

VIZUÁLIS INGEREK HATÁSA MOZGÓ HANGFORRÁSOK ÉSZLELÉSÉRE

Szombathy Péter, Hajdu Botond

Konzulensek:

Bóhm Tamás

MTA TTK Kognitív Idegtudományi és Pszichológiai Intézet
és BME VIK Távközlési és Médiainformatikai Tanszék

Babarczy Anna

BME TTK Kognitív Tudományi Tanszék

Budapest

2012. október

Absztrakt

Mindennapi életünk egy meghatározó észlelési folyamata a környezetünkben lévő tárgyak egymástól való elkülönítése (ld. akár az utcai közlekedést vagy a szakirodalomban gyakran hivatkozott koktélparti-jelenséget). Az ennek hátterében álló kognitív mechanizmusok egyelőre csak részben feltártak. A tárgyak megkülönböztetését elsősorban látási és hallási folyamatok biztosítják, melyek (visszacsatolásokon keresztül) egymást segítve működhetnek.

Kísérletünkkel azt vizsgáltuk, vajon a hangforrások mozgásával együtt érkező vizuális ingerek képesek-e befolyásolni a hangok észlelését. A kísérlet résztvevői egy virtuális térbe kerültek. Speciális fülhallgatók segítségével műfejes felvétellel rögzített, a hallgató körül mozgó illetve álló szinuszos hangforrások által keltett hangsorokat hallottak; eközben a látóterük nagy részét kitöltő kijelzőn a hanggal szinkronban pulzáló gömbök voltak láthatóak, melyek a hangforrásoknak megfelelően – vagy éppen azoktól eltérően – mozogtak a térben, az észlelő nézőpontjából.

A kísérletben résztvevők elsődleges feladata annak visszajelzése volt, hogy az általuk hallott két szinuszos, alternáló hangsort mikor hallják egyetlen hangforrásból származó dallamnak (integrált szerveződés), és mikor észlelik őket két külön hangforrásként (szegregált szerveződés). A választ korábbi vizsgálatok alapján a hangforrások egymáshoz viszonyított elhelyezkedése jelentősen befolyásolhatja. Jelen kísérlet célja azonban annak kiderítése volt, hogy vajon a hallott téri mozgásnak megfelelően – vagy attól eltérően – mozgó vizuális inger hatással van-e a hangok integrált illetve szegregált észlelésére.

Kezdeti eredményeink azt mutatták, hogy a mozgó tárgyak hang alapján történő szétválasztását a hozzájuk kapcsolható téri vizuális ingerek nem befolyásolják. Ez egybevág más kísérleti elrendezésben kapott korábbi eredményekkel [SPENCE ÉS DRIVER, 1997] és információkkal szolgálhat a halló- és a látórendszer által alkalmazott téri reprezentációkról.

1 Bevezetés

Az embert körülvevő tér olyan, különböző forrásokból származó hangokkal van tele, melyek gyakran átfedik egymást időben, spektrális tulajdonságokban vagy mindkettőben. Annak érdekében, hogy az akusztikai környezetünkben található objektumokat érzékelni tudjuk, a hallórendszerünk feladata, hogy a fülünkbe érkező jeleket elválassza egymástól. Habár a hangok szétválasztása egy bonyolult számítási feladat, tapasztalatból tudjuk, hogy az emberi hallórendszer hatékonyan képes ezt elvégezni. A hallott ingerek alapján a környezet rendezett reprezentációjának kialakítását BREGMAN [1990] után hallásijelenet-elemzésnek (auditory scene analysis, ASA) nevezzük. Ezen problémakör keretein belül a hallott ingerek számos akusztikai tulajdonságának hatását vizsgálták aszerint, hogy hogyan befolyásolják a hangok koherens egységekre, ún. hallási láncokba szervezését. Például a hangszint hatékony ismertetőjegynek találták a hangok forrásaik szerinti megkülönböztetéséhez [lásd például SMITH ÉS MUNKATÁRSAI, 1982].

Ezen kutatások túlnyomó részében a „hallási láncokra bomlás” paradigmát alkalmazták [VAN NOORDEN, 1975], melyben a hallgatók olyan ismétlődő hanghármasokat hallanak, amelyek két, egymástól bizonyos akusztikai tulajdonság(ok)ban eltérő hangból állnak. Az így elrendezett ingereket a megfigyelő hallhatja egyetlen hangsornak (amit integrált észleletnek neveznek) vagy két egyidejű hangsornak (szegregált észlelet). A kétféle hang hasonlósága erősen befolyásolja, hogy az ilyen hangsorokat milyenek észleljük [MOORE ÉS GOCKEL, 2002]. Ilyen hangsorok hosszabb időn át történő folyamatos hallgatása esetén a résztvevők spontán váltásokat észleltek a két különböző észlelet között [például PRESSNITZER ÉS HUPÉ, 2006]. Ezek a váltások akkor is megtörténtek, ha az ingerek elrendezése sokkal nagyobb mértékben segítette elő az egyik észleletet, mint a másikat [DENHAM ÉS MUNKATÁRSAI, nyomtatás alatt].

VAN NOORDEN [1975] és JUDD [1977] egyaránt azt találták, hogy a két fülbe felváltva lejátszott hangsorokat a hallgatók két különböző hallási láncként észlelték [lásd BREGMAN, 1990]. Továbbá DENHAM ÉS MUNKATÁRSAI [2009] és SZALÁRDY ÉS MUNKATÁRSAI [nyomtatás alatt] egyaránt kimutatták, hogy hanghármasok téri elhelyezkedésbeli különbsége is megkönnyítette a szegregált észlelet kialakulását.

BÖHM ÉS MUNKATÁRSAI [nyomtatás alatt] térben elválasztott helyzetű és térben mozgó hangforrások hatását vizsgálták a hallási láncok szétválasztására. Ezzel azt is

vizsgálták, hogy a "közös sors" Gestalt-elve [KÖHLER, 1947] érvényesül-e térben mozgó hangforrások esetén (a Gestalt "hasonlóság elvét" hallási láncokra bomlás esetén a fentebb hivatkozott kutatások már kimutatták). Különösképpen az képezte vizsgálat tárgyát, hogy az együtt mozgó hangforrásokból érkező hangokat nagyobb valószínűséggel észlelik-e a hallgatók egyetlen hangsornak, mint azokat, amelyek együtt állnak; valamint az ellenkező esetet is vizsgálták, miszerint a különböző pályákon mozgó hangforrásokat nagyobb valószínűséggel észlelik-e a hallgatók két külön hangsornak, mint a két, egymástól távolabb elhelyezkedő, álló hangforrásból érkezőket. Kísérletükkel továbbá kimutatta a „közelség” Gestalt-elve alapján feltételezhető

BŐHM ÉS MUNKATÁRSAI [nyomtatás alatt] kísérletében a felhasznált ingeranyagot emberi mozgáshoz kötött hangforrásokkal hozták létre, fejre erősített hangszórók használatával. Ez az összeállítás a téri elhelyezkedés és a mozgási pályák kevésbé pontos kezelését teszi lehetővé, mint a szimulált, vagy a pontosan irányított motorok által létrehozott mozgás (mely módszereket a hallott téri mozgást vizsgáló pszichofizikai és fiziológiai kísérletekben gyakran használják). Ennek ellenére használatát az indokolja, hogy az így keletkezett mozgásminták jobban tükrözik a természetes körülmények között szerzett tapasztalatokat.

BŐHM ÉS MUNKATÁRSAI [nyomtatás alatt] eredményei a "közös sors" Gestalt-elvére alapozott elvárásokkal ellentétben nem mutattak szignifikáns interakciót a hangforrások egymáshoz viszonyított helyzete és mozgó vagy álló helyzetük között. Ezen adatok alapján a hallórendszer nem használta fel a hallott mozgásból kinyerhető jellemzőket a felhasznált hangsorok láncokra bontásához. Ezzel ellentétben a térben egymástól távolabb elhelyezett források elősegítették a szegregált érzet kialakulását mind álló, mind mozgó hangok esetén (a korábbi kutatásokkal egybevágóan). Mozgó hangok esetén (a hangforrások relatív helyzetétől függetlenül) a szegregált és integrált észleletek átlagos hossza szignifikánsan csökkent, ami azt jelenti, hogy a résztvevők mozgó hangok esetén gyorsabban váltottak a két észlelet között, mint álló hangok esetén. Feltételezéseik szerint a mozgásból kinyerhető jellemzők stabilizáló hatásának hiánya abból következhet, hogy pusztán a hangforrások által bejárt pályák közötti egyezőségek és különbségek nem elegendőek a hallási tárgyak stabil reprezentációinak kialakulásához.

Jelen kísérlet célja az volt, hogy megállapítsuk, a hallott környezethez kapcsolt vizuális tér segíthet-e a hallási tárgyak megbízhatóbb elkülönítésében. Az egyes észlelési modalitások között sokféle egymásrahatást mutattak ki. A vizuális tér ilyen hatását a hallási környezetre jól illusztrálja az ún. McGurk-effektus [MCGURK ÉS MACDONALD, 1976], mely szerint bizonyos hallott szótagok percepcióját módosítani tudja egy egyidejűleg bemutatott videó egy eltérő szótag artikulációjáról. Ez a hatás viszont nem a hallási tér elemzésén, hanem a beszédpercepción alapul, így az általunk vizsgált feldolgozási mechanizmus esetleges modalitásközi hatásaira nem enged következtetni.

BROOKS ÉS MUNKATÁRSAI [2007] kísérletének eredményei alapján az álló hangokhoz viszonyítva a látott mozgással azonos irányban mozgó hangok segítették, míg az ellentétes irányban mozgókat gátolták a vizuális ingerek mozgásának érzékelését. Bár ez az eredmény az általunk vizsgálthoz képest fordított irányú modalitásközi hatást írt le (hanginger hatását vizuális érzékelésre), azt jelzi, hogy a mozgó vizuális és hallási objektumok reprezentációja interakcióba léphet egymással.

KUBOVY ÉS VAN VALKENBURG [2001] a releváns szakirodalmat összefoglalva azt feltételezi, hogy a hallásnak a látási rendszerhez hasonlóan külön pályái lehetnek a tárgyak azonosítására illetve helyzetük megállapítására. Ezek közül a hangok helyzetére vonatkozó érzetek egyértelműen segítik a látott tárgyak helyzetének megállapítását, ellenben a látási oldalról egyedül a helymeghatározás van hatással és csak a hallott tárgyak azonosítására. Ez BROOKS ÉS MUNKATÁRSAI [2007] eredményeit támogatja, továbbá feltételezi viszont egy vizuális–hallási hatás hiányát is.

MEYER ÉS MUNKATÁRSAI [2005] annak az eltérésnek az okát vizsgálta, hogy míg az álló képi és hanginger egyidejű bemutatása megkönnyíti az ingerek érzékelését, addig ugyanez az ingerküszöb-csökkenés nem jelentkezett korábbi, mozgó ingerekkel folytatott kutatásokban. Eredményeik azt mutatják, hogy az azonos irányú és helyzetű ingerek esetén ez az előny visszatér, de csak abban az esetben, ha a helyzetük meghatározásához minden olyan jellemző rendelkezésre áll, amelyek a valódi életben érzékelt fizikai objektumok esetében is.

Ezen eredmények alapján feltételezzük, hogy a hallási tárgyak reprezentációját az ezekkel együtt álló ill. mozgó látott tárgyak segíthetnek stabilizálni. Kísérletünkben ez

alapján az alábbi hipotéziseket vizsgáljuk a hallási láncokra bomlás paradigma használatával:

1. Együtt mozgó vagy álló hangforrások esetén a velük azonos pályán (ún. kongruens módon) mozgó vizuális térbeli ingerek a hangsorok integrált érzetének kialakítását, míg

2. az egymástól függetlenül mozgó illetve külön álló hangforrásokkal kongruens mozgó vizuális ingerek a szegregált hangérzet kialakítását segítik elő, szemben a hangforrásokétól eltérő pályán (inkongruens módon) mozgó látott tárgyakkal.

3. A hangforrásokkal kongruens módon mozgó vagy álló vizuális ingerek elősegítik a hallási tárgyak stabilabb reprezentációjának kialakulását, növelve ezzel az észlelettípusok közötti váltások között eltelt időt a megfelelő, inkongruens esetekhez viszonyítva.

Emellett kíváncsiak voltunk arra is, hogy ki tudunk-e mutatni a Gestalt „közös sors” szerveződési elvnek megfelelő hatást együtt mozgó hangforrásokra. Bár BÖHM ÉS MUNKATÁRSAI [nyomtatásban] kísérletében pusztán hangingerekkel ezt nem sikerült kimutatni, lehetséges, hogy a látott ingerek segítségével megjelenhet a „közös sors” elve alapján várt eredmény [lásd MEYER ÉS MUNKATÁRSAI, 2005]:

4. A hangforrások együtt mozgása elősegíti az integrált észlelet kialakulását az együtt álló esetekhez képest, valamint

5. a hangforrások egymástól függetlenül való mozgása az egymástól távolabb álló esetekhez képest a szegregált észlelet nagyobb arányú megjelenését okozza.

2 Módszerek

2.1 Résztvevők

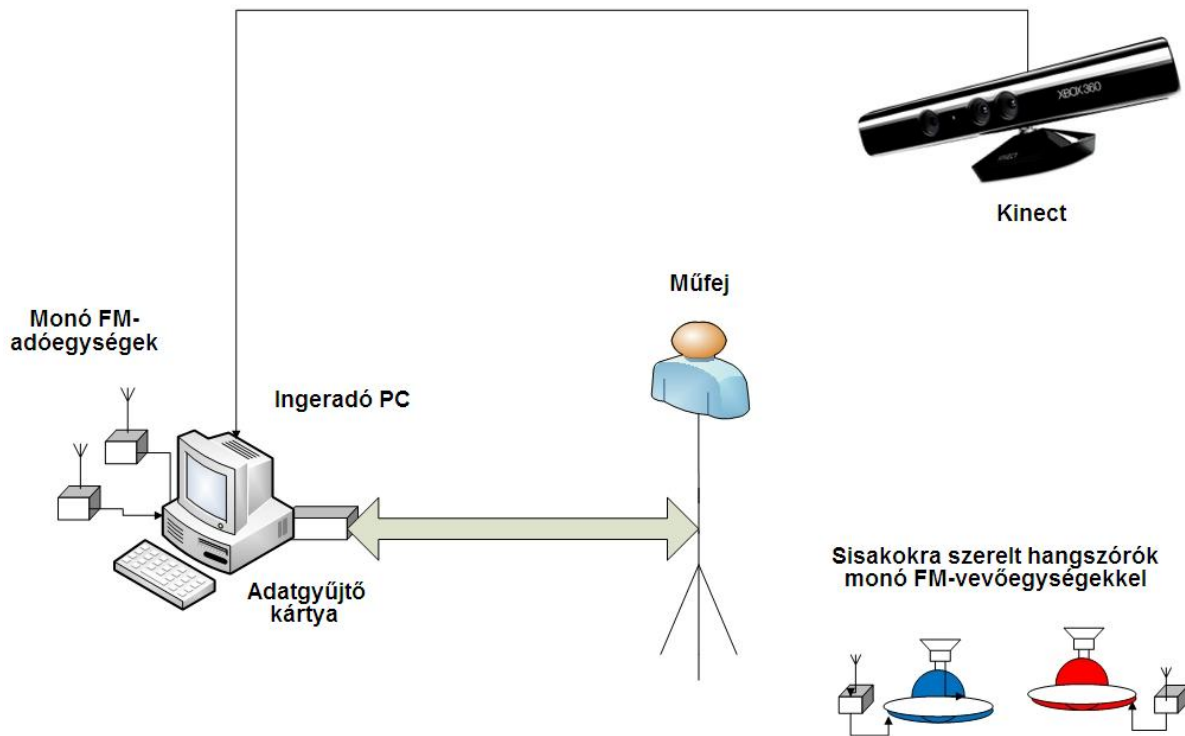
A vizsgálatot eddig 19 kísérleti személyen végeztük el, az adatfelvétel azonban még nem zárult le. A résztvevők közül 11 férfi, 17 jobbkezes, életkoruk 18 és 26 év közötti, átlagosan 21,4 év volt; közreműködésükért csekély juttatásban részesültek. Először egy hallásküszöbmérést végeztünk el rajtuk, és csak azok vehettek részt a kísérlet további részében, akiknek a hallása megfelelő volt: a vizsgált 250, 500, 1000, 2000 és 4000 Hz-es frekvenciákon a hallásküszöbnek a bal és a jobb fül esetén is 25 dB_{HL} alatt kellett lennie, a fülenként mért érték pedig nem különbözhetett egymástól 10 dB-nél (500 és 1000 Hz esetén 5 dB-nél) nagyobb mértékben. A résztvevők egy, az MTA Etikai Bizottsága által előzetesen jóváhagyott nyilatkozatot is aláírtak a kísérlet megkezdése előtt.

2.2 Eszközök és ingerek

A kísérlet során egy korábbi vizsgálatához készült hangfelvételeket használtunk fel, melyek rögzítése GEORGIU ÉS MUNKATÁRSAI [2011] által leírtakhoz hasonlóan történt [BŐHM ÉS MUNKATÁRSAI, 2012]. A rögzített hangminták két alternáló, különböző hangmagasságú, azonos amplitúdójú szinuszos hangból álltak: 100 ms hosszúságú, felváltva 400 illetve 475,7 Hz frekvenciájú tiszta szinuszos hangjelek követték egymást emelt koszinuszos fel- és lefutással, közöttük 25 ms szünetekkel (interstimulus interval, ISI); így összességében 250 ms-ig tartott egy két hangból és két szünetből álló periódus.

A felvételek során a különböző frekvenciájú szinuszos jelek egy-egy Anthony Gallo Acoustics A'Diva Ti típusú hangszórón szólaltak meg. Ezek munkavédelmi sisakokon voltak rögzítve, hogy így könnyen hordozhatók legyenek, és két ember által viselve a mozgásuk a mindennapi életben megszokott jellemzőkkel rendelkezzen. A hangszórók függőlegesen felfelé voltak irányítva, hogy a vízszintes síkban egyenletes legyen azok iránykarakterisztikája. A jeladó PC-től FM adó-vevők segítségével jutott el a jel a hangszórókig.

A hangfelvételek egy jól csillapított színházteremben készültek egy Head Acoustics HSU III.2 műfejes mikrofon segítségével. A mikrofonok jelét egy National Instruments 4462 adatfeldolgozó kártya rögzítette két csatornán, 96 kHz mintavételezési frekvenciával és 24 bites felbontással. A felvételi elrendezést a *2.1 ábra* mutatja [BŐHM ÉS MUNKATÁRSAI, 2012].



2.1 ábra: A felvételi elrendezés

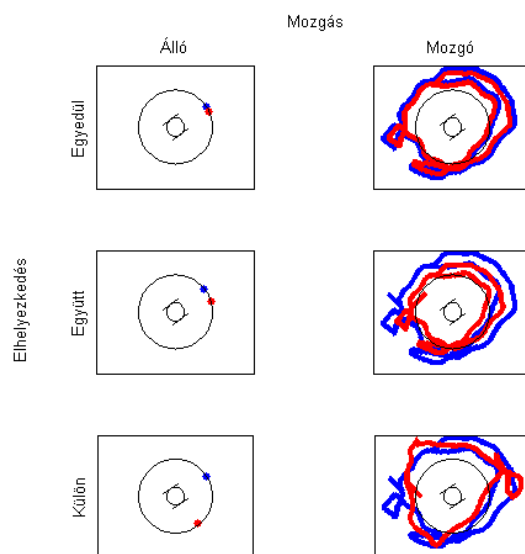
A rögzített hang utófeldolgozása során egy 350 és 526 Hz közötti sávszűrés történt (harmadfokú Butterworth-szűrővel), mely a két szinuszos hangingerre nincs hatással, de az esetlegesen beszűrődő egyéb zajokat, lépések vagy a szellőzés hangját megfelelően képes kiszűrni.

A felvételek készítésekor két ember viselte illetve mozgatta a sisakra erősített hangszórókat. A sisakok különböző színűek voltak, melyek a felvételek során végig ugyanahhoz a hangmagasságú hanghoz tartoztak: a piros színű a 400 Hz, illetve a kék a 475,7 Hz frekvenciájú hangot szolgáltatta meg (kivéve természetesen azokat az eseteket, ahol mindkét hangjel egyetlen hangszóróból szólt). A későbbiekben a színek jelentőségére még visszatérünk.

A korábban említett vizsgálathoz készült felvételek közül hat bejárást használtunk fel a kísérletünkben. (A hangforrások által bejárt pályát trajektóriának nevezzük.) Az egyes bejárások különböztek egymástól a hangforrások mozgásában: azok vagy egy helyben álltak (eltekintve az akaratlan fejmozgásokból adódó apró elmozdulásoktól), vagy

véletlenszerű pályán mozogtak a terem közepén elhelyezett műfej körül. Ez alapján a **Mozgás** függő változó (faktor) kétféle értéket (szintet) vehet fel: lehet *álló* illetve *mozgó*. Különböző lehetett a hangforrások egymáshoz viszonyított elhelyezkedése is, ezeket a **Relatív helyzet** függő változó szintjeivel jellemezhetjük: *azonos* (egyetlen hangszóróból szólt mind a két hang), *közeli* (a két hangszórót hordozó ember egymás mellett állt / sétált), illetve *független* (a hangszórókat hordozók egymástól távol álltak, illetve függetlenül sétáltak). A kondíciókat a 2.2 ábra példái szemléltetik.

A kísérlet során a hangingerekhez képanyagot is készítettünk, mely a hangforrásoknak megfelelő, vagy azoktól eltérő mozgást ábrázolt. Ez alapján a kísérletben vizsgált harmadik, ún. **Kongruencia** függő változó értéke rendre lehet *kongruens* vagy *inkongruens*. A kondíciók összesítését a 2.1 táblázat mutatja.



2.2 ábra: A hangszórókat hordozó emberek által bejárt trajektóriák sematikus illusztrációja.

A felvételek során a műfej mikrofon a terem közepén volt, annak sarka felé irányítva.

A műfej körül rajzolt kör sugara 80 cm.

2.1 táblázat: A kísérletben felhasznált kondíciók

#	Kongruencia	Mozgás	Relatív helyzet
1	kongruens	álló	azonos
2	kongruens	álló	közeli
3	kongruens	álló	független
4	kongruens	mozgó	azonos
5	kongruens	mozgó	közeli
6	kongruens	mozgó	független
7	inkongruens	álló	azonos
8	inkongruens	álló	közeli
9	inkongruens	álló	független
10	inkongruens	mozgó	azonos
11	inkongruens	mozgó	közeli
12	inkongruens	mozgó	független

Nem állt rendelkezésre két egymáshoz közel, középen álló hangforrást tartalmazó hangfelvétel, azonban szükségünk volt ilyen ingerre. Ehhez egy olyan felvételt használtunk fel, amelyben a műfej előtt illetve mögött helyezkedett el a két hangforrás, kihasználva azt az ismert jelenséget, hogy hallórendszerünk nehezen tud különbséget tenni az előlről és hátulról azonos beesési szögben érkező hangok között [WALLACH, 1939]. Informális meghallgatásos előkísérletünk szerint a kiválasztott hangfelvételen is érvényesül ez a jelenség, így használható az említett célra.

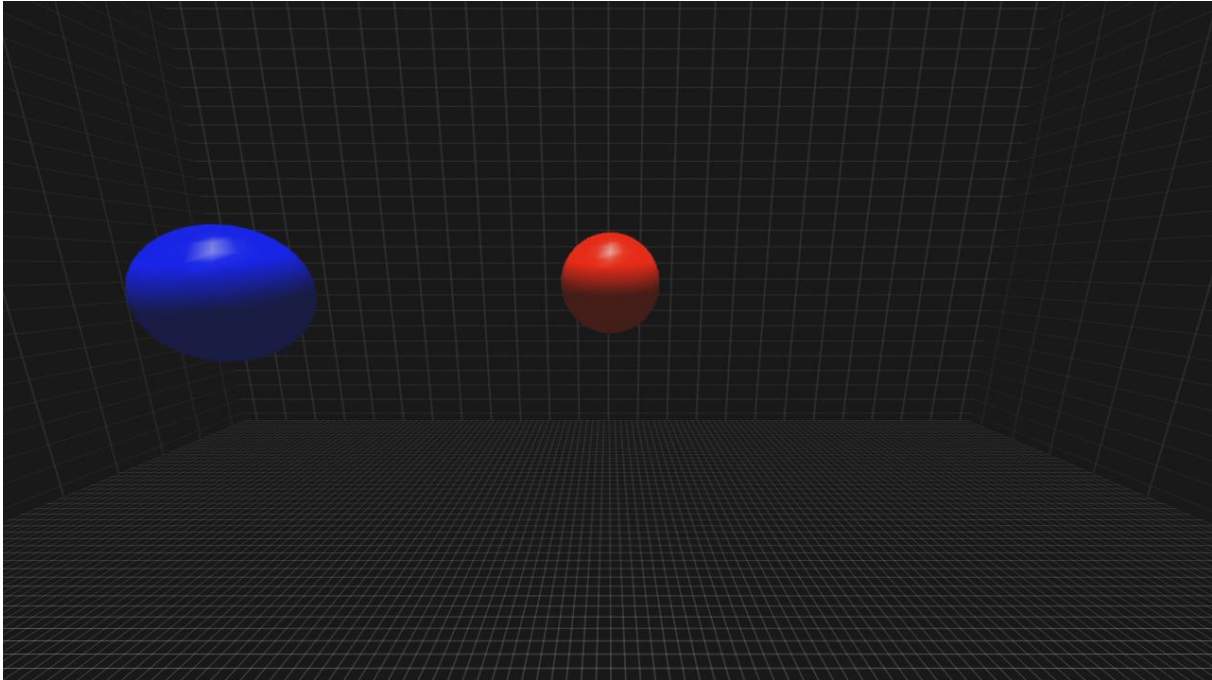
A hangfelvételekből az utófeldolgozás után 4 perces hangsorozatok készültek. Ezek olyan, egy percnél rövidebb részletek periodikus ismétléseiből álltak össze, melyek rögzítése során a legkevesebb probléma merült fel, és az elejük úgy volt kapcsolható a végük után, hogy a vágás minél kevésbé legyen észrevehető (az elmozdulásvektor és annak deriváltja minél kevésbé különbözzék). Az összeillesztés a 25 ms időtartamú szünetek alatt történt, 1 ms hosszúságú átúsztatásokkal (a nemkívánatos kattanások elkerülése végett).

Végül a kísérletek során az így kapott 96 kHz-es mintavételi frekvenciájú, 24 bites felbontású, tömörítetlen, kétcsatornás sztereó audiofájlokat használtunk hangingerként.

A képanyag előállításánál vált fontossá a hanganyagok rögzítésekor az egyes hangszórókhoz tartozó sisakok színe. A bejárások egy Microsoft Kinect mozgásérzékelő beviteli eszközzel is rögzítésre kerültek, másodpercenként körülbelül 14 frissítéssel. A Kinect infravörös érzékelője segítségével képes távolságokat megállapítani az általa belátott térben. Ezek alapján egy egyedi képfeldolgozó programmal határoztuk meg a mozgó tárgyak távolságinformációit, amely folyamatosan és viszonylag pontosan képes kinyerni és eltárolni a néhány méteren belül mozgó tárgyak, esetünkben a feltűnő színű sisakok pillanatnyi pozícióját. A felvételekhez időbélyeg tartozik, továbbá a kép-hang szinkronitás biztosítása érdekében a bejárások kezdetét egy taps jelezte, melynek segítségével a képanyag és a hangfelvétel időben pontosan összeilleszthető volt.

Az így kapott mozgási trajektóriákon természetesen szükség volt némi korrekcióra. Egyrészt a magasságinformációt fixnek vettük, mivel a függőleges irányérzékelésünk sokkal kevésbé érzékeny a vízszinteshez képest, és a segítők magasságkülönbsége és lépésekből eredő függőleges mozgásának látványa zavaró lett volna a kísérlet szempontjából. Másrészt előfordultak olyan esetek, amikor valamelyik sisak a Kinect számára takarásba került egy rövid időre, ilyenkor a megfelelő hangfelvétel téri ismertetőjegyei alapján lehetett következtetni a hangforrás mozgására, és ez alapján helyre lehetett állítani a hiányos trajektóriát.

A hangforrások vizuális megjelenítésére a Unity 3.4 videojáték-fejlesztő programot használtuk [<http://unity3d.com>]. A Kinect által rögzített adatokból kinyert mozgási trajektóriákat egy piros és egy kék színű, fejmagasságban lebegő gömb mozgatására használtuk fel egy virtuális szobában, ahol a nézőpont a műfej pozíciójában lett kijelölve. Ennek köszönhetően az animációk pontosan azt ábrázolták (lásd a 2.3 ábrát), ahogyan a műfej nézőpontjából láthatóak, illetve – a megfelelő hangminták hozzárendelése után – hallhatóak voltak a bejárások (pontos kép-hang szinkron esetén). Azokban az esetekben, amikor mindkét szinuszos hangjel egyetlen hangforrásból származott, egy félig piros, félig kék színű gömbbel ábrázoltuk a forrást.



2.3 ábra: Minta az egyik videóból

A videók felbontása 1280×720 pixel volt, képfrekvenciájuk 40 Hz, így egy képkocka időtartama 25 ms-nak adódott, amivel mind a 100 ms-os szinuszos hangok, mind a 25 ms-os szünetek hossza egész számú képkocka időtartamával egyezett meg. A négyperces videók elejére további 5 másodpercnyi, csak a virtuális termet ábrázoló videó került, ahová 4 másodperc után egy egyetlen képkocka (25 ms) hosszúságú hangjelet helyeztünk el. Utóbbinak később, az adatrögzítés pontosságának biztosítása során volt szerepe. Szükséges volt továbbá az is, hogy a gömbök a résztvevők által látható térrészben több alkalommal fehéren felvillanjanak; ennek szerepére szintén visszatérünk a későbbiekben.

A kísérlet során szükség volt olyan videókra is, ahol a képanyag nem egyezett meg a felvételek során rögzített valós trajektóriával. Az ilyen, ún. inkongruens videók esetében új mozgási trajektóriák készültek, egy videónként és résztvevőnként különböző mértékű geometriai transzformációsorozat segítségével (többek között forgatás a tér középpontja körül és tükrözés valamelyik tengely mentén). Az így kapott inkongruens mozgástrajektóriák statisztikai tulajdonságai megegyeznek az eredeti, kongruens trajektóriákkal (pl. az álló gömbök transzformálás után sem mozognak), azonban mégis

kellően különböznek tőlük ahhoz, hogy a látottak és hallottak összerendelése ne legyen egyértelmű (pl. az álló gömbök nem ott állnak, ahonnan a hangjuk szól).

A színes gömbök mozgatásán túl azok pulzálását is biztosítanunk kellett a kísérlet során a másodlagos kísérleti feladathoz (lásd 2.3.2 szakasz). Ezért ameddig az egyik hangforrás hangja szólt (100 ms, 4 képkocka), pontosan addig a hozzá tartozó gömb méretét is megnöveltük kissé, ezzel segítve a látottak és hallottak összekapcsolását.

A kísérlet során az ingeradó számítógép az animációk hangját egy Audiotrak Prodigy HD2 hangkártya sztereó kimenetén adta ki, ez egy egyedi építésű keverő-erősítőn haladt keresztül, majd a kísérleti személyekhez Etymotic Research ER-2 típusú, hallójáratba helyezhető fülhallgatón jutott el. Ez utóbbi speciális hallgatóeszköz biztosította azt, hogy a műfejes felvételek során rögzített, a fej, a vállak és a fülkagyló hangtartománybeli hatását (fejhez kötött átviteli függvény, head-related transfer function, HRTF) is tartalmazó hangfelvételeket pontosan úgy hallhassák a résztvevők, mintha a felvétel során ők ültek volna a mikrofon helyén, a terem középpontjában.

Az ingeradó számítógépben egy NVIDIA GeForce 9500 GT típusú videokártya volt megtalálható, melynek kimenő jele egy ATEN VS-132 videojel-elosztón (splitter) keresztül került a gép monitora mellett a vizsgálat során használt Samsung LE40C530F1W típusú, HD-képmegjelenítésre képes LCD televízióra is. A kísérleti személyek egy elsötétített és hangszigetelt helyiségben foglaltak helyet a kijelzőtől 35 cm távolságban, állukat egy álltámaszra helyezve. Ez utóbbi biztosította azt, hogy a vizsgálat során ne mozgassák, ne fordítsák el a fejüket, hiszen ez a látási és hallási ingeranyag egymástól való térbeli szétcsúszásához vezetett volna. A 35 cm-es távolság a megfelelő látószög biztosításához volt szükséges. A vizsgálat során egyedül tartózkodtak a helyiségben, a kísérletvezetővel mikrofonon keresztül tartották a kapcsolatot.

2.3 Kísérleti eljárás

2.3.1 Beállítás és betanítás

A kísérletben való részvétel előfeltétele volt a megfelelő hallás mindkét fülre, ennek vizsgálata egy hallásküszöbmérő eszközzel történt a korábban ismertettek szerint. Ezután a hangerősség beállítása következett egy a kísérletben is szereplő, álló helyzetű hangforrásokat tartalmazó hangrészlet felhasználásával: egyszerűsített lépcsőmódszerrel

(ami Békésy-féle lengőkiegyenlítéses módszerként is ismert) állapítottuk meg a hallásküszöböt, majd efelett 50 dB-lel állítottuk be a továbbiakban használt hangerősséget.

A kísérlet megkezdése előtt egy rövid, 20-25 perc hosszúságú betanításra volt szükség, hogy a résztvevőknek legyen idejük hozzászokni a kísérlet körülményeihez, illetve hogy a kép és a hang összerendelése biztossá váljon számukra, és valóban az animációban szereplő gömböket tekintsék a hangok forrásainak. Először 7, egyenként 20 másodperc hosszúságú videót néztek végig; mindegyik előtt elmondtuk, hogy mit fognak látni, majd megkértük őket, hogy számoljanak be a tapasztalataikról. Amennyiben minden érthető volt, továbbléptünk a következő animációra; arra az esetre, ha valami nem lett volna egyértelmű a résztvevő számára, lehetőség volt a kérdéses videók többszöri megtekintésére is. A betanítóvideók kondícióit a 2.2 táblázat mutatja.

A betanítóvideókban sorban előbb egy, majd két gömb volt látható, melyek kezdetben ritkán, majd egyre gyakrabban pulzáltak: 100 ms-on keresztül hangot adtak, miközben méretük is megváltozott. Először egyetlen, középen álló piros gömb volt látható, amely másodpercenként 1 alkalommal pulzált (ISI = 900 ms) (1), majd ellenpéldaként egy a hangtól jelentősen elcsúsztatott képanyagú videót is bemutattunk (2). Ezt követően egy gyakrabban (másodpercenként 2-szer, ISI = 400 ms) pulzáló gömb volt látható (3). Ezután kétgömbös példák következtek, ahol a piros gömb továbbra is középen, a kék gömb a kép bal szélén helyezkedett el. Ezt az elrendezést bemutattuk kisebb pulzálási gyakoriság mellett (4-4 pulzálás 3 másodperc alatt, ISI = 275 ms) (4), illetve a végleges kísérlet folyamán használt sebességen is (4-4 pulzálás másodpercenként, ISI = 25 ms) (5). Végül mozgó gömböket is bemutattunk az utóbbi két pulzálási gyakorisággal (6 és 7). Ha egy résztvevő számára elsőre nem volt világos a gömbök és a hangok összerendelhetősége, a videó egy-kétszeri megismétlése már legtöbbször elegendőnek bizonyult ehhez; erre jellemzően a mozgó gömbök esetén volt szükség. A videókhoz szükséges, ritkított hangingereket már meglévő hangfelvételek periodikus elnémításával készítettük el.

Ezután hasonló ingeranyaggal egy ellenőrző tesztet végeztünk a kísérleti személyeken, egyrészt annak ellenőrzésére, hogy képesek-e az audiovizuális szinkronitás megfelelő megítélésére, másrészt azért, hogy a képi és a hangingerek összekapcsolását tovább mélyítsük. Ennek során a résztvevőknek azt kellett eldönteniük, hogy az általuk látott, az 5. betanítóvideóhoz hasonló animációkban vajon szinkronban van-e egymással a

kép és a hang a gömbök pulzálása alapján. A két álló gömböt tartalmazó videók egy részénél ugyanis, míg a hangok megegyeztek az eredetivel, a képkockák számának módosításával a gömbök méretváltozásának sebességét gyorsabbra vagy lassabbra állítottuk be (az eredetileg 10 képkockából álló periodicitást ± 3 -mal változtattuk meg). Ennek köszönhetően a pulzálás képi és hangbeli frekvenciája eltérő volt, így a különbség mértékétől függő idő alatt teljesen elcsúszott egymástól a kép és a hang. A kísérleti személyek egy példa (8) és egy ellenpélda (9) igény szerinti számú megtekintése után kezdhettek bele a tesztbe (A). Amikor a legutolsó 8 videó szinkronitását helyesen (maximum egy hibával) állapították meg, ± 3 képkockáról ± 2 -re csökkentettük a periodicitás különbségét. A tényleges kísérletbe csak azok kezdhettek bele, akik ezen a nehezebb teszten (B) is jól (az utolsó nyolc videóban maximum két hibával) teljesítettek.

A betanításhoz használt videófájlok megtalálhatóak az [online](#) és a lemez mellékleten.

2.2 táblázat: A betanítóvideók kondíciói

#	Feladat	Hangforrások	Mozgás	Pulzálás	Szinkronitás
1	tanítás	1	álló	1 Hz	szinkron
2	tanítás	1	álló	1 Hz	aszinkron
3	tanítás	1	álló	2 Hz	szinkron
4	tanítás	2	álló	2,67 Hz	szinkron
5	tanítás	2	álló	8 Hz	szinkron
6	tanítás	2	mozgó	2,67 Hz	szinkron
7	tanítás	2	mozgó	8 Hz	szinkron
8	példa	2	álló	8 Hz	szinkron
9	példa	2	álló	8 Hz	aszinkron
A	teszt 1	2	álló	8 Hz	szink. / aszink.
B	teszt 2	2	álló	8 Hz	szink. / aszink.

2.3.2 Főkísérlet

A kísérlet fő részének elkezdése előtt néhány perces pihenési lehetőséget kaptak a résztvevők. Ezalatt ismertettük meg őket a feladatukkal, és ekkor hallhattak néhány hangpéldát is a későbbiekben megfigyelendő jelenségekre [DENHAM ÉS MUNKATÁRSAI, nyomtatás alatt]. Hasonló, pár perces szünetre legalább egy alkalommal a későbbiek folyamán is lehetőségük volt. A hangpéldák megtalálhatóak az [online](#) és a lemezmellékleten.

A kísérlet során a résztvevők összesen 12, egyenként 4:05 perc hosszúságú kísérleti blokkot végeztek el (egy videót néztek végig), egyet minden kísérleti kondícióhoz: egy-egy *kongruens* és *inkongruens* animációt a hatféle mozgási trajektória mindegyikéhez, véletlenszerű sorrendben. A kísérleti személyek elsődleges feladata annak folyamatos visszajelzése volt, hogy (a) észlelnek-e különböző magasságú hangokból álló ismétlődő mintázatot (integrált észlelet), (b) észlelnek-e azonos magasságú hangokból álló ismétlődő mintázatot (szegregált észlelet), vagy (c) esetleg ezek valamilyen kombinációját észlelik (vegyes észlelet). Ennek jelzésére a kezükben tartott nyomógombok használatával volt lehetőségük, melyet folyamatosan nyomva kellett tartaniuk, amíg az adott típusú észlelet fennállt, és váltaniuk kellett, amint az megváltozott. A bal és a jobb oldali gombok hozzárendelése az egyes válaszlehetőségekhez kiegyenlített volt a résztvevők között, hogy ezzel csökkentsük a kezességi hatást.

Az előkészületek során felhívtuk a résztvevők figyelmét arra, hogy nincs jó vagy rossz válasz: az általuk ténylegesen észlelt percepcióra vagyunk kíváncsiak. A blokkok utáni rövid szünetekben elmondhatták tapasztalataikat a látottakkal-hallottakkal kapcsolatban, ezzel is biztosítva, hogy megértették az instrukciókat, és valóban azt a nyomógombot nyomták, amelyikhez tartozó percepciót érzékelték.

A kísérleti személyek egy másodlagos, figyelemfenntartó feladatot is kaptak. A videóknak szereplő gömbök többször fehérén felvillantak a négyperces blokk alatt, a résztvevőknek ezen felvillanásokat kellett összeszámolniuk és a blokkok végén bediktálniuk. Ennek a feladatnak a célja az volt, hogy a kísérleti személyek figyelmét a képi ingereken tartsuk, hogy ne tudják kizárni azokat az észlelésükből, miközben a hangokra vonatkozó elsődleges feladatra koncentrálnak.

Az egyik résztvevő adatfelvételekor használt blokkok videófájljai megtalálhatóak az [online](#) és a lemezmellékleten.

2.4 Adatrögzítés és elemzés

A TTL nyomógombok jeleinek feldolgozása egy SynAmps² NeuroScan System 64-csatornás erősítőn keresztül történt, amely gameporton keresztül juttatta el a gombnyomások jelzéseit az ingeradó számítógépbe. A kísérlet lebonyolításáért felelős, MatLab (R2009b) környezetben futtatott program 25 ms-onként (képkockánként) rögzítette a nyomógombok pillanatnyi állapotát, illetve az aktuális időbélyeget; később az elemzés során ezekből határoztuk meg az egyes gombnyomások eloszlását és időtartamát. Az időbélyegek pontosságának biztosítására használtuk a videók 5. másodpercének legelején elhelyezett hangjelet, melynek megérkezése esetén egy Arduino Uno R2 mikrokontroller jelet küldött a program számára, és ez indította el az időzítőt.

Az elemzés során két eltérő gombnyomás által határolt észleleti fázisok ezredmásodpercben mért időtartamának tízes alapú logaritmusait és a jelzett észleleteket vizsgáltuk. A 300 ms alatti fázisokat nem elemeztük, mivel ezek feltehetőleg a gombok véletlen megnyomásból, illetve lenyomásának és elengedésének pontatlanságaiból erednek, mintsem két különböző, egymást gyorsan követő észleletváltásból [MORENO-BOTE ÉS MUNKATÁRSAI, 2010]. Ezen adatok alapján kiszámoltuk minden egyes észlelet átlagos arányát (egy blokkon belül jelzett észlelet időtartamának aránya a 4:05 perc hosszúságú blokk hosszához viszonyítva) és az észleleti fázisok átlagos hosszát, külön a résztvevők, észlelettipusok (integrált, szegregált és vegyes észleletek) és kondíciók szerint. Az „egyik sem” észleleteket (amikor a kísérleti személy egyik gombot sem tartotta lenyomva) nem elemeztük, mivel az összes kondícióban csupán a blokkok hosszainak 3,7 %-át tették ki, és hosszuk túlnyomó részben 1 mp alatt volt. Megjelenésük valószínűleg az észlelt szerveződés pillanatnyi szétesésének következménye (pl. észleletváltásokkor).

Ismételt méréses varianciaanalízist (repeated measures ANOVA) végeztünk a kapott hat adatsor (arány vagy időtartam \times integrált vagy szegregált vagy vegyes észlelet) mindegyikén, függő változókként a **Kongruencia**, a **Mozgás** és az **Elhelyezkedés** kategóriákat használva. **Kongruencia** szerint megkülönböztettünk *kongruens* és *inkongruens*, **Mozgás** szerint *álló* és *mozgó*, valamint **Elhelyezkedés** szerint *egyedül*, *együtt* és *külön* szinteket (lásd a 2.1 táblázatot).

A szabadsági fokokat lehetőség szerint a Greenhouse—Geisser korrekció elvégzésével módosítottuk. Ezek (1-től eltérő) ε értékeit, valamint a részleges η^2 -értékeket is feltüntettük. A post-hoc összehasonlításokat a Tukey-teszt segítségével végeztük. Az elemzések 95%-os konfidenciaszinttel készültek.

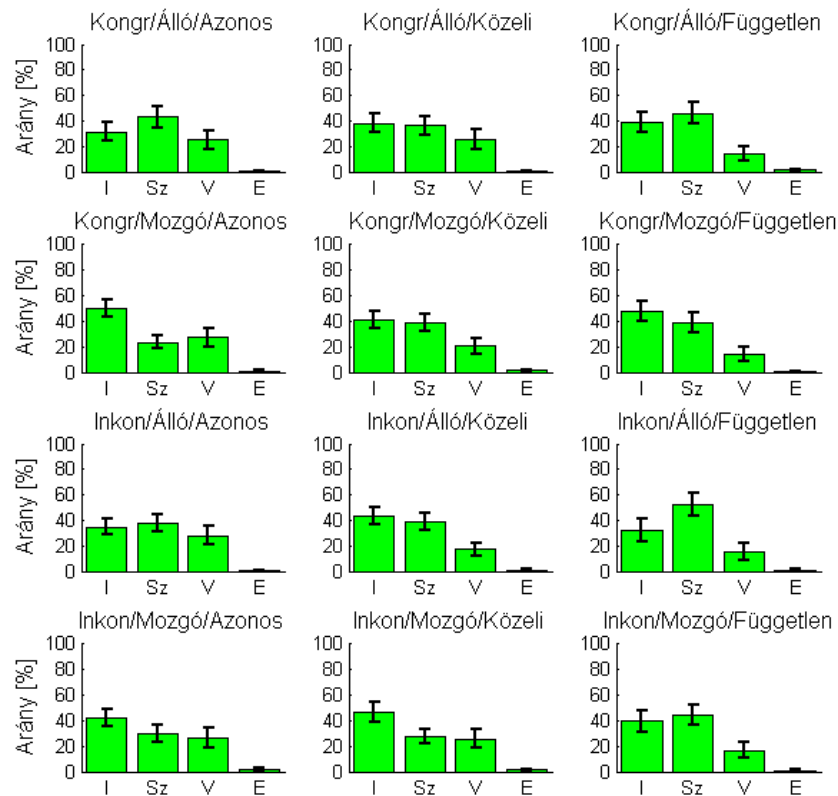
A bevezetésben tárgyalt hipotéziseket a következő hatások illetve interakciók megjelenése tükrözheti elemzésünkben:

Kongruencia \times **Relatív helyzet** interakciója: *Együtt* álló ill. mozgó hangforrásokkal *kongruens* vizuális ingerek megnövelik az integrált észleletek arányát az *inkongruens* esetekkel szemben, míg az egymástól külön álló ill. *függetlenül* mozgó hangokkal *kongruens* vizuális ingerek esetében a szegregált észleletek aránya nő meg *inkongruens* párjaikkal szemben (1-2. hipotézis).

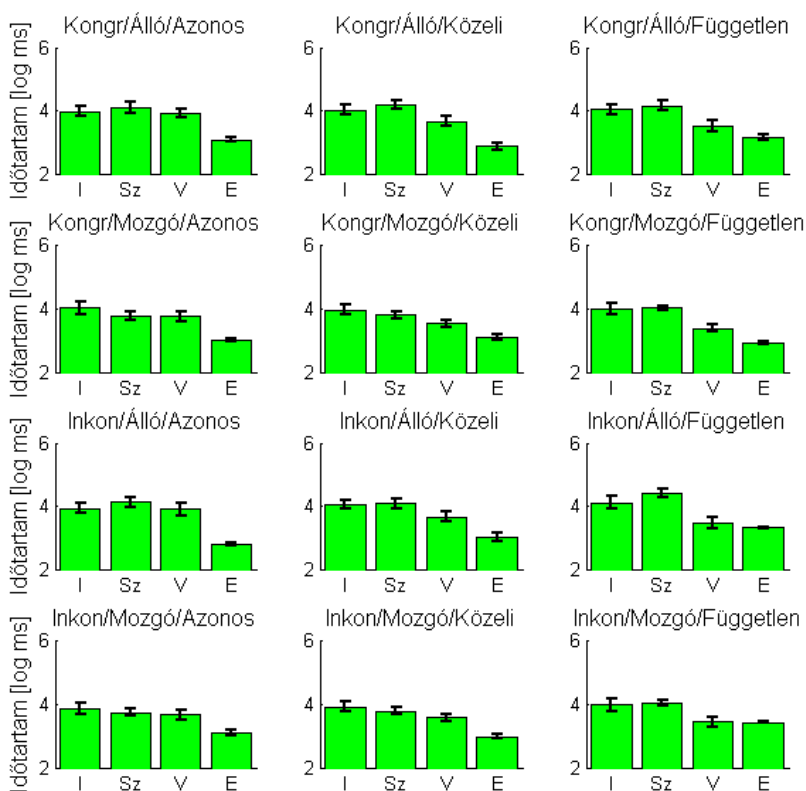
Kongruencia főhatás: A *kongruens* hallott és látott mozgások kondícióiban megnőnek az egyes észlelettípusok átlagos fázishosszai az *inkongruens* kondíciókhoz képest (3. hipotézis).

Mozgás \times **Relatív helyzet** interakciója: Az együtt mozgó hangforrások az együtt állóknál nagyobb arányban váltanak ki integrált észleletet, míg az egymástól függetlenül mozgó hangforrások a szegregált észleletek arányát növelik meg az egymástól külön állókhoz képest (4-5. hipotézis).

3 Eredmények



3.1 ábra: Az egyes észleletek időeloszlása kondíciónként, ahol a különböző észleletek: Integrált, Szegregált, Vegyes és Egyik sem; a **Kongruencia**-szintek: *kongruens*, *inkongruens*; a **Mozgási** szintek: *álló*, *mozgó*; a **Relatív helyzet** szintjei: *azonos*, *közeli*, *független*



3.2 ábra: Az egyes észleletek átlagos fázishossza kondíciónként, ahol a különböző észleletek: Integrált, Szegregált, Vegyes és Egyik sem; a **Kongruencia**-szintek: *kongruens*, *inkongruens*; a **Mozgási** szintek: *álló*, *mozgó*; a **Relatív helyzet** szintjei: *azonos*, *közeli*, *független*

A fenti ábrákon a kísérlet során a résztvevők által folyamatosan visszajelzett észleletek összesítése látható. A 3.1 ábra az egyes percepciók százalékos eloszlását mutatja a különböző kondíciókapcsolatokra, míg a 3.2 ábra ugyanezekre az észleletek átlagos hosszát adja meg; a hibavonalak standard hibának felelnek meg.

Elvégeztük a három függő változó ismételt méréses varianciaanalízisét (ANOVA), melyhez a Statistica 11.0 programot használtuk. Az időeloszlásokra vonatkozó eredményeket a 3.1 táblázat, az átlagos fázishosszokra vonatkozó eredményeket a 3.2 táblázat tartalmazza. Ezek alapján a következő eredményeket fogalmazhatjuk meg:

A **Kongruencia** hatását egy esetben sem tudtuk egyértelműen kimutatni, meg sem közelítették a várt $p < 0,05$ -ös szignifikanciaszintet.

A **Mozgás** jelentős hatással volt a szegregált és a vegyes észleletek átlagos időtartamára. A post-hoc tesztek eredményei alapján szignifikáns különbség mutatkozott az említett észleletek esetében: a mozgó hangforrások átlagosan rövidebb észlelési időtartamokat eredményeztek, mint az álló források ($dF = 1$, $p < 0,001$ a szegregált észleletek átlagos időtartamának eltérésére az álló és a mozgó eset között; $dF = 1$, $p < 0,05$ a vegyes észleletek átlagos időtartamának eltérésére az álló és a mozgó eset között).

Bár a **Relatív helyzet** hatása sehol nem érte el a szignifikanciaszintet, a szegregált és vegyes észleletek átlagos hossza és eloszlása esetében megközelítette azt.

A három függő változó sem páronként, sem együttesen nem mutatott **interakciót**.

3.1 táblázat: A varianciaanalízis (ANOVA) időeloszlásokra vonatkozó eredményei. A „Mérés” oszlop a különböző észleleteket sorolja fel: Integrált [%], Szegregált [%] és Vegyes [%]. A „Függő változó” oszlop tartalmazza a **Kongruencia, Mozgás és Relatív helyzet** változókat, illetve ezek különböző, modalitások közötti interakcióit. A további oszlopokban „dF” jelöli a korrigált szabadsági fokot (G-G adjusted degree of freedom), „F” az F értéket, „p” a korrigált szignifikanciaszintet, „ ϵ ” a Greenhouse—Geisser korrekciós tényezőt, „ η^2 ” pedig a hatás nagyságát.

Mérés	Függő változó	dF	F	p	G-G ϵ	η^2
Integrált [%]	Kongruencia	1; 18	0,179	0,677		0,010
	Mozgás	1; 18	1,887	0,186		0,095
	Relatív helyzet	1,39; 25,00	0,217	0,725	0,694	0,012
	Kongruencia \times Mozgás	1; 18	0,463	0,505		0,025
	Kongruencia \times Relatív helyzet	1,54; 27,73	1,754	0,196	0,770	0,089
	Mozgás \times Relatív helyzet	1,98; 35,72	1,269	0,293	0,992	0,066
	Kongr. \times Mozg. \times Rel.helyzet	1,38; 24,82	0,354	0,626	0,689	0,019
Szegregált [%]	Kongruencia	1; 18	0,097	0,759		0,005
	Mozgás	1; 18	3,035	0,099		0,144
	Relatív helyzet	1,22; 22,02	3,155	0,083	0,612	0,149
	Kongruencia \times Mozgás	1; 18	0,004	0,951		0,000
	Kongruencia \times Relatív helyzet	1,99; 35,80	1,649	0,207	0,995	0,084
	Mozgás \times Relatív helyzet	1,97; 35,48	0,971	0,388	0,986	0,051
	Kongr. \times Mozg. \times Rel.helyzet	1,99; 35,76	1,676	0,202	0,993	0,085
Vegyes [%]	Kongruencia	1; 18	0,023	0,881		0,001
	Mozgás	1; 18	0,132	0,720		0,007
	Relatív helyzet	1,46; 26,30	2,867	0,089	0,730	0,137
	Kongruencia \times Mozgás	1; 18	1,209	0,286		0,063
	Kongruencia \times Relatív helyzet	1,73; 31,08	0,167	0,816	0,863	0,009
	Mozgás \times Relatív helyzet	1,94; 35,00	0,042	0,956	0,972	0,002
	Kongr. \times Mozg. \times Rel.helyzet	1,94; 34,84	1,244	0,300	0,968	0,065

3.2 táblázat: A varianciaanalízis (ANOVA) átlagos fázishosszakra vonatkozó eredményei. A varianciaanalízis során talált szignifikáns (illetve közel szignifikáns) hatásokat a táblázat **félkövérrel** szedett sorai mutatják. A „Mérés” oszlop a különböző észleleteket sorolja fel: Integrált [log ms], Szegregált [log ms] és Vegyes [log ms]. A „Függő változó” oszlop tartalmazza a **Kongruencia, Mozgás és Relatív helyzet** változókat, illetve ezek különböző, modalitások közötti interakcióit. A további oszlopokban „dF” jelöli a korrigált szabadsági fokot (G-G adjusted degree of freedom), „F” az F értéket, „p” a korrigált szignifikanciaszintet, „ε” a Greenhouse—Geisser korrekciós tényezőt, „η²” pedig a hatás nagyságát.

Mérés	Függő változó	dF	F	p	G-G ε	η ²
Integrált [log ms]	Kongruencia	1; 18	0,923	0,349		0,049
	Mozgás	1; 18	1,740	0,204		0,088
	Relatív helyzet	1,56; 28,17	0,184	0,779	0,782	0,010
	Kongruencia × Mozgás	1; 18	0,163	0,691		0,009
	Kongruencia × Relatív helyzet	1,67; 30,00	0,433	0,616	0,833	0,024
	Mozgás × Relatív helyzet	1,75; 31,47	0,022	0,967	0,874	0,001
	Kongr. × Mozg. × Rel.helyzet	1,85; 33,34	0,126	0,867	0,926	0,007
Szegregált [log ms]	Kongruencia	1; 18	1,842	0,191		0,093
	Mozgás	1; 18	19,862	<0,001		0,525
	Relatív helyzet	1,35; 24,28	3,413	0,066	0,674	0,159
	Kongruencia × Mozgás	1; 18	0,837	0,372		0,044
	Kongruencia × Relatív helyzet	1,98; 35,58	1,483	0,241	0,988	0,076
	Mozgás × Relatív helyzet	1,47; 26,40	1,264	0,289	0,733	0,066
	Kongr. × Mozg. × Rel.helyzet	1,28; 23,09	0,247	0,683	0,641	0,014
Vegyes [log ms]	Kongruencia	1; 18	0,094	0,762		0,005
	Mozgás	1; 18	4,858	0,041		0,213
	Relatív helyzet	1,43; 25,66	2,809	0,094	0,713	0,135
	Kongruencia × Mozgás	1; 18	0,000	0,993		0,000
	Kongruencia × Relatív helyzet	1,84; 33,08	0,506	0,592	0,919	0,027
	Mozgás × Relatív helyzet	1,87; 33,57	0,204	0,802	0,933	0,011
	Kongr. × Mozg. × Rel.helyzet	1,69; 30,38	0,606	0,525	0,844	0,033

4 Értelmezés és következtetések

Jelen kísérlet célja az volt, hogy megállapítsuk, hatással van-e a mozgó forrású hangok észlelésére, ha azokhoz vizuális ingerek is társulnak.

Azt feltételeztük, hogy azokban az esetekben, amikor a hangforrások együtt mozognak (*együtt* mozgási kondíció), az ezekkel kongruens (*kongruens* kondíció) vizuális ingerek hatására az integrált, egymástól függetlenül mozgó hangforrások esetén (*külön* kondíció) pedig a szegregált percepció aránya jelentősen megnő az adott kondíciók inkongruens párhuzamához képest. Ezzel szemben a varianciaanalízis során a **Mozgási és Relatív helyzet** faktorok között szignifikáns interakciót nem találtunk.

Ezen felül stabilabb percepciók kialakulását vártuk a hangokkal kongruens vizuális ingerek hatására. Ezek hosszabb fázisidőtartamokban mutatkoztak volna meg a **Kongruencia** főhatása miatt, de elemzésünk szerint ez a hatás elenyésző volt és távol maradt a szignifikanciától.

A gondos előkészületek – a résztvevők alapos betanítása és a tényleges kísérletet megelőző tesztelése (előválogatása) – feltételezhetően biztosították, hogy a kísérleti személyek képesek lesznek a vizsgálat során a hallott és látott ingerek összekapcsolására. Így, bár a szignifikáns eredmények hiánya nem utal egyértelműen a várt jelenség hiányára, valószínűtlen, hogy a fenti hatásokat, amennyiben léteznek, ne mutattuk volna ki.

Ezek az eredmények nem felelnek meg az elvárásainknak, hiszen azt mutatják, hogy a vizuálisan érzékelt tárgyak nem segítik elő a hozzájuk kapcsolható hallási tárgyak stabilizálódását. Ezzel ellentétben viszont összhangban vannak azokkal a kutatási eredményekkel, melyek szerint a látórendszer által kialakított érzetek nem befolyásolják a hallórendszer működését (legalábbis a hallott tárgyak helyzetére vonatkozóan; lásd KUBOVI ÉS VAN VALKENBURG [2001]).

Vártuk továbbá a Gestalt „közös sors” elve alapján, hogy az egymáshoz közeli pályán mozgó hangok megnövelik a kialakult integrált észleletek arányát az egymáshoz közel álló hangokhoz képest, valamint, hogy az egymástól függetlenül mozgó hangok megnövelik a szegregált észleletek arányát az egymástól külön álló hangokhoz képest. A „közös sors” elvének ilyen megjelenése a **Mozgás** és **Relatív helyzet** faktorok interakciójában lehetne tettenérhető, de ilyen szignifikáns interakciót egyetlen esetben sem találtunk. Ez egyrészt megegyezik a BÖHM ÉS MUNKATÁRSAI [nyomtatás alatt] által kapott

eredményekkel, miszerint a hallórendszer képtelennek tűnik felhasználni a hallott mozgást, mint ismertetőjegyet a hallási láncok szétválasztására, jelen eredményeink alapján még abban az esetben is, ha a hallott mozgást vizuálisan is követhetővé tettük.

Kísérletünk eredményei alapján két további, nem várt hatást is megfigyeltünk. Az egyik, szignifikanciát el nem érő, de azt minden esetben nagy hatással közelítő főhatást a **Relatív helyzet** mutatta, a szegregált észleletek időtartamának esetében majdnem átlépvén a szignifikancia szintjét. Ez a hatás hasonlít a BÖHM ÉS MUNKATÁRSAI [nyomtatás alatt] által egy vizuális ingerek nélküli, de egyébként hasonló paradigmában kimutatott szignifikáns főhatásra. Mindkét esetben feltételezhető a „közelség” Gestalt-elvének érvényesülése, mely szerint az egymáshoz közel található tárgyak összekapcsolása, az egymástól távolabbi tárgyaknak pedig a szétválasztása a könnyebb feladat. Érdeemes volna több kísérleti személyt bevonni az adatfelvételbe, hogy ennek a nagy hatású, de szignifikanciát el nem érő jelenségnek a meglétét egyértelműen igazolni vagy cáfolni tudjuk.

A másik megfigyelt hatás a **Mozgás** főhatása volt, ami a szegregált és vegyes észleletek esetében szignifikánsan, de egyéb esetekben is jellemzően csökkentették az egyes észleletek átlagos hosszát. Ezt a hatást szintén kimutatta BÖHM ÉS MUNKATÁRSAI [nyomtatás alatt] kísérlete is. Az észleletek rövidebb volta azt mutatja, hogy a résztvevők gyakrabban váltottak két észlelettípus között mozgó tárgyak esetén, mint álló tárgyaknál. Ez alapján azt mondhatjuk, hogy a mozgás nemhogy elősegítette, de több helyen egyenesen gátolta a stabil hallási tárgyak kialakulását.

Kísérletünkkel azt vizsgáltuk, milyen módon befolyásolja a mozgó hangforrások alapján kialakított hallási tárgyak stabilitását, ha ezekhez velük együtt vagy tőlük függetlenül mozgó vizuális ingereket kapcsolunk. A kapott eredmények bár nem felelnek meg a körültekintően megalapozott és megfogalmazott elvárásainknak, emellett viszont összhangban vannak más, hasonló jelenségeket vizsgáló kutatások eredményeivel.

Mindenképpen érdemesnek tartjuk ezt a multimodális jelenséget akár ugyanennek a paradigmának a több résztvevő bevonásával történő vizsgálatát a kis elemszámból adódó bizonytalanságok egyértelműsítésére, akár a jelenség más paradigmákba ágyazott vizsgálatát az elvárások és az eredmények disszociációjának átfogóbb kísérleti felderítésére.

Mellékletek

Az [online](#) és a lemezmellékleten megtalálhatóak a betanítás során használt videófájlok, a főkísérlet megkezdése előtt bemutatott hangpéldák és az egyik résztvevő adatfelvételekor használt blokkok videófájljai is.

Irodalomjegyzék

Böhm, T. M., Shestopalova, L., Bendixen, A., Andreou, A. G., Georgiou, J., Garreau, G., Pouliquen, P., Cassidy, A., Denham, S. L., & Winkler, I. The role of perceived source location in auditory stream segregation: Separation affects sound organization, common fate does not. *Learning and Perception*, (nyomtatás alatt).

Böhm, T. M., Shestopalova, L., Garreau, G., Andrijevic, D., Georgiou, J., Andreou, A., Bendixen, A., Denham, S. L., & Winkler, I.: Do visual cues of motion affect the perceptual organization of sound sources moving in space? *Making Sense of Sounds 2012 – Workshop on Sounds and Sound Processing in Natural and Artificial Systems*, Plymouth, 2012. február 20-21.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, Massachusetts: MIT Press.

Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia*, 45, 523-530.

Denham, S. L., Gyimesi, K., Stefanics, G., & Winkler, I. Perceptual bi-stability in auditory streaming: How much do stimulus features matter? *Learning and Perception*, (nyomtatás alatt).

Georgiou, J., Pouliquen, P., Cassidy, A., Garreau, G., Andreou, C., Stuarts, G., d'Urbal, C., Andreou, A. G., Denham, S., Wennekers, T., Mill, R., Winkler, I., Böhm, T. M., Szalárdy, O., Klump, G. M., Jones, S., & Bendixen, A. (2011). A multimodal-corpus data collection system for cognitive acoustic scene analysis. In *45th Annual Conference on Information Sciences and Systems (CISS 2011)* (pp. 1-6).

Grantham, D. W. (1995). Spatial hearing and related phenomena. In B.C.J. Moore (Ed.), *Hearing* (pp. 297-346). San Diego: Academic Press.

Judd, T. (1977). An explanation of Deutsch's scale illusion. Unpublished manuscript. Department of Psychology, Cornell University.

Köhler, W. (1947). *Gestalt Psychology*. (2 ed.) New York: Liveright.

Kubovi, M. & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97-126.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, 166, 538-547.

Middlebrooks, J. C. & Green, D. M. (1991). Sound Localization by Human Listeners. *Annual Review of Psychology*, 42, 135-159.

Moore, B. C. J. & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88, 320-333.

Moreno-Bote, R., Shpiro, A., Rinzel, J., & Rubin, N. (2010). Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *Journal of Vision*, 10, 1-18.

Noesselt, T., Fendrich, R., Bonath, B., Tyll, S. & Heinze, H. J. (2005). Closer in time when farther in space: Spatial factors in audiovisual temporal integration. *Cognitive Brain Research*, 25, 443-458.

Perrott, D. R. & Marlborough, K. (1989). Minimum Audible Movement Angle - Marking the End-Points of the Path Traveled by A Moving Sound Source. *Journal of the Acoustical Society of America*, 85, 1773-1775.

Pressnitzer, D. & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16, 1351-1357.

Smith, J., Hausfeld, S., Power, R. P., & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Percept.Psychophys.*, 32, 454-464.

Spence, C. & Driver, J. (1997). Audiovisual links in exogenous overt spatial orienting. *Perception & Psychophysics*, 59, 1-22.

Szalárdy, O., Bendixen, A., Tóth, D., Denham, S. L., & Winkler, I. Modulation frequency acts as a primary cue for auditory stream segregation. *Learning and Perception*, (nyomtatás alatt).

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Ph.D. Eindhoven University of Technology, Leiden, The Netherlands.

Wallach, H. (1939). On Sound Localization. *Journal of the Acoustical Society of America*, 10(4), 270-274.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, 21, 147-163.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13, 532-540.