgozott szintaktikai sértések ugyanis kiváltják az N400 ERP-választ (Kutas & Hillyard, 1983), míg az ignoráltak nem (Szalárdy et al., 2018).

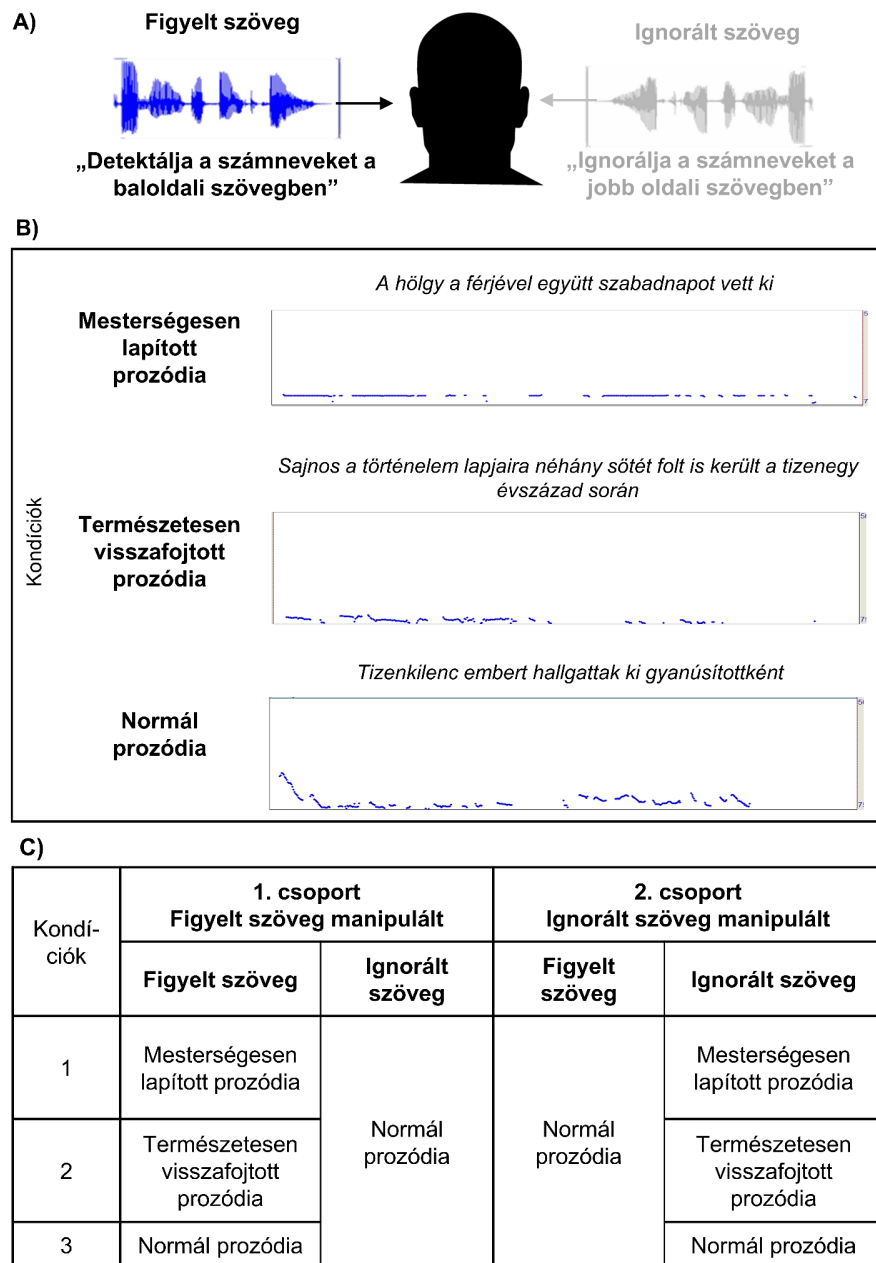
3.3. Eljárás

A kísérlet során az ingeranyag két hangszórón keresztül került bemutatásra, az egyik a résztvevő balján, a másik a jobbán helyezkedett el. A kísérlet alatt a résztvevőnek egy monitoron elhelyezett fixációs kereszten kellett tartania a tekintetét, hogy minél több szem- és fejmozgást szűrjünk ki. A résztvevő feladata az volt, hogy a balról érkező szövegben detektálja a számneveket, és ezt gombnyomással jelezze, miközben a jobbról érkező szöveget figyelmen kívül hagyja. A két beszélőnek állandó helye volt, vagyis a figyelt (bal) oldalról mindig az egyik beszélő hangja érkezett, az ignorált (jobb) oldalról mindig a másik beszélőé. A figyelendő beszélő személyének és elhelyezkedésének állandósága segítette a hallgatót a figyelt szöveg követésében.

A résztvevők minden blokk után egy emlékezeti tesztet kaptak arról a szövegről, amely abban a blokkban hangzott el a figyelt oldalon. Minden teszt 5 feleletválasztós kérdésből állt, és a résztvevőnek jeleznie kellett, ha a választ nem az iménti szövegből jegyezte meg, hanem más forrásból tudta. Ebben az esetben a választ figyelmen kívül hagytuk. A teszt kérdéseit a kísérletvezető olvasta föl a résztvevőnek, aki szóban válaszolt.

A résztvevők mind a három kísérleti kondícióra 4-4 szövegpárt hallottak külön blokkokban, így a kísérlet 12 blokkból állt (3 kondíció * 4 blokk). A blokkok kondíció

alapján pszeudorandom sorrendben követték egymást, azzal a megkötéssel, hogy két egymást követő blokk nem tartozhat egyazon kondícióhoz. A 6. blokk után minden esetben szünet következett, a többi blokk között pedig a résztvevő kérése szerint tarthatott pihenőt. A kísérleti elrendezést az 1. ábra szemlélteti.



1. ábra. Kísérleti elrendezés. A) A résztvevők feladata volt a figyelt oldalon elhangzó számnevek detektálása, és a másik oldal ignorálása. B) Az ingeranyag három kondíciója egy-egy példamondat F0-kontúrájával. C) Kondíciók a figyelt szöveg manipulált és az ignorált szöveg manipulált csoportban.

3.4. Adatelemzés

3.4.1. Viselkedéses adatok

A kísérletben két viselkedéses feladat volt: egy számnév-detekciós és egy felismerési emlékezeti feladat. Az emlékezeti feladatnál a helyes válaszok arányát számítottuk ki az egyes kondíciókban, míg a számnév-detekció esetén négyféle különböző változót vettünk föl. A *találatok* megállapításakor korábbi tanulmányaink protokollját követtük (pl. Tóth et al., 2019), miszerint a célinger (figyelt szövegben elhangzott számnév) kezdetétől mért 5000 ms-os időablakban történt válaszadások közül azokat tekintettük találatnak, melyek az összes válaszadás 5% és 95% közötti időintervallumába estek. Így kiszűrtük a túl gyors és a túl lassú válaszokat. A találatok tényleges időablaka így 405-1689 ms volt.

A *téves riasztásokat* (false alarm, FA) megkülönböztettük a *disztraktor-hatástól*. Ez utóbbit úgy definiáltuk mint az ignorált szövegben megjelenő számnevekre adott válaszok, míg FA-nak tekintettünk minden, nem számnévre adott választ. A kettőt jelen vizsgálat kontextusában fontos külön tárgyalni, hiszen a disztraktorok által előidézett gombnyomások a szelektív figyelmi folyamatoknak a sikertelenségére utalnak. A disztraktor-hatás időablakának meghatározásakor is a találatoknál leírtaknak megfelelően jártunk el.

Hogy elkerüljük a résztvevők esetleges torzító válaszadási taktikáinak a hatásait (pl. aki minden számnévre gombot nyom, annak 100%-os lesz a találati aránya, de ugyanígy a disztraktor-hatás is; Green & Swets, 1988), d' -ot is számítottunk a z-transzformált találatokból és a disztraktor hatásból a következőképpen: $d' = z(\text{TALÁLATOK}) - z(\text{DISZTRAKTOR})$.

3.4.2. EEG felvétel és előfeldolgozás

Az EEG adatok felvétele 64 csatornás BrainAmp DC EEG rendszerrel történt. Az adatfelvétel pár másodperccel a szövegpár kezdete előtt indult el, és a rövidebb szöveg vége után pár másodperccel zárult. Az elektródákat a nemzetközi 10/20-as rendszer szerint helyeztük föl, referencia-elektrodaként pedig az FCz csatorna szolgált. A mintavételi frekvencia 1 kHz volt.

Az EEG adatok előfeldolgozását az EGGlab szoftverrel végeztem (Delorme, Sejnowski & Makeig, 2007). A 0.5 és 80 Hz közötti tartományon kívül eső frekvenciákat kiszűrtem. A műtermékek eltávolítására független komponens analízist (independent component analysis, ICA) használtam (Delorme, Sejnowski & Makeig, 2007). A szemmozgással kapcsolatos műtermékeket tartalmazó ICA komponenseket manuálisan távolítottam el.

3.4.3. ERP elemzés

Az EEG adatokban a kiváltott potenciálok (ERP-k) is elemzésre kerültek. A célingerek (figyelt szövegben elhangzó számnevek), disztraktorok (ignorált szövegben elhangzó számnevek) és az ignorált oldalon elhangzó szintaktikai sértések kezdetétől számított -200 és +2400 ms közötti epochokat vettük figyelembe. A -200-0 ms-os időablakban mért átlagos feszültség szolgált baseline-ként. Kizártuk az elemzésből azon epochokat, melyek tartalmaztak +/-100 μ V nagyságúnál nagyobb feszültségváltozást bármely elektródán.

3.4.4. EEG forráslokalizáció és funkcionális konnektivitás

A forráslokalizációhoz először 2048 ms-os epochokra osztottuk az EEG felvételeket, majd kizártuk az elemzésből azokat az epochokat, melyek tartalmaztak célingert, disztraktort, vagy bármilyen gombnyomást - ezeket csak az ERP elemzésnél vettük figyelembe. A +/-100 μ V nagyságúnál nagyobb feszültségváltozást tartalmazó epochokat is eltávolítottuk. Ezen az adathalmazom a Brainstorm szoftver (Tadel et al., 2011) segítségével végeztünk forrásrekonstrukciót előző tanulmányok protokollja alapján (Huang et al., 2016; Pizzagalli, 2007; Song et al., 2015).

Az agyterületek közötti funkcionális konnektivitást a BrainWave szoftverben (Stam et al., 2007) számítottuk ki. Az FC-t a két régió között lévő fázisszinkronizáció segítségével fejeztük ki. Ennek keretében először megállapítottuk a releváns agyi régiókat (region of interest, ROI) az alapján, hogy a korábbi irodalom mely régiókat köti figyelmi, hallási vagy beszédfeldolgozási folyamatokhoz (Saur et al., 2008, 2010). A ROI-kat és rövidítésüket a 2. táblázat mutatja. Ezután minden ROI között phase lag indexet (PLI) számoltunk hat frekvenciatartományban (delta: 0.5-4 Hz, theta: 4-8 Hz, low alpha: 8-10 Hz, high alpha: 10-13 Hz, beta: 13-30 Hz, gamma: 30-80 Hz). A PLI 0 és 1 közötti érték lehet, ahol a 0 random fáziskülönbséget jelent (ez a minimum FC erősség), az 1 pedig konstans fáziskülönbséget (maximum FC erősség). Így 36*36-os mátrixok jöttek létre, melyek az egyes ROI-párok közötti PLI értékeket tartalmazták. Végül a mátrixokat kísérleti személy, kondíció és frekvenciatartomány alapján átlagoltuk.

Anatómiai régió	EEG forrásrégiók angol neve	Rövidítés
Frontális	Caudal Middle Frontal Gyrus	MFGc
	Rostral Middle Frontal Gyrus	MFGr
	Orbitofrontal Gyrus	OFG

	Inferior Frontal Gyrus	IFG
	Superior Frontal Gyrus	SFG
	Precentral Gyrus	PrCG
Cinguláris	Anterior Cingulate Cortex	ACC
	Posterior Cingulate Cortex	PCC
Temporális	Fusiform Gyrus	FFG
	Inferior Temporal Gyrus	ITG
	Middle Temporal Gyrus	MTG
	Superior Temporal Gyrus	STG
Parietális	Inferior Parietal Gyrus	IPG
	Superior Parietal Gyrus	SPG
	Supramarginal Gyrus	SMG
	Paracentral Gyrus	PCG
	Postcentral gyrus	PoCG
	Precuneus	PCUN

2. táblázat. EEG forrásrégiók és rövidítésük a nagyobb anatómiai régiók szerint.

3.5. Statisztikai elemzés

A viselkedési és ERP adatok statisztikai elemzését a Statistica 13.1 szoftver segítségével végeztük. Külön-külön ismételt méréses ANOVA elemzésekkel vizsgáltuk a találati arányokat, d' -ot, FA-t, a disztraktor-hatást és a felismerési memóriateljesítményt a prozódia (mesterségesen lapított vs. természetesen visszafojtott vs. normál) mint within-subject változó és a csoport (figyelt szöveg manipulált vs. ignorált szöveg manipulált) mint between-subject változó hatásának megállapítására.

Az ERP-elemzés során a célingerek által kiváltott parietális N2b és P3b amplitúdókat hasonlítottuk össze a kondíciók között, külön a két csoportban. A disztraktorok és szintaktikai hibázások által kiváltott parietális ERP-komponenseket a 0-val hasonlítottuk össze egyszélű t-teszttel FDR-korrekción (false discovery rate) alkalmazásával. Az FDR-korrekcióna a többszörös összehasonlítások miatt volt szükség.

A szignifikancia elvárt szintje 0.05. A szfericitás sérülése esetén a szabadságfok Greenhouse-Geisser korrekcióját alkalmaztuk. A hatásnagyság jelentésére a parciális eta-négyzetet (η^2) használjuk. Az utólagos összehasonlításokhoz Tukey HSD post-hoc tesztet alkalmaztunk.

A funkcionális konnektivitásra vonatkozó adatok statisztikai elemzésére a Network Based Statistics (NBS) eszköztárban került sor (Zalesky, Fornito & Bullmore, 2010). A prozódia hatását a következő páros összehasonlításokkal vizsgáltuk, külön a két csoportban és minden frekvenciatartományban: 1) mesterségesen lapított vs. normál prozódia, 2) természetesen visszafojtott vs. normál prozódia. Az FC elemzés vizualizálására a BrainNet Viewer eszköztárat (Xia & Wang, 2013) használtuk.

4. Eredmények

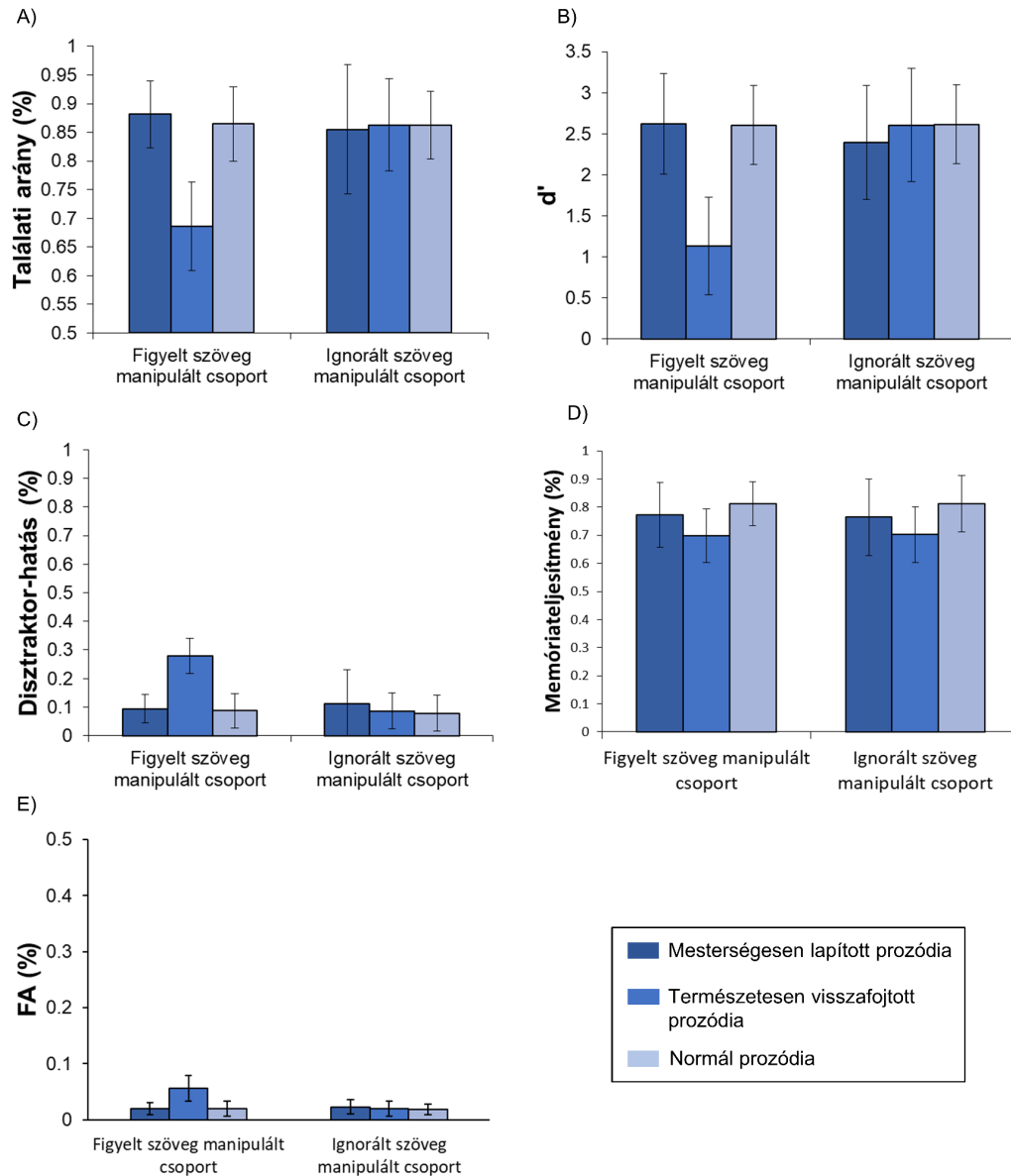
4.1. Viselkedéses válaszok

A 2. ábrán látható a kísérleti manipulációk hatása a találatokra (A), d' -ra (B), a disztraktor-hatásra (C), a felismerési emlékezetre (D) és a téves riasztásokra (E). A 3. táblázat tartalmazza ezen elemzések F, p és η^2 értékeit az ANOVA teszt szerint. Mind a csoportnak, mind a prozódia hatás szignifikáns főhatása volt minden viselkedéses válaszra a felismerési emlékezet kivételével, ám ezeket a főhatásokat a szignifikáns interakciós hatás eredményezte. A post-hoc teszt kimutatta, hogy a figyelt szöveg manipulált csoportban a természetes prozódia-visszafojtás kondícióban a hallgatók szignifikánsan eltérő válaszokat adtak, mint bármely más csoport-kondíció kombinációban: a találati arány és a d' alacsonyabb, a disztrakció és a téves riasztások aránya pedig magasabb volt ebben az esetben (minden $p < 0.001$). A felismerési memóriateljesítmény a természetes prozódia-visszafojtás kondícióban szignifikánsan alacsonyabb volt, mint a másik két kondícióban, azonban mesterségesen lapított prozódia esetén is romlott, ahol szignifikánsan alacsonyabb volt a normál prozódia kondíciónál. Csoport-főhatás vagy interakciós hatás ebben az esetben nem volt.

	CSOPORT	PROZÓDIA	CSOPORT * PROZÓDIA
Találatok	$F(1, 48) = 7.2236$, $p < 0.01$, $\eta^2 = 0.13$	$F(2, 96) = 46.254$, $p < 0.001$, $\eta^2 = 0.49$	$F(2, 96) = 51.074$, $p < 0.001$, $\eta^2 = 0.51$
d'	$F(1, 48) = 8.2951$, $p = 0.01$, $\eta^2 = 0.15$	$F(2, 96) = 49.670$, $p < 0.001$, $\eta^2 = 0.51$	$F(2, 96) = 54.538$, $p < 0.001$, $\eta^2 = 0.53$
Disztraktor-hatás	$F(1, 48) = 13.887$, $p < 0.001$, $\eta^2 = 0.22$	$F(2, 96) = 58.191$, $p < 0.001$, $\eta^2 = 0.55$	$F(2, 96) = 58.191$, $p < 0.001$, $\eta^2 = 0.58$
FA	$F(1, 48) = 11.661$, $p = 0.001$, $\eta^2 = 0.2$	$F(2, 96) = 60.368$, $p < 0.001$, $\eta^2 = 0.56$	$F(2, 96) = 63.111$, $p < 0.001$, $\eta^2 = 0.57$

Felismerési memória		$F(2, 96) = 22.479$, $p < 0.00$, $\eta p^2 = 0.32$	
---------------------	--	--	--

3. táblázat. A statisztikailag szignifikáns viselkedéses eredmények F, p és ηp^2 értékei.



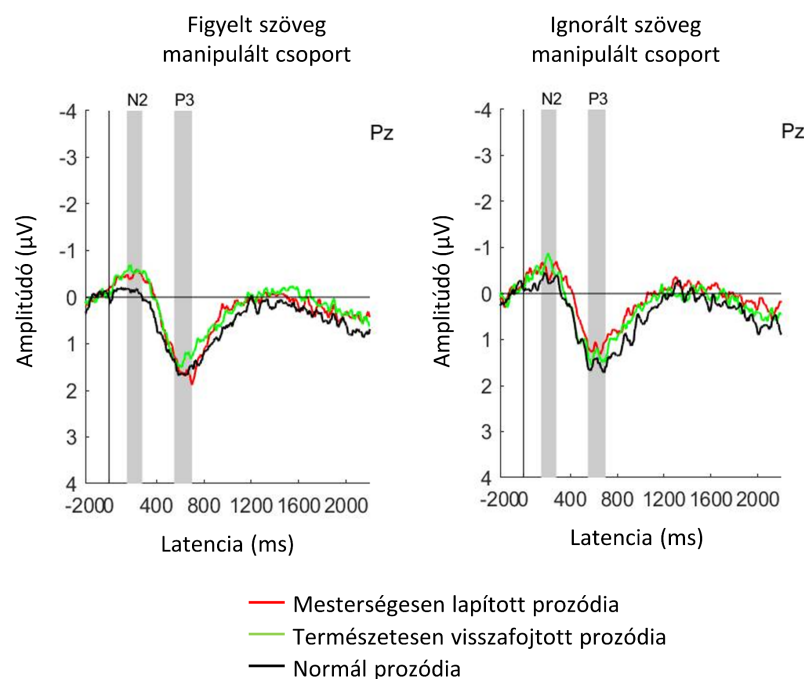
2. ábra. A prosódia és a csoport hatása a találati arányra (A), d' -ra (B), disztraktor-hatásra, (C) felismerési memóriára (D) és téves riasztásokra (E).

4.2. ERP eredmények

A 3. ábra a célingerek (figyelt szövegben elhangzó számnevek) által kiváltott agyi válaszokat mutatja a két csoportban a kondíciók szerint. A figyelt szöveg manipulált csoportban a prosódiai manipulációnak szignifikáns főhatása volt a célingerek által kiváltott N2b

ERP-komponensre ($F(2,48) = 8.056$; $\varepsilon = 0.671$; $p < 0.01$; $\eta^2 = 0.251$). A post-hoc összehasonlítás szerint az N2b amplitúdó szignifikánsan kisebb volt a normál prozódia kondíciónál, mint mesterségesen lapított vagy természetesen visszafojtott prozódiával felolvasott szöveg esetén ($p < 0.01$), bár utóbbi két kondíció között nem volt szignifikáns különbség. Ezzel szemben a prozódia főhatása a P3b amplitúdóra nem volt szignifikáns. A disztraktorok (ignorált szövegben elhangzó számnevek) szintén nem eredményeztek kiváltott választ. A ignorált szövegben megjelenő szintaktikai sértések csak akkor váltottak ki a 0-tól szignifikánsan eltérő P3b-t, amikor a figyelt oldalon természetesen visszafojtott prozódiájú beszéd hangzott el ($p < 0.05$).

A ignorált szöveg manipulált csoportban a prozódia szignifikáns hatása volt a célingereknél mért P3b-re ($F(2,48) = 4.373$; $\varepsilon = 0.769$; $p < 0.05$; $\eta^2 = 0.154$). A post hoc összehasonlítás azt mutatja, hogy a célingerek szignifikánsan nagyobb amplitúdójú P3b-t váltottak ki akkor, amikor a ignorált szöveg normál prozódiájú volt, mint amikor mesterségesen lapított prozódiájú ($p < 0.05$). A célingerek szignifikáns N2b-t nem váltottak ki, és a disztraktor események sem eredményeztek 0-tól eltérő kiváltott válaszokat.



3. ábra. A célingerek által kiváltott parietális (Pz) ERP-k csoportátlaga ($N = 25$) a figyelt szöveg manipulált (balra) és az ignorált szöveg manipulált csoportban (jobbra). A zero latencia a célinger (számnév) kezdete.

4.3. EEG funkcionális konnektivitás

A prozódia-manipuláció által szignifikánsan érintett funkcionális hálózatokat a 4. ábra szemlélteti. A függelék táblázatában látható, hogy a hálózatok egyes csomópontjai hány kapcsolattal rendelkeznek.

A prozódia manipulációja különböző funkcionális hálózatok aktivitását eredményezte a két csoportban. Ami a figyelt szöveg manipulált csoportot illeti, egy delta frekvenciatartományban (0.5–4 Hz) működő hálózat szignifikánsan eltérő aktivitást mutatott a mesterségesen lapított és a normál prozódia kondícióban. A hálózat (18 kapcsolat, 10 csomópont) a jobb agyféltekében volt domináns. A legtöbb kapcsolattal rendelkező területek a mediális frontális gyrus caudalis része (MFGc), valamint a superior parietális gyrus (SPG) voltak.

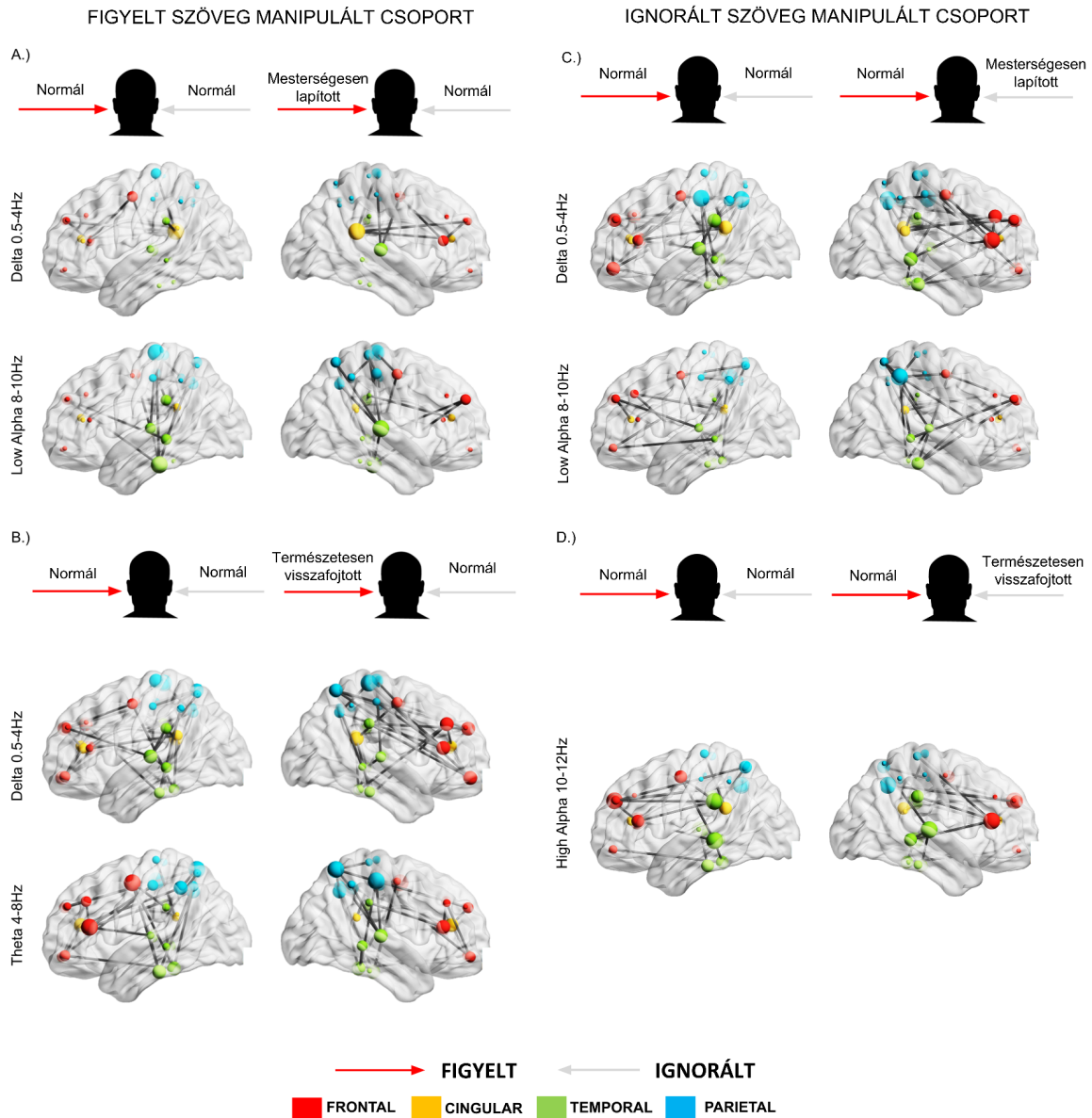
Ugyanezen kondíciók összevetésében a low alpha frekvenciatartományban (8–10 Hz) is megjelent egy, a parietális, frontális és temporális lebenyeket is magába foglaló funkcionális hálózat (36 kapcsolat, 17 csomópont). Ennek a legtöbb kapcsolattal rendelkező csomópontjai a bal féltekében a gyrus temporalis inferior (ITG), az orbitofrontális kéreg (OFC) és a paracentralis gyrus (PCG), a jobb féltekében pedig a superior temporális kéreg (STC) voltak.

Ezzel szemben amikor a figyelt oldalon nem mesterségesen lapított, hanem természetesen visszafojtott prozódiával felolvasott beszéd hangzott el, más funkcionális hálózatok jelentek meg. A delta frekvenciatartományban egy kiterjedt hálózatot látni (66 kapcsolat, 32 csomópont), melynek központi régiói a frontális lebenyben az OFC, az MFG és a superior frontális gyrus (SFG); a parietális lebenyben a supramarginalis gyrus (SMG) és az SPG; a temporális lebenyben a fusiform gyrus (FFG); valamint a posterior cingulum (PCC).

Ugyanebben az összehasonlításban (természetesen visszafojtott prozódia vs. normál prozódia a figyelt szövegben) a theta tartományba (4–8 Hz) eső oszcillációk fázisszinkronizációja is szignifikáns hálózatot alkotott (66 kapcsolat, 32 csomópont). Ennek központi csomópontjai a jobb féltekében a PCG és SMG, a bal féltekében pedig a precuneus (PCUN) és a PCC. Ezek a területek frontális és temporális kapcsolatokkal is rendelkeztek.

A ignorált szöveg prozódiai manipulálása szintén hatással volt a megjelenő funkcionális hálózatokra. A normál és a mesterségesen lapított prozódia ignorálása közben mért agyi aktivitás összevetésében a delta és low alpha tartományokban jelentek meg hálózatok, ugyanúgy, mint a másik csoportban ugyanezen kondíciók kontrasztjában. Ennek a delta hálózatnak a legtöbb kapcsolattal rendelkező csomópontjai a PCC, valamint két baloldali parietális terület, a PCG és PoCG. A low alpha hálózat központi területei a jobb féltekei inferior parietális gyrus (IPG) és ITG voltak.

Végezetül a figyeltek oldalon elhangzó természetesen visszafojtott és normál prozódiajú beszéd közötti különbség egy high alpha tartományú (10–13 Hz) hálózatban mutatkozott meg (42 kapcsolat, 21 csomópont), melynek központi területei frontálisan az MTG és SFG, parietálisan a postcentralis gyrus (PoCG) és SPG, temporálisan az STC, valamint a PCC.



4. ábra. A prozódia-manipuláció által szignifikánsan érintett funkcionális hálózatok a figyeltek szöveg manipulált (bal oldali oszlop, A és B ábra) és az ignoráltak szöveg manipulált (jobb oldali oszlop, C és D ábra) csoportban. A pontok a ROI középpontját jelölik (csomópontok). A csomópontok mérete a csomópont kapcsolatainak a számát mutatja az adott hálózaton belül.

5. Diskusszió

Tanulmányunkban a prozódia hatását vizsgáltuk két megszólaló beszédének láncra bontására és figyelmi szelekciójára, két külön csoportban manipulálva vagy a figyelt, vagy az ignorált szöveget. Azt vártuk, hogy a prozódia megváltoztatása hatással lesz a szelektív figyelemre, de a láncra bontásra kevésbé (1. hipotézis), valamint hogy a beszéd alaphfrekvenciájának teljes kiátlagolása erősebb hatásokhoz vezet, mint hogyha a beszélő maga próbálja visszafojtani a prozódiát (2. hipotézis). Eredményeink azt mutatják, hogy a normál prozódia valóban inkább a figyelmi szelekciót segíti elő, míg a hallási láncra bontás céljából kevésbé hasznos, mivel eltérő hatásokat figyeltünk meg akkor, amikor a prozódia-manipuláció a figyelt oldalon volt, mint akkor, amikor az ignorált oldalon. Ugyanakkor a 2. hipotézis nem igazolódott, hiszen a beszélők által természetesen előidézett prozódia-visszafojtás jóval nagyobb hatásokat eredményezett, mint az F0 mesterséges lapítása, főként ha mindez a figyelt szöveget érintette. A következőkben ezeket az eredményeket értelmezzük a viselkedéses és neurofiziológiai (ERP, FC) adatok alapján.

5.1. A prozódia megváltozásának hatása a figyelmi szelekcióra és a hallási láncra bontásra

5.1.1. Viselkedéses hatások

Hipotézisünk alapján azt vártuk, hogy beszédfeldolgozás során a leromlott prozódia a figyelmi szelekciót jobban érinti, mint a hallási láncra bontást. Viselkedéses szempontból ez a hipotézis nagyobb hatásokat feltételez a figyelt szöveg manipulálása, mint az ignorált szöveg manipulálása esetén. Eredményeink alátámasztják ezt a feltételezést, mivel az ignorált szövegen alkalmazott prozódiai változtatások nem vezettek szignifikáns célinger-detekciós hatásokhoz, míg a figyelt szöveg prozódiájának lerontása jóval alacsonyabb találati aránnyal és magasabb disztrakcióval járt.

Ugyanakkor az is látható, hogy amikor a figyelt szöveget manipuláltuk, a téves riasztások aránya is megnőtt valamelyest a disztraktor-hatással együtt. Emellett pedig volt egy olyan viselkedéses mérés, amelyre az ignorált szöveg manipulációja is hatással volt, mégpedig a felismerési memóriateljesítmény. A téves riasztások növekedése legalább részben az ignorált szöveg feldolgozásba való betörésének tulajdonítható, az emlékezeti teljesítmény romlása pedig arra utalhat, hogy a résztvevők a célinger-detekció fokozott nehézsége miatt stratégiát váltottak, és kevesebb erőforrást szántak az információ emlékezeti kódolására. Emiatt arra következtethetünk, hogy a prozódia valamelyest a hallási láncra bontásra is hatással van, habár ez a hatás kisebb, mint amit a figyelmi szelekcióra gyakorol.

5.1.2. ERP hatások

A célinger-detekcióhoz köthető ERP hatásokat illetően azt láttuk, hogy a figyelt szöveg prozódiajának bármilyen formájú módosítása nagyobb N2b amplitúdót eredményezett, mint amikor a figyelt szöveg prozódiaja normál volt (3. ábra, balra). Mivel az N2b amplitúdója nagyobb feldolgozási teher hatására növekszik (Patel & Azzam, 2005), ezért eredményeink alátámasztják, hogy az atipikus prozódiajú beszéd feldolgozása kihívást jelent a hallgatóknak. Kérdés, hogy ez a figyelmi szelekció vagy a láncra bontás nehézségét mutatja. Ennek eldöntésében segít, ha megnézzük a szintaktikai P3b válaszokat. Szintaktikai hibázások mindig csak az ignorált szövegben fordultak elő. Az ezek által kiváltott P3b válaszok amplitúdója a természetes prozódia-visszafojtás kondícióban elérte a szignifikancia-szintet, ami azt jelenti, hogy amikor prozódia-visszafojtott szövegre kellett figyelni, akkor az ignorálandó szövegben megjelenő szintaktikai sértések mégis gyakran mentek át tudatos feldolgozáson. Ez a figyelt szöveg szelekciójának a nehézségét mutatja. Ugyanakkor a másik, mesterségesen manipulált kondícióban a szintaktikai P3b-k nem mutatkoztak szignifikánsnak, bár ez lehet amiatt, hogy a hatás megvan, csak kisebb.

A célingerek (figyelt szövegben elhangzó számnevek) által ebben a csoportban kiváltott P3b válaszok is hasonló irányba mutatnak: ezek amplitúdója akkor volt a legalacsonyabb, amikor a figyelt szöveg természetesen prozódia-visszafojtott volt. Polich (2007) szerint a P3b esetében az alacsonyabb amplitúdó azt mutatja, hogy a célinger detekciójakor a figyelem jobban megoszlott a bejövő ingerek között, kevésbé volt fókuszált. Így ez az eredmény is figyelmi szelekciós nehézségekre utal.

Az ignorált szöveg manipulált csoportban - ahol a figyelt szöveg mindig normál prozódiaval rendelkezett - a célingerek által kiváltott P3b válaszok amplitúdója nagyobb volt akkor, amikor az ignorált szöveg is normál prozódiajú volt, és kisebb akkor, amikor mesterségesen lapított. Ezért a figyelem jobban megoszlott a mesterségesen lapított prozódia kondícióban, vagyis a lapított alaphézfkvenciájú beszéd valamivel jobban elterelhette a figyelmet, mint a normál beszéd. Azonban ez a hatás nem volt elég erős ahhoz, hogy megmutatkozzon a viselkedési válaszokban (a disztraktor-hatás itt nem volt szignifikáns). Ugyanakkor a memóriateljesítményben ugyanaz a mintázat rajzolódott ki, mint a figyelt szöveg manipulált csoportban, azaz bármely prozódia-manipulált kondícióban romlott az emlékezet. Ez ismét afelé mutat, hogy kis mértékben a hallási láncra bontás is akadályozva volt.

5.1.3. FC hatások

Több-beszélős szituációban a lassú hullámú oszcillációk (1-7 Hz) támogatják a figyelmi szelekciót azzal, hogy fázisszinkronizáción keresztül lekövetik a figyelt szöveget (Golumbic et al., 2013). Ezért ha a prozódia inkább a szelekciót befolyásolja, mint a láncra bontást, akkor a jelen kísérleti manipulációk hatására emelkedett funkcionális konnektivitást várunk a delta (0.5-4 Hz) és theta (4-8 Hz) tartományú hálózatokban. Ennek megfelelően eredményeink azt mutatják, hogy mesterségesen lapított prozódijú szöveg követésekor a delta tartományban nő a konnektivitás, természetesen visszafojtott prozódijú szöveg esetén pedig a delta hálózaton felül a theta tartományban is megjelenik egy hálózat. Ugyanakkor a viselkedéses és ERP hatásokhoz hasonlóan hallási láncra bontási nehézségekre is találunk némi bizonyítékot a mesterségesen lapított prozódia kondícióban, minthogy a low alpha (8-10 Hz) tartományban is emelkedett funkcionális konnektivitást látunk.

5.2. Különböző prozódiai jegyek különböző hatással vannak a több-beszélős szituációban történő beszédfeldolgozásra

A 2. hipotézis szerint az F0 teljes kiátlagolása hátrányosabban érinti a beszédfeldolgozást több-beszélős szituációban, mint a természetes prozódia-visszafojtás, hiszen a két eset közül az utóbbiban maradt nagyobb F0 variancia az ingerekben, és az intonációs kontúr a prozódia egyik prominens aspektusa (Meyer, 2018). Azonban ez a hipotézis megdőlt: a természetesen prozódia-visszafojtott beszédet nehezebb volt követni és az emlékezetbe kódolni, mint a mesterségesen lapított intonációs kontúrral rendelkezőt. Ez azt mutatja, hogy nem az intonációs kontúr variabilitása az, ami önmagában meghatározza a szelektív figyelmet vagy a hallási láncra bontást: csupán ennek a variabilitásnak az eltávolítása nem érintette hátrányosan ezeket a folyamatokat a viselkedés szintjén.

Ahogy az 1. táblázat is mutatja, a két prozódia-manipuláció közötti különbség a beszédtempóban rejlik: a természetes prozódia-visszafojtás gyorsabb beszédtempót eredményezett rövidebb szünetekkel, míg a mesterséges manipuláció a normál beszédhez képest kizárólag az F0-t érintette, a beszédtempót nem. A gyorsabb beszédtempó vélhetően akadályozta a beszéd értelmes egységekre való bontását, elemzését, ezzel is nehezítve a beszédértést, és hátráltatva az adott beszédinger szelekcióját. Ezért azt mondhatjuk, hogy a beszédtempó és szünetezés fontos prozódiai jegyek a több-beszélős szituációban történő beszédfeldolgozás szempontjából. Ez nem jelenti azt, hogy az F0 ne játszana ebben szerepet. A mesterségesen lapított prozódia kondícióban tapasztalt emlékezeti kódolási nehézségek, alacsony P3b amplitúdó és a low alpha aktiváció mind arra utalnak, hogy az F0-variancia

hiánya hátráltatja a hallási láncra bontást. Ez a hatás azonban a viselkedésben kompenzálható, vélhetően nagyobb kognitív erőfeszítéssel.

6. Összefoglalás

Kutatásunkban arra kerestük a választ, hogy több-beszélős szituációban a beszéd prozódiaja milyen szerepet játszik a beszédfeldolgozásban: a hallási láncra bontásra vagy a figyelmi szelekcióra hat inkább? Összességében azt találtuk, hogy elsősorban a figyelmi szelekciót segíti elő a beszéd egységeinek a prozódia általi körülhatárolása, míg ez a szerep a hallási láncra bontásban másodlagos. Arra is fény derült, hogy a prozódia különböző jegyei különböző súllyal vesznek részt a több-beszélős szituációban történő beszédfeldolgozásban. Az intonációs kontúr variabilitása a figyelmi szelekcióban nem játszik pótolhatatlan szerepet, mivel ez az F0 kiátlagolása mellett is sikeres volt. A hallási láncra bontás szempontjából már fontosabb jegy az F0-variancia, ám hiányát is ellensúlyozni tudja a hallgató a beszédfeldolgozás során. Ezzel szemben a beszédtempó – amelyre a beszédszünetek hossza is kihat – már jelentősebb szerepet játszik a figyelmi szelekcióban. A megrövidült vagy kihagyott beszédszünetek megnövelték a munkamemória-terhet, megnehezítették az adott szöveg szelektív feldolgozását. Ezért több beszélő esetén a prozódia számos aspektusa közül a beszédtempó segíti leginkább a hallgatót a hatékony figyelmi szelekcióban.

Köszönetnyilvánítás

Köszönet illeti a tanulmány minden társszerzőjét – Tóth Brigittát, Szalárdy Orsolyát, Honbolygó Ferencet, Kohári Annát, Mády Katalint és Winkler Istvánt –, akiktől annyit tanultam.

Hivatkozott irodalom

- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.
- Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, 44(1), 51.
- Broersma, P., & Weenink, D. (2005). Praat software (version 5.2. 01).
- Carlson, K. (2009). How prosody influences sentence comprehension. *Language and Linguistics Compass*, 3(5), 1188-1200.
- Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4), 1443–1449.
- Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29), 11854–11859. doi:10.1073/pnas.1205381109
- Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. In bioRxiv. <https://doi.org/10.1101/238642>
- Frazier, L., Carlson, K., & Clifton Jr, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244-249.
- Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, 77(5), 980–991.
- Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Peninsula Publication.
- Hill, K. T., & Miller, L. M. (2009). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, 20(3), 583–590. doi:10.1093/cercor/bhp124
- Huang, Y., Parra, L. C., & Haufe, S. (2016). The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting. *NeuroImage*, 140, 150–162.
- Keitel, A., Prinz, W., Friederici, A. D., Hofsten, C. V., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children’s development.

- Journal of Experimental Child Psychology*, 116(2), 264–277.
<https://doi.org/10.1016/j.jecp.2013.06.005>.
- Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40(2), 153–194.
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, 11(5), 539–550.
<https://doi.org/10.3758/BF03196991>
- LaCroix, A. N., Blumenstein, N., Tully, M., Baxter, L. C., & Rogalsky, C. (2020). Effects of prosody on the cognitive and neural resources supporting sentence comprehension: A behavioral and lesion-symptom mapping study. *Brain and language*, 203, 104756.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010.
doi:10.1016/j.neuron.2007.06.004
- Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7), 2609–2621. <https://doi.org/10.1111/ejn.13748>
- Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain sciences*, 9(3), 70.
- Obleser, J., Wise, R. J., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, 27(9), 2283–2289. doi:10.1523/JNEUROSCI.4663-06.2007
- Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the Event-Related Potential. *International Journal of Medical Sciences*, 2(4), 147–154.
<https://doi.org/10.7150/ijms.2.147>
- Pizzagalli, D. A. (2007). Electroencephalography and high-density electrophysiological source localization. *Handbook of Psychophysiology*, 3, 56–84.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Rimmele, J. M., Golumbic, E. Z., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, 68, 144–154.

- Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PloS one*, 8(2), e56000.
- Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 25(23), 3079–3085.
- Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, 105(46), 18035–18040. doi:10.1073/pnas.0805234105
- Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Küpper, H., Kellmeyer, P., Kümmerer, D., Klöppel, S., Glauche, V., Lange, R., Mader, W., Feess, D., Timmer, J., & Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *NeuroImage*, 49(4), 3187–3197.
- Sheppard, S. M., Love, T., Midgley, K. J., Holcomb, P. J., & Shapiro, L. P. (2017). Electrophysiology of prosodic and lexical-semantic processing during sentence comprehension in aphasia. *Neuropsychologia*, 107, 9–24.
- Song, J., Davey, C., Poulsen, C., Luu, P., Turovets, S., Anderson, E., & Tucker, D. (2015). EEG source localization: Sensor density and head surface coverage. *Journal of Neuroscience Methods*, 256, 9–21.
- Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28(11), 1178–1193. doi:10.1002/hbm.20346
- Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8(350), 1–7. <https://doi.org/10.3389/fnhum.2014.00350>
- Szalárdy, O., Tóth, B., Farkas, D., Hajdu, B., Orosz, G., & Winkler, I. (2020). Who said what? The effects of speech tempo on target detection and information extraction in a multi-talker situation: An ERP and functional connectivity study. *Psychophysiology*. 2020;00:e13747. <https://doi.org/10.1111/psyp.13747>
- Szalárdy, O., Tóth, B., Farkas, D., Kovács, A., Urbán, G., Orosz, G., Szabó, B. T., Hunyadi, L., Hajdu, B., & Winkler, I. (2018). The effects of attention and task-relevance on the processing of syntactic violations during listening to two concurrent speech streams. *Cognitive, Affective and Behavioral Neuroscience*, 18(5), 932–948. <https://doi.org/10.3758/s13415-018-0614-4>

- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011, 879716-879716. doi:10.1155/2011/879716
- Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, 50(11), 3831–3842.
- Tóth, B., Honbolygó, F., Szalárdy, O., Orosz, G., Farkas, D., & Winkler, I. (2020). The effects of speech processing units on auditory stream segregation and selective attention in a multi-talker (cocktail party) situation. *Cortex*, 130, 387-400.
- Tóth, B., Farkas, D., Urban, G., Szalárdy, O., Orosz, G., Hunyadi, L., Hajdu, B., Kovács, A., Szabó, B. T., Shestopalova, L. B., & Winkler, I. (2019). Attention and speech-processing related functional brain networks activated in a multi-speaker environment. *PloS One*, 14(2), e0212754.
- Wöstmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, 113(14), 3873–3878.
- Wöstmann, M., Lim, S. J., & Obleser, J. (2017). The human neural alpha response to speech is a proxy of attentional control. *Cerebral cortex*, 27(6), 3307–3317.
- Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PloS One*, 8(7), e68910. doi:10.1371/journal.pone.0068910
- Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4), 1197–1207. doi:10.1016/j.neuroimage.2010.06.041