A hippokampális formáció funkcionális modellezése

Doktori disszertáció

Szirtes Gábor

Témavezető

dr. habil. Lőrincz András

Eötvös Loránd Tudományegyetem, Informatika Kar

Információs Rendszerek Tanszék

2004

Köszönetnyilvánítás

A tudományos publikációkkal ellentétben a disszertáción csak az én nevem szerepelhet, mégis, ha bármilyen igazi teljesítményt mutatnak e lapok, akkor az valójában rengeteg ember összefogásának és segítségének eredménye (a hibákért viszont csak engem illet felelősség). Nagyon szerencsésnek gondolom magam, mert fel sem tudom sorolni mindazok nevét, akiktől életem minden fontosabb pillanatában jószándékot és támogatást kaptam. Ezért remélem, azok sem sértődnek meg, akiket nem említek.

Szüleimmel kezdem és legszívesebben a végén is hozzájuk térnék vissza, mert azt a szeretetet és bizalmat, amit mindig is kaptam tőlük, úgysem tudom szavakban megköszönni. Köszönöm feleségemnek, Anikónak, hogy nem fél a bizonytalantól és hisz bennem. Minden volt tanáromnak köszönöm, hogy hagyták, hogy a saját utamat járjam. Szeretném megköszönni Reményi Piroskának, hogy megmutatta, a számítógép is olyan mint egy nagyító vagy egy laboratórum: lehetővé teszi, hogy gyönyörködhessünk a világban. Köszönöm Simon Istvánnak és az Enzimológia Intézetnek, hogy amikor gondban voltam, befogadtak, amikor mennem kellett, elengedtek.

Bár idegen területről érkeztem, jó szívvel fogadtak az Információs Rendszerek tanszéken, amiért hálás vagyok. A csoportomról csak annyit: fantasztikus veletek dolgozni! Szerettem azt az intellektuális pezsgést, amit Póczos Barnabással, Szatmáry Botonddal és Takács Bálinttal folytatott „szoba-csevegéseink" alatt éreztem. Örültem, hogy megismerhettem Palotai Zsoltot, remélem, ragadt rám valami az ő fegyelmezettségéből és kitartásából. Hévizi Gyurinak pedig köszönöm, hogy olyan sokszor feltételezte, hogy vagyok olyan okos, mint ő.

Nyilvánvalóan az embernek meg kell köszönni témavezetője támogatását. Én azonban egy hivatalos köszönetnél sokkal többet szeretnék mondani, mert Lőrincz Andrástól sokkal többet kaptam, mint puszta támogatást. Egyetemi tanulmányaim alatt általa nyerhettem betekintést a mesterséges intelligencia és a neurobiológia különböző területeire és ő biztatott, hogy járjak utána mindennek, amit tudni szeretnék, de sosem merem megkérdezni. Ő mindig bízott bennem, hiányos háttértudásom ellenére is felvett doktoranduszának, tanított, kérdezett, noszogatott, mikor mi kellett. Remélem, nem bánta meg választását és később is büszke lehet rám.

Tartalomjegyzék

1. Irodalmi összefoglaló és tudományos célkitűzések

Munkám során az emlős agyban található hippokampusz és környékének az emlékezetben és a tanulásban betöltött feladatainak funkcionális matematikai modellezésével foglalkoztam. Ezen régiók együttese, az úgynevezett hippokampális formáció az emlős agy egyik legtüzetesebben vizsgált területe, amely magában foglalja az ősi, limbikus kéreghez tartozó hippokampuszt és az entorhinális kérget, mely összeköttetést teremt a neokortex és a hippokampusz között. Nagyszámú kísérlet és tapasztalat bizonyította, hogy e területek kulcsfontosságú szerepet töltenek be az emlékezet és tanulás különböző folyamataiban, sérülésük retrográd (emlékek eltűnése) és anterográd (új információk raktározási képességének sérülése) amnéziához vezethet. Szintén e területek leépülése vagy rendellenes működése figyelhető meg a manapság mind több embert érintő Alzheimer- és Parkinson-kórban, valamint a temporális lebenyben jelentkező epilepszia esetében. Nemcsak neurobiológiai, orvosi szempontból fontos e régió vizsgálat, hiszen különleges kapcsolatrendszere, alrégióinak bonyolult aktivitás-mintázata elméleti szempontból is igen izgalmas. Mivel az emlős agy memória szerveződése egészen más jellegű, mint az ember tervezte számítógépeké, ezért feltehető, hogy saját agyunk vizsgálata akár hatékonyabb számítógépes rendszerek tervezését is lehetővé teszi. Ráadásul a különböző agyterületek fejlődése önszervező jellegű, így tanulmányozásukkal talán közelebb kerülhetünk a ténylegesen adaptív rendszerek megértéséhez. Mivel az emlékezet és tanulás folyamatai nem választhatók el sem az érzékeléstől, sem a cselekvéstől, ezért e folyamatok kölcsönhatásainak vizsgálata például egészen új ember-számítógép interfészek tervezését is lehetőség teszi.

Az agy funkcionális modellezése interdiszciplináris kutatási terület, mely egyrészt a szélesebb értelemben vett mesterséges intelligencia kutatás célkitűzéseit és eszköztárát használja, másrészt az idegtudományok (neurobiológia, neurofiziológia és kognitív tudományok) tudástárára épít. E tudományos területen belül több megközelítés létezik. Az általam is követett irányzat az agyat és az agy által végzett folyamatokat az evolúció mérnöki eredményeiként értelmezi és alapvető célja egy „konkurens” tervezési folyamat levezetése. Más szóval axiómarendszereket keresünk, amelyek meghatározzák egy, a környezetével dinamikus kölcsönhatásban lévő rendszer szükséges funkcióit, majd e funkciókhoz szükséges számítási módokat, algoritmusokat a tényleges környezet jelentette kényszerek által próbáljuk szűkíteni, pontosítani. A kényszereket a biológiai, az információelméleti

1. ábra. **Információáramlás az entorhinális kéreg és a hippokampusz alkotta hurokban. (A)** Az EC-HC hurok általunk modellezett területei és legfontosabb kapcsolataik. Míg az entorhinális kéreg (EC) a neokortexhez hasonlóan többé-kevésbé jól szeparálható rétegekből épül fel, a hippokampusz (HC) különböző, egymástól gyakran nehezen elválasztható régiókból áll, melyek egyes fajoknál akár teljesen azonosíthatatlanok. Az EC rétegeit római számokkal jelölik. DG: gyrus dentatus, CA: Ammon-szarv. **(B)** A funkcionális modell leképezése. A gyrus dentatus szerepe feltehetőleg az észlelésben fellépő időbeli késések javítása, időbeli dekonvolúció [Lőrincz és Buzsáki, 2000]. Mivel téziseimben e területtel nem foglalkozom, ezért kihagytam a leképezésből. ICA: független komponens analízis. HTM: a közép- illetve hosszútávú memória feltételezett helye. Rek. input: rekonstruált (visszaállított) input.

és a matematikai ismeretek adják közösen. Ezen a módon az absztrakttól a létezőig haladunk, ezért ezt a megközelítést „bottom-up" kényszerekkel korlátozott „top-down" modellezésnek is hívhatjuk. Téziseimben a Lőrincz András és Buzsáki György által javasolt emlős memóriamodell (továbbiakban L-B modell) [Lőrincz és Buzsáki, 2000] kiterjesztésére tettem kísérletet, ezért a következő alfejezetben röviden összefoglalom az L-B modellt és központi feltevéseit.

1.1. A hippokampális formáció Lőrincz-Buzsáki féle funkcionális modellje

A fontosabb rétegek és alrégiók kapcsolati struktúráját és az információ modell szerinti áramlását szemlélteti az 1. ábra:

- A szenzoros információfeldolgozás célja, hogy elősegítse, pontosítsa a belső (rejtett) modelleket, melyek a rendszer akcióinak tervezésében vesznek részt, illetve a környe-

zet változásait jósolják.

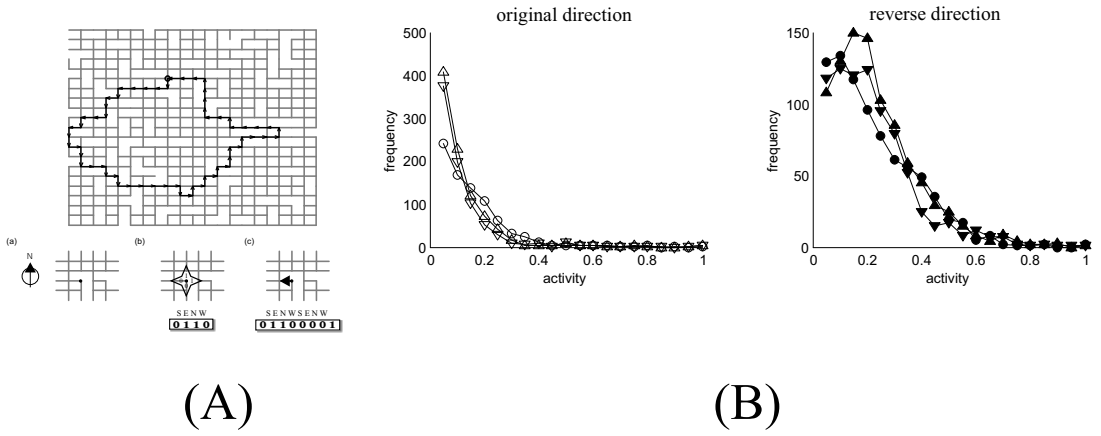
- A hatékony információfeldolgozás és -továbbítás érdekében érdemes úgynevezett faktoriális kódokat keresni.
- Párhuzamos és hierarchikus felépítés és működés jellemzi a rendszert.
- A belső modell és a külvilág változásai állandó összehasonlítást igényelnek: a „komparátor-elméletet" [Grastyán és mtsai., 1959] követve feltételezhető, hogy a memóriaszerveződésért felelős központi rész, a hurokba szerveződött entorhinális kéreg és a hippokampusz lényegében egy dinamikus rekonstrukciós hálót képez, mely folyamatosan összeveti a beérkező jeleket a jósolt, „várt" jelekkel. A számolt eltérések alakítják a belső reprezentációt, valamint ezek képezik a magasabb feldolgozási szintek bemenő jelét. A rekonstrukciós hálóban az input belső reprezentációját iterációs dinamika alakítja, melynek célja az input és a rekonstruált input közötti eltérés csökkentése.
- A biológiai leképezés során (azaz a különböző funkciók anatómiai területekhez való hozzárendelésekor) mindenképpen figyelembe kell venni, hogy
  - a hippokampusz részterületei és a környező területek több, párhuzamosan futó hurkot képeznek,
  - lényegesen több lefelemenő kapcsolat létezik, ugyanakkor e kapcsolatok csak ritkán aktívak,
  - pszichofizikai kísérletek szerint mind ún. „gyors", mind „lassú" tanulásra képes a vizsgált terület,
  - mindenféle plaszticitás (tanulást lehetővé tévő kapcsolati átrendeződés) legfontosabb formája az ún. Hebb-tanulás
  - az idegsejtek többféle szinkron működésre képesek, melyek pontos (de még nem értett) kölcsönhatása szükséges a sikeres tanuláshoz.

E modell kiterjesztését és vizsgálatát kétféle kényszer határozta meg. A biológiai kényszerek, melyek egyrészt anatómiai (struktúrális), másrészt neurofiziológiai (funkcionális) jellegűek, korlátozzák a „top-down" modellezési eljárásunk során hipotézisként megfogalmazott szükséges funkciók lehetséges megvalósítását. Ha azonban úgy találtuk, hogy az

elképzelt funkció neurálisan megvalósítható az adott biológiai ismereteink alapján, akkor a sikeresen leképezett funkció támaszt további (funkcionális) követelményeket, melyeket ismét le kell tudnunk képezni az adott biológiai rendszerre. A következő tézisek e hosszú folyamat különböző állomásait jelentik. Főbb céljaim az L-B modell egységes elméleti hátterének kidolgozása és a javasolt funkciók pontosabb összeillesztése, leképezése volt. A következő pontokban a fontosabb eredményeket kíséreltem meg egységes keretben összefoglalni, a részletes matematikai és szimulációs eredmények a csatolt publikációkban találhatók meg.

2. Eredmények I. A Lőrincz-Buzsáki modell kiterjesztése temporális független komponens analízis kódolással

Egyes vélemények szerint (pl. [Barlow, 1972, Field, 1987]) az idegrendszer faktoriális kódokat használ a szenzoros jelfeldolgozás során az információ szűréséhez és további feldolgozásához. Különböző vizsgálatok megmutatták [Field, 1994, Olshausen és Field, 1996, Bell és Sejnowski, 1997, Olshausen és Field, 1997], hogy faktoriális kódolást kialakító ún. független komponens analízis (ICA, pl. [Comon, 1994]) módszerek segítségével és különböző ritkasági kényszerek használata mellett természetes kép inputokon olyan szűrők alakultak ki, melyek szerkezete hasonlít az emlősök elsődleges látókérgét meghatározó sejtek ún. receptormezőihez. (Az ICA algoritmusok képesek a komponensek között a másodrendű korrelációk mellett a magasabb rendű korrelációkat is eltűntetni, így maximalizálva az információtovábbítás hatékonyságát.) Hasonló tulajdonsággal rendelkeznek a hippokampusz különböző részein megtalálható un. „tér-sejtek" is: ezek a sejtek akkor aktívak, amikor az állat a tér egy bizonyos pontja körül van, illetve áthalad e ponton („tér-mezők"). Érdemes megjegyezni, hogy éppen e különleges sejtek jelenléte miatt gondolták sokáig, hogy a hippokampusz elsősorban a térbeli mozgáshoz szükséges memóriáért felelős. Bár ICA algoritmusok megvalósíthatók neurális formában (azaz olyan mesterséges neuronhálókkal, melyekben a transzformációk lokálisak és a tanulási szabályok a neurobiológiai tapasztalatoknak megfelelnek), működésükhöz szükséges, hogy a bemenő inputokat egymástól függetlenül, véletlenszerűen válasszuk. E feltétel természetesen biológiai rendszerek esetében állandóan változó környezetben nem tartható, ráadásul a bemenő információk egyidejű raktározását is igényelné. Ennek alapján felmerül a kérdés, hogy az L-B modell alapján megjósolt faktoriális kódolás vajon hogyan valósul meg és időben korrelált



2. ábra. (**A**) A nyilak jelzik a véletlen pályát. Az inputok első négy bitje a falak meglétét írja le abszolút irányok szerint, a második négy bit pedig az útvonal irányát.(**B**) A független komponensek nagyságeloszlása. Bal ábra: tanult irány, jobb felső: új (ellentétes) irány esetében, 8, 16 és 24 input öszefűzésekor. Míg a tanult irányban haladva a kapott eloszlás exponenciális, addig az új irányra (csonkolt) Gauss-eloszlás jellemző.

inputok esetén is képes-e kialakítani a hippokampusz sejtjeire jellemző tér-mezőket. E kérdésekre adott válaszainkat a [Lőrincz és mtsai., 2000a, Lőrincz és mtsai., 2000b, Lőrincz és mtsai., 2000c] közleményekben ismertettük. A numerikus kísérletekben mesterséges labirintus rácson (2.(**A**) ábra) különböző pályákon haladó 'virtuális patkányt' szimuláltunk, a „látott" információ a poziciót és a haladási irányt írta le redundáns módon. Különböző mélységben összefűzött inputsorozatokon tanítottuk a rendszert, majd azt vizsgáltuk, hogy a függetlenítés eredményeképpen kapott belső reprezentációk aktivitás-eloszlása hogyan függött a (i) a haladási iránytól, (ii) a poziciótól és (iii) az összefűzés mélységétől.

Eredményül kaptuk, hogy (i), kialakulnak irányszelektív tér-mezők temporális ICA alkalmazásával a tanítás során bejárt iránnyal egyező irányú inputok esetében, ellenkező irányban haladva viszont nem, és így a rendszer képessé vált eltérő inputhalmazok megkülönböztetésére, Szintén megfigyeltük, (ii) hogy az aktivitás-eloszlás a tanult irányban haladva exponenciális jellegű, míg a másik irányban csonkolt Gauss-eloszlást kaptunk (2.(**B**) ábra). A hippokampusz principális sejtjei mindkét eloszlást mutatják [Treves és mtsai., 1999], amit eredményeink talán képesek magyarázni (iii) Az összefűzés mélysége pedig jobban szétváló (élesebb) tér-mezőket eredményezett. Szintén érdekes

eredmény, hogy (iv) bár az ICA alkalmazásakor a transzformációk előjelére nem tettünk megkötést, a tér-mezők középpontja környékén nem tapasztaltunk előjelváltást, ami egy lehetséges párhuzamot jelent az un. pozitív-kódolás elmélettel [Charles és Fyfe, 1998]. Érdemes megjegyezni, hogy az időbeli inputok összefűzésére a dendritfák integráló hatása valószínüleg lehetőséget teremt ([Henze és mtsai., 1996, Jaffe és Carnevale, 1999, Lisman, 1999]). Tapasztalataink szerint az inputok időbeli összefűzése nélkül a kapott szűrők tulajdonságai kevésbé egyeznek a valósággal. A temporális összefűzés ráadásul hatékonyabb kompressziót is lehetővé tesz, ami feltehetően kiegészíti a hierarchikus kódolást. E kérdéskör biológiai jelentősége abban rejlik, hogy a hippokampusz az összes szenzoros információt feldolgozza, ugyanakkor a különböző rétegeiben lévő sejtek száma messze kevesebb, mint az agy különböző területeiről információt küldő sejtek száma.

3. Eredmények II. Az „Ockham borotvája"-elv alkalmazása az EC-HC hurok modellezésében

A basalis-ganglionok és thalamocorticalis hurkok modellezésében használt elvekhez hasonlóan, a rekonstrukciós hálók modellezésében bevezettük az „Ockham borotvája" modellezési eljárást [Lőrincz és mtsai, 2002a, Lőrincz és mtsai, 2002b] és a kapott szerkezetet leképeztük a hippokampusz és az entorhinális kéreg alkotta hurokra. Az Ockham-elv lényegében azt állítja, hogy ugyanazon probléma lehetséges modelljei közül az a valószínűbb, amelyik kevesebb független feltételre épül. Ezt szem előtt tartva a visszafejtéses mérnöki megközelítéshez hasonlóan kijelöltünk egy alapvető funkciót és azt próbáltuk meghatározni, hogy e központi elem milyen egyéb szerkezeti és funkcionális kényszerekhez vezet. E modellezési eljárás segítségével sikerült a rekonstrukciós hálók jelenlétének általános indoklását adni, valamint meghatározni azt a minimális feltevés halmazt, amelynek segítségével az általunk modellezett agyi területek és funkciók tulajdonságai koherens módon levezethetők. A felépített modell azon jóslatait, melyek már kísérletileg igazoltnak tekinthetők, verifikációs jóslatnak tekintjük, míg a még nem igazolt tulajdonságok valódi predikciói a modellnek. A rendszer fokozatos felépítése során mindvégig pusztán matematikai kényszerek korlátozták az építőelemek kiválasztásában a szabadságunkat. A végeredményül kapott megoldások közül pedig ejtettük azokat, amelyek vagy nem képezhetők le a modellezni kívánt rendszer anatómiai kapcsolataira, vagy pedig mai tudásunk alapján az idegrendszer élettana nem kínál olyan mechanizmust, amellyel az adott funkció

megvalósítható lenne.

Modellezési eljárásunk kiindulási hipotézise az volt, hogy a neurális számítások során a rendszer belső reprezentációkat használ. A reprezentációk problémája nemcsak filozófiai jellegú kérdés, a kognitív tudományok egyik központi vitatémája. Míg általában a különböző kognitív modellek természetesnek veszik a reprezentációk használatát és csak a a különböző értelemben vett hatékonyságukat vizsgálják, mi elsősorban az ún. „homunkulusz"-paradoxonra [Searle, 1992] kerestük a választ. A paradoxon abból indul ki, hogy minden reprezentáció lényegében egy jelsorozat, mely egy másik jelsorozat transzformációjaként áll elő. A létrejött reprezentációt azonban valakinek (a homunkulusznak) „értelmezni" kell. Az értelemzés azonban csak egy újabb reprezentációt eredményez és így végtelen regresszióhoz jutottunk. Feltevésünk szerint az értelmezés határozatlanságából ered az ellentmondás. A megoldáshoz a következő funkcionális hipotézisből indultunk ki: tegyük fel, hogy az értelmezés nem újabb reprezentációk készítését jelenti, hanem azt a képességet, hogy a külvilágból érkező információt a kapcsolatrendszerben tárolt tudás segítségével a rendszer képes újra előállítani, rekonstruálni. Azaz a belső reprezentációk helyett maguk az inputok tekinthetők értelmesnek, ha (a belső reprezentációk segítségével) rekonstruálhatók. E fordított gondolatot már Horn is megfogalmazta: „a látás inverz rajzolás" [Horn, 1977]. Ha az értelmezést a rekonstrukció sikeressége jelenti, akkor a nem értelmezhető jelek rekonstrukciós hibát képeznek. A rekonstrukciós hálók tehát iteratív módon a rekonstrukciós hiba csökkentését próbálják elérni. A hiba eredhet tényleges zajból illetve az adott rendszer számítási kapacitásának korlátaiból. Ennek alapján természetesen adódik, hogy (i) az értelmezéshez szükséges rekonstrukciós hálót ki kell egészíteni zajszűrő mechanizmusokkal, (ii) érthetővé (és szükségessé) válik, hogy az így kiegészített hálókat többrétegű hierarchiába lehessen szervezni.

A javasolt struktúrában lefele és felfele menő kapcsolatok is vannak, így a tárolási kapacitás korlátok mellett az információ-átvitelt befolyásoló csatorna kapacitást is figyelembe kell venni. Több, párhuzamos csatornát feltételezve a hatékony információtovábbítás (maximization of information transfer) úgy érhető el, ha sikerül minimalizálni az átfedést (a kölcsönös információt) a csatornák között (MMI, minimization of mutual information). E feltevést szem előtt tartva újabb kényszer adódik a hálóban található csatornák működésére, azaz a tényleges információ kódolásra: javaslatunk szerint a második és magasabbrendű korrelációk egyidejű eltűntetésére képes független komponens analízis (Independent Component Analysis, [Jutten és Herault, 1991, Comon, 1994]) valósul

meg a felfele menő kapcsolatrendszerben. Röviden összefoglalva e transzformáció képes nem Gauss-eloszlással jellemző források lineáris keverékeinek szétválasztására és így Gauss-eloszlással jellemezhető zaj szűrésére. A bemenő jelek függetlenítésének mélyreható következményei vannak:

- az előző tézispontban leírt temporális kódolás segítségével időben és térben elkülöníthető események definiálhatók
- a transzformáció eredményéül kapott reprezentációk statisztikája alapján könnyen integrálható a rendszer zajszűrő mechanizmusokkal (küszöbölés, ritka reprezentációk kialakítása, lásd pl. [Olshausen és Field, 1997]).
- Szintén pusztán a statisztikák alapján természetes módon beilleszthető egy újdonság-detektáló mechanizmus.

A rendszer rekonstrukciós képessége a kapcsolatokban tárolt tudás és a feldolgozás alatt álló input statisztikai hasonlóságán alapul. Ha adaptív, önszervező rendszert modellezünk, „a priori" tudás feltevése nélkül, akkor minden új, még nem „látott" input befolyásolja a már elraktározott tudást. Ha azonban a raktározási kapacitását már elérte a rendszer, akkor különbséget kell tenni a már ismert információk torzított (zajos) verziója és a még nem látott információk között. Az „újdonság-érzékelés" (novelty detection) problémakör szorosan kapcsolódik a pszichológiában használt kategóriális percepció témához, illetve a mesterséges intelligencia kutatásokban használt osztályozás problémájához. Az így levezetett információfeldolgozási kényszerek és tulajdonságok alapján sikerült a funkcionális leképezés több elemét is pontosítani, valamint adtunk egy lehetséges leírást a hippokampuszra jellemző kétfázisú viselkedésről. Ennek alapján feltehető, hogy:

- az EC-HC hurok rekonstrukciós hálót képez.
- a CA3 réteg a függetlenítés első lépcsője, valamint az úgynevezett éles hullám fázisban a HTM hangolásához szükséges „visszajátszás" (pl. [Levy, 1996]) elinditója.
- A CA1 rétegben történik a függetlenített jelek ritkítása az EC-ből az úgynevezett közvetlen úton (EC III–CA1 direct path) érkező jelek segítségével.
- A ritkitáshoz szükséges kapcsolati rendszer hangolási sebessége a felfelé és lefele tartó kapcsolatok hangolási sebessége *közé* kell essen.

- A rekonstrukciós dinamika asszimetriájából következik, hogy a felfelé tartó kapcsolatok hangolása gyors legyen és megelőzze a lefelé tartó kapcsolatokét. Ez a gyors tanulás és a ritkitás együtt lehetővé teszi új információk fokozatos megtanulását (bázis elemmé válását a lefele tartó kapcsolatok mátrixában) katasztrófális interferencia [Grossberg, 1982, McCloskey és Cohen, 1989] kialakulása nélkül, azaz nem sérülnek vagy tünnek el már megtanult elemek.

- A függetlenített és ritkított jelek mintázatkiegészítése feltehetőleg a CA1–EC V-VI kapcsolatrendszerben történik.

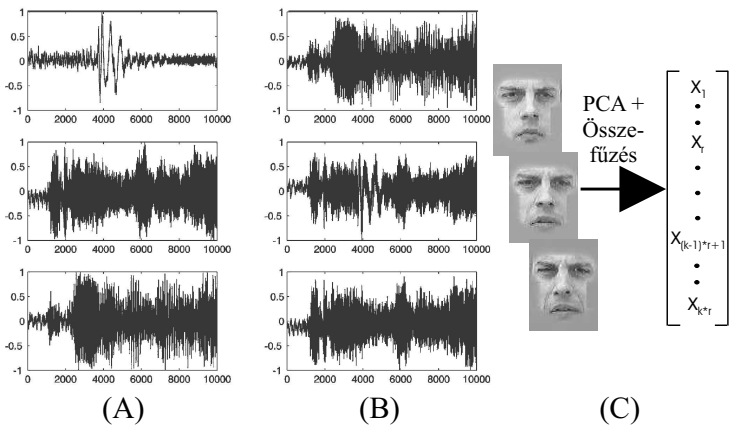
- A belső reprezentációkat kialakító modell az EC mély rétegeiben található és a magasabbrendű területekről érkező hatások is elsősorban ezen a ponton módositják a rekonstrukciós háló működését.

- A modell réteg asszociatív struktúrája feltehetőleg valamilyen időbeli predikcióra képes, melynek célja lehet az iterációból származó késleltetés kompenzációja.

Egyszerű numerikus szimulációkkal azt is megmutattuk, hogy a felfele és lefele mutató transzformációkra vonatkozó nem túl szigorú megkötések mellett sem az aktivitások, sem az aktivitásváltozások normálása nem zavarja meg jelentősen a rendszer konvergens viselkedését.

4. Eredmények III. Alacsony- és magasabbrendű memóriafolyamatok kapcsolata a EC-HC hurokban

Az előző tézispontban leírtak alapján arra jutotam, hogy az L-B modell magját jelentő rekonstrukciós hálónak a már megtapasztalt inputok terében kell hatékonynak lenni, miközben az újonnan érkező inputok lehetnek (i), zajosak, (ii), hiányosak. Tehát garantálni kell, hogy az új információ ne tehesse tönkre a már eltárolt információkat (ne jöhessen létre 'katasztrofális interferencia'). Az L-B modellt kiegészítettem [Szirtes és Lőrincz, 2002d] egy zajtalanítást és ritkítást végző második hurokkal, mely az aktuálisan feldolgozott input statisztikai tulajdonságai alapján működik. Az így kiegészített hurok működését természetes és mesterséges hangok keverékén, illetve arckifejezéseket bemutató filmeken vizsgáltam (3. ábra), különös tekintettel a belső reprezentációk kialakulásának sebességére és az aktivitás-eloszlására.



3. ábra. A szimulációk során használt inputok. Természetes és mesterséges hangok keverékének használata lehetővé tette a kölcsönös információ (átfedés) mértékének szabályozását. (**A**) Három független hangminta. (**B**) A három minta lineáris keveréke zaj hozzáadása nélkül. (**C**) Az érzelmeket mutató filmek pedig a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikájával közös (Simon-csoport) kutatásainkat segítették. A filmek előfeldolgozása többlépcsős volt. Az arcokat középre helyeztük, normáltuk, a képkockák méretét Főkomponens Analízis (PCA) módszerrel csökkentettük, majd az így kapott vektorokat összefűztük.

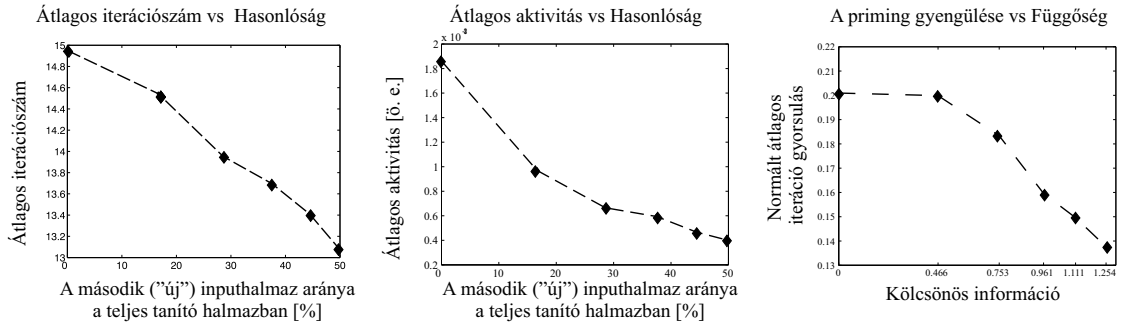
Mindezek alapján a következő eredményekre jutottam:

- Azt jósolom, hogy a hosszútávú memóriát (HTM) a rekonstrukciós háló top-down - azaz a rejtett rétegből induló - kapcsolatai adják.

- Megmutattam, hogy a hálóban egyértelmű a különböző transzformációkat végző kapcsolati mátrixok hangolási sorrendje.

- Megmutattam, hogy a felfele és lefele menő kapcsolatok hangolásának eltérő sebessége *és* a küszöbként működő másodlagos hurok működésének együttes eredménye, hogy a rendszer képessé válik új információk gyors érzékelésére („one-shot learning"), zajtól való elválasztására és további feldolgozására a hosszútávú memória átalakítása (felülírása) nélkül is.

- Szimulációk segítségével megmutattam (4.(**A**)), hogy az inputoknak a már megtanult információkhoz való hasonlósága függvényében gyorsul a rekonstrukciós iteráció. Ez azt jelenti, hogy már az HTM módosítása előtt is tanul a rendszer. E jelenség biológiai megfelelője az ún. implicit memória [Graf és Schachter, 1985, Schachter, 1987, Squire, 1992], mely magában foglal több, alacsonyabbrendű



(A) (B) (C)

4. ábra. Szimulációk eredményeinek összegzése. (**A**) A rekonstrukciós dinamika relaxációjának függése a tanító és a teszt halmaz hasonlóságától. Két lineáris keveréket két-két további részre osztottunk. A tanító halmaz a két eltérő halmaz egy-egy felének különböző arányú keverékéből jött létre. A teszt halmaz pedig az egyik halmaz még fel nem használt része volt. Látható, hogy minél inkább megfelel az adott input a tanító halmaznak, annál kevesebb iteráció szükséges az önkényesen beállított konvergencia-kritérium eléréséhez. (**B**) A relaxált állapothoz tartozó belső reprezentáció normájának függése a tanító és a teszt halmaz hasonlóságától. Tapasztalataink szerint nagyobb hasonlóság nemcsak gyorsabb relaxációt eredményezett, hanem kisebb aktivitást is a belső reprezentáció szintjén, azaz ritkább kódolást nyertünk. (**C**) A források függetlenségének hatása az iteráció változására. Azt is megvizsgáltuk, hogy a források függetlenségére vonatkozó feltétel sérülése milyen hatással van a tapasztalt dinamikai változásokra. A három forráshoz hozzákevertünk egy negyedik forrást, így egy háromdimenziós szeparáló mátrix semmiképpen sem tudja tökéletesen „szétkeverni" az eredeti forrásokat. A kísérletek azt mutatják, hogy a függetlenségi feltétel bizonyos határig elengedhető, azon túl még nagyon hasonló tanító és teszthalmazok esetében sem tapasztalunk gyorsulást az iteráció folyamán (priming eltűnése).

memóriafolyamatot. Ennek alapján a modell lehetővé teszi, hogy egységes keretben tárgyaljuk az explicit (HTM formálás) és az implicit (ismétléses (nem tudatos) rögzülés, repetition priming) memóriafolyamatokat.

- Szimulációk segítségével mutattam meg azt is, hogy, hogy miközben gyorsul az iteráció, az ún. belső modell réteg számítási egységeinek zömében csökkent az átlagos aktivitás (4.(**B**)). Ugyanakkor néhány elem esetében határozott növekedést tapasztaltam. Ha a biológiai leképezésnek megfelelően a számítási egységeket mint neuronokat tekintjük, a megfigyelt populációaktivitás időbeli változásai az ún. „repetition suppression" és „repetition enhancement" folyamatok együttes megjelenésének feleltethetők meg [Lőrincz és mtsai, 2002c]. Azt a következtetést is levontam, hogy e folyamatok tekinthetők a priming neurobiológiai korrelátumának és nem pusztán együttes előfordulásról beszélhetünk, melynek oka továbbra is ismeretlen.

5. Eredmények IV. Alzheimer-kórra jellemző kategória-tanulási képesség változásának modellezése

Különböző vizsgálatok bizonyították, hogy a szenzoros információk csoportosításához, kategóriák kialakításához több agyterület összehangolt működése szükséges. Általánosan elfogadott nézet [Knowlton és Squire, 1993], hogy a kategória-tanulást a neokortex implicit mechanizmusai biztosítják, míg a felismerési feladatokban a mediotemporális (az explicit memóriáért felelős területek) vesznek részt. A megtanult kategóriák hatékony tárolása azonban még nem ismert, mind az úgynevezett példaalapú modell, mind a prototípus alapú modell mellett és ellen is szólnak érvek. Mindenesetre feltételezhező, hogy idővel a prototípusok önálló tárolása is kialakul. A [Kéri és mtsai, 2002] publikációban egy olyan meglepő kisérleti eredményt próbáltunk megmagyarázni, amely mindezen ismereteknek ellentmondani látszik: Kéri Szabolcs és munkatársai a Szegedi Egyetem Pszichiátriai és Fiziológiai Tanszékén az explicit memória betegségének tekintett Alzheimer-kórban szenvedők kategória tanulási képességeit vizsgálták és eredményeik szerint a betegek zöme még erősen torzított elemeket is helyesen tudott kategorizálni, ugyanakkor a „legtipikusabb" prototípust *nem ismerte fel*.

Klasszikus értelmezés szerint az Alzheimer-kór az explicit memóriarendszer betegsége, a korai szakaszra jellemző mediotemporális (halántéklebeny középső része) amnézia fo-

5. ábra. Prototípus felismerésének gyengülése. (**a**) A kialakult memória mátrix prototípus inputokon tanulva. (**b**) A kialakult memória mátrix gyengén torzított ($d = 1$) inputok esetében. (**c**) A kialakult memória mátrix erősen torzított ($d = 2$) mátrixok esetében. A jobb láthatóség érdekében az értékeket simitottuk. Fehér (fekete): 1 (0) kapcsolati erősség. (d) Az osztályhoz tartozás mértéke a torzítottság függvényében. Két szimuláció erredménye látható, az egyik esetben (fehér kör,($d = 1$)) gyengén torzított, a másik esetben (fekete kör,($d = 2$)) erősen torzított inputokat használtunk a tanításhoz.

kozatosan kiteljesedik és a kognitív funkciók beszűkülésével, viselkedési zavarral és az elbutulást követôen kialakuló gyors biológiai leépüléssel jellemezhetô. A neuropatológiai vizsgálatok jellemzôen a halántéklebeny és a hippokampusz területén a cholinerg receptorok pusztulását mutatták ki, valamint az entorhinális kéreg térfogata is jelentősen csökkent.

Mivel az L-B modell a memóriaszerveződés alapvető leírását célozza, így azt vizsgáltuk, vajon e speciális jelenséget képesek vagyunk a modell segítségével megmagyarázni. A rekonstrukciós hálóban a felfele és lefele ható transzformációs kapcsolatokat most nem hangoltuk, egyedül a belső reprezentációt kialakító modell réteg asszociatív struktúrái tanulhattak Hebb-kölcsönhatások révén. Az előre rögzített memória vektorok szolgáltak lokális szűrőként egy 10x10-es rács világban. A szűrök 3x3-as területen voltak csak „érzékenyek" (nem nulla elemek csak adott pozíciókban voltak), 100 szűrőt alkalmaztunk (tehát a belső reprezentáció dimenziója is 100 volt) úgy, hogy mindegyik középpontja megfelelt a rács-világ egy pontjának. A tanítás során egy tetszőleges, de rögzített Gauss inputot tekintettünk prototípusnak és ennek eltolt (torzított) változatait használtuk. A tesztekben pedig azt viszgáltuk, hogy (i) a prototípusra, (ii) az azonos mértékben torzított inputokra, (iii) eltérő mértékben torzított inputokra milyen belső reprezentáció alakul ki a rekonstrukciós hálóban. A kategóriához tartozást egy olyan mérőszámmal (d) jellemeztük, mely arányos volt a modellréteg aktivitásával. Az (5). ábra két szimuláció sorozat eredményét mutatja. A kisebb mértékű ($d = 1$) torzított példák esetében a vártnak megfelelően d a torzítás monoton csökkenő függvénye,legerősebben a prototípisnál mutatja a kategóriához tartozást. Nagyobb torzításnál ($d = 2$) azonban a görbének maximuma van a tanult torzításnál és *lokális minimuma* a prototípusnál. A modell réteg asszociatív hálójának gyengítése is ugyanehhez a meglepő függéshez vezetett. Ennek alapján arra jutottunk, hogy a belső reprezentációs réteg asszociatív kapcsolatainak sérülései eredményezhetik a prototípus-felismerés gyengülését, miközben a torzított kategória példák felismerése továbbra is sikeres. A kapott eredmények egyrészt közvetve bizonyítják modellünk helyességét, másrészt egy erősen vitatott jelenségkörre adnak egyszerű és elegáns magyarázatot. Ez a tézis két kutatócsoport együttműködésének eredménye.

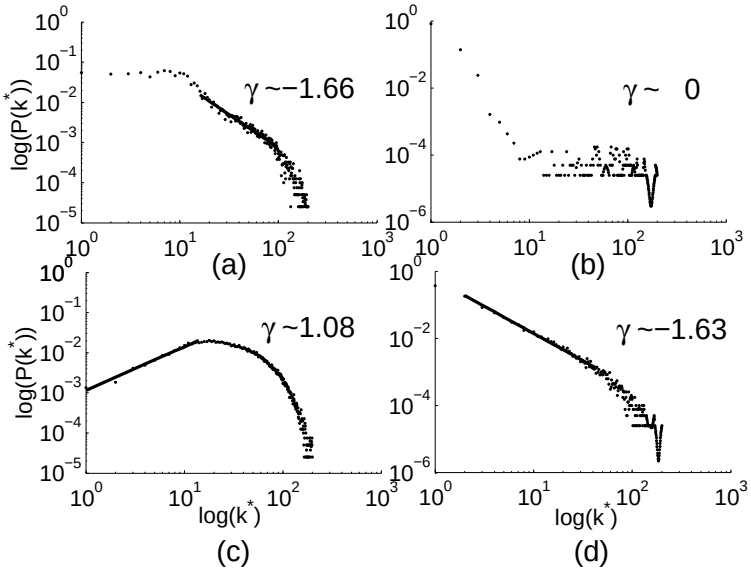
6. Eredmények V. Zaj indukált önszervező hálózatban kis világ struktúrák megjelenése

A modellben kulcsfontosságú szerepet kap a zaj (szűk értelemben struktúra nélküli jel, tágabb értelemben a rendszer számára értelmezhetetlen vagy irreleváns jel) kiszűrése. Azonban az idegrendszer kialakulása (embrionális és újszülött korban megfigyelhető ingerfüggetlen terjedő aktivitáshullámok) és működése (sejtszintű aktivitás áttevődés nagy variabilitása) során fellépő endogén eredetű zaj látszólag ellentmond az eddig felépített elveknek, létezésének oka nem ismert és erre a kiegészített L-B modell sem ad magyarázatot. Ráadásul a 'mindent vagy semmit' jellegű, zajjal szemben elméletileg robusztus,impulzusszerűen terjedő aktivitást torzítja, valamint a pontos időzítésre épülő, ma már klasszikusnak tekintett Hebb-tanulás hatékonyságát is látszólag lecsökkenti az egyes neuronok viselkedésbeli ingadozása. Mivel az eddig leírt információfeldolgozó modell nem használta ki a tényleges kapcsolati struktúrákban lévő lehetőségeket (topografikus kapcsolatok, funkcionális szerveződések) felmerült, hogy a zaj szerepe elsősorban a szerkezeti változásokban lehet jelentős. E kérdés megválaszolásához egy ún. minimális modellel közelítettem az idegrendszeri kapcsolatokat, melyben nem különböztettem meg gerjesztő és gátló neuronokat és a a neuronok közötti kölcsönhatást (azaz a kapcsolati erősségek változást, tehát a tanulást) általánosított Hebb-tanulással (STDP, spike timing dependent plasticity, részletes leírás és referenciák a [Rieke és mtsai., 1996]-ban találhatók) jellemeztem [Szirtes és mtsai, 2003a]. A következő egyenletek definiálták a rendszert:

$$\frac{\Delta a_i}{\Delta t} = \sum_j w_{ij} a_j^s + x_i^{(ext)}, \tag{1}$$

$$\frac{\Delta w_{ij}}{\Delta t} = \sum_{(t_i,t_j)} K(t_j - t_i) a_i^{t_i,s} a_j^{t_j,s}, \tag{2}$$

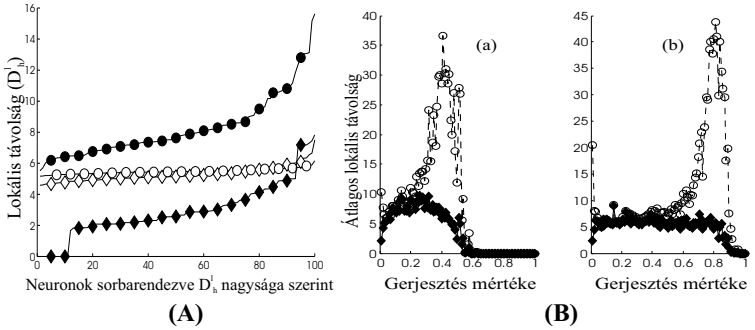
Az alkalmazott szabály alapján két neuron között a kapcsolat erősödik, ha az inputot küldő és fogadó neuron bizonyos időablakon belül aktív, és gyengül, ha a sorrend fordított, illetve az időablakon kívül aktívak. A gyengítés és erősítés arányát, időfüggését és aszimmetriáját is vizsgálták különböző fajokban és szövetekben és több különböző szabályt is meg tudtak fogalmazni, [Abbott és Nelson, 2000] . Mivel elsősorban arra voltam, kiváncsi, hogy „értelmes", szerkezettel rendelkező inputok hiányában miként alakul egy ilyen rendszer, kizárólag véletlen zajjal (szinaptikus bombázás) indukáltam változásokat a kapcso-



6. ábra. **Súlyozott kimeneti kapcsolatok eloszlása log-log skálán különböző paraméterek mellett.**

A négy diagrammhoz tartozóparaméterek a következők (a): $r_{A^+/A^-} = 0.1$ $r_{ex} = 0.3$, (b): $r_{A^+/A^-} = 0.1$ $r_{ex} = 0.6$,(c): $r_{A^+/A^-} = 0.6$ $r_{ex} = 0.3$ és (d): $r_{A^+/A^-} = 0.6$ $r_{ex} = 0.75$, ahol r_{A^+/A^-} a gerjesztés és gátlás aránya a K függvényben, r_{ex} pedig a gerjesztett nódusok arány. Az (a) és (d) eset két példa a „kis-világokat" jellemző kitevő szabályt követő eloszlásokra.

lati súlyokban. Mivel a gerjesztő jelekben sem térbeli (hiszen topológiai kényszereket nem vizsgáltunk), sem időbeli korreláció nem volt, ezért az általunk választott kölcsönhatást definiáló függvényben (K) is kizárólag a gerjesztés és gátlás aránya volt fontos. Összefoglalva tehát a nódusok a külvilágból és egymástól kapnak jeleket és egy alkalmas küszöbfüggvényen keresztül maguk is bocsátanak ki jeleket (aktivitásterjedés). Bár a valódi idegsejtek aktivitás-terjedése sokkal összetettebb, matematikai analízissel megmutatható [Burkitt és mtsai., 2003], hogy ez az egyszerű modell is stabil fix-ponthoz vezető dinamikával rendelkezik. Az ily módon felépített rendszer kapcsolaterősségeinek változását vizsgáltuk az idő függvényében gráfelméleti eszközökkel. A kapcsolati háló irányított, súlyozott gráf, melyben kezdetben véletlen eloszlású, gyenge kapcsolatok vannak. Egyfelől vizsgáltuk a súlyozott ki- és bemeneti kapcsolatok eloszlását az idő változásával, másrészt a lokális és globális összekötöttséget tanulmányoztuk a világháló (world wide web) vizsgálatához bevezetett mérőszámok segítségével. Az úgynevezett hálózati hatékonyságot mértük az irodalomban elterjedt karektirisztikus útvonalhossz és fürtképző együttható



7. ábra. **(A)** Lokális összekötöttség. A könnyebb értelmezhetőség kedvéért nem minden adatot ábrázoltam, az összetartozó pontokat pedig összekötöttem egy vonallal. A gyémánttal jelzett vonalak az STDP hangolással kialakult szerkezetekhez taroznak, míg a körök a kialakult súlyok véletlen újraelosztásával létrejött véletlen hálókhoz tartoznak. Az üres jelek az 1. régióhoz, a fekete jelek a 2. régióhoz tartoznak. Látható, hogy a 2. régióhoz tartozó esetben a kialakuló hálózatot sokkal kisebb lokális távolság (nagyobb összekötöttség) jellemzi, mint az összehasonlításhoz használt véletlen hálüt. A globális összekötöttség a kialakult háló és a randomizált hálók esetében: 1. régió: $D_h = 5.25$ and $D_h^r = 5.01$, 2. régió: $D_h = 6.28$ and $D_h^r = 4.98$ **(B)** Zaj hatása a szerkezetre. Két különböző kernel esetében vizsgáltuk, hogy a gerjesztés növelése milyen hatással van a szerkezetre. (a) $r_{A^+/A^-} = 0.33$, (b) $r_{A^+/A^-} = 0.6$ Gyémánt: átlagos lokális távolság az STDP hálóban. Kör: átlagos lokális távolság a megfelelő véletlen hálóban. Látható, hogy az STDP hálóban még igen nagy gerjesztés esetén is megmarad a nagy lokális összekötöttség, nem „zilálódnak" szét a hálók.

helyett, mivel e mérőszám mind a lokális mind a globális összekötöttség mérésére azonos formát használ és így értelmezése is egységesebb.

Különböző paraméterekhez tartozó néhány jellegzetes eloszlást mutat a 6. ábra. Kitevő szabályt (power law) követő elsozlást feltételezve ($P(k^*) \approx k^{*\gamma} e^{-k^*/\xi}$, ahol k^* a kapcsolati erősség diszkretizált értékeit jelöli), lineáris illesztést számoltunk, hogy megbecsülhessük γ-t. Eredményeink szerint viszonylag széles paraméter-tartományban kisvilág eloszlás alakul ki.

A hálózati hatékonyságot (7.(**A**) ábra) és a zajjal szembeni ellenállóképességet (7.(**B**) is tanulmányoztuk.

A hálózati hatékonyság vizsgálata megmutatta, hogy (i) a létrejött skálamentes struktúrák egyben kis világok is, a súlyok átrendezésével kapott véletlen hálók lokális összekötöttsége sokkal kisebb. A szimulációk azt is megmutatták (ii), hogy e nagymértékű fürtösödést még nagy zaj esetében is megtartják a kialakult hálók. Mindezek alapján azt mondhatjuk, hogy az eredményül kapott hálózatok skála-mentesek kis-világok, melyek hatékonyak információfeldolgozás és -továbbítás szempontjából és a működési hibákkal szemben kevésbé érzékenyek. Ezen eredmények egyfelől választ adhatnak fejlődéstani

kérdésekre (embrionális korban endogén zaj hullámok szerepe), másfelől az asszociatív memória hatékonyságának egyik magyarázatát adhatják. A modell bizonyos közelítések mellett alkalmas a Web matematikai leírására is egyben. A skálamentes rendszerek viselkedését leíró népszerű modellekkel (mint pl. [Barabási és Albert, 1999]) ellentétben azonban nem tételezi fel a rendszer állandó (ezért nem reális) növekedését és nem igényel a kapcsolatok változtatásához „globális" tudást (azaz egy nódusnak elegendő a vele összeköttetésben lévő egyéb nódusok állapotát ismerni).

7. Eredmények VI. A Lőrincz-Buzsáki modell kiegészítése predikcióra képes struktúrával

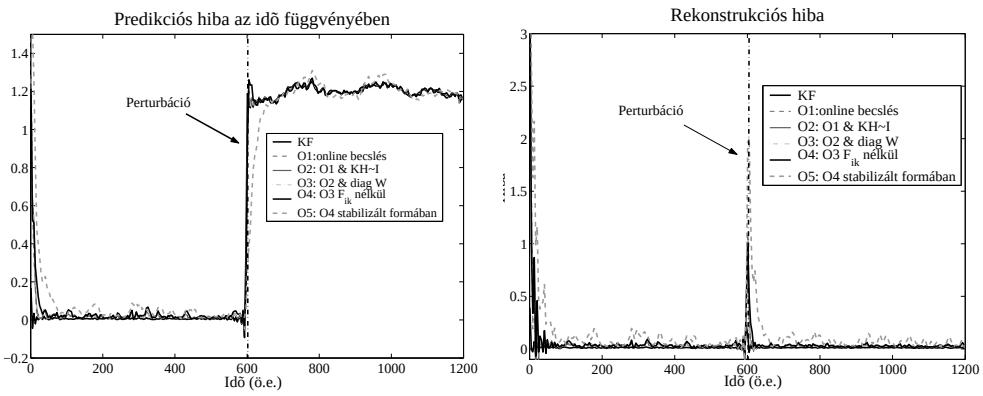
Az L-B modell további kiterjesztésével lehetővé tettük, hogy a rekonstrukció során az inputok térben és időben is kiegészíthetők legyenek. Ezáltal a modell képessé válik a külvilág változásainak korlátozott predikciójára, ami egy gyorsan változó környezetben a környezetével kölcsönható ügynök számára rendkívül előnyös tulajdonság. A szenzoros információfeldolgozás prediktív jellegét a neurobiológiában is többen hangsúlyozzák. Ismét az Ockham borotvája elvet követve célszerű a lehető legegyszerűbb dinamikát „feltenni" a világ változásairól (illetve feltesszük, hogy az agyunkban működő predikciós rendszer egyszerű dinamikát keres a világ változásai mögött). Tekintsük a következő *lineáris* dinamikai rendszert:

$$\mathbf{y}^t = \mathbf{H}\mathbf{x}^t + \mathbf{n}^t \quad \text{megfigyelési folyamat} \tag{3}$$

$$\mathbf{x}^{t+1} = \mathbf{F}\mathbf{x}^t + \mathbf{m}^t \quad \text{rejtett folyamat} \tag{4}$$

ahol $\mathbf{m}^t \propto \mathcal{N}(0, \Pi)$, $\mathbf{n}^t \propto \mathcal{N}(0, \Sigma)$ független Gauss-zajok. A feladat a rejtett változó $\mathbf{x}^t \in \mathbf{R}^n$ becslése a megfigyelést változókból $\mathbf{y}^\tau \in \mathbf{R}^p$, $\tau \leq t$. Négyzetes hibában a legjobb becslést az ún. Kálmán-szűrő adja [Kalman, 1960], melynek „predikciós" egyenlete a következő:

$$\hat{\mathbf{x}}^{(t+1|t)} = \mathbf{F}\hat{\mathbf{x}}^{(t|t-1)} + \mathbf{K}^t\left(\mathbf{y}^t - \mathbf{H}\hat{\mathbf{x}}^{(t|t-1)}\right) = \mathbf{F}\hat{\mathbf{x}}^{(t|t)} \tag{5}$$



8. ábra. A Kálmán-szűrő és közelítéseinek összevetése. 'Predikciós hiba': $\|\mathbf{x}-\hat{\mathbf{x}}\|$. 'Rekonstrukciós hiba': $\|\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\|$. 'KF' and 'O1-O5' az optimális Kálmán szűrőt és az 5 levezetett online modellt jelölik. A lineáris dinamikai rendszer kétdimenziós rotációs mátrixokból áll, $t = 600$-nál a megfigyelési mátrix $\mathbf{H}$ hirtelen megváltozott. (A váltás körül a negatív értékek a jobb megjelenítéshez használt Csebisev-szűrő miatt jelentek meg.)

amelyben $\mathbf{K}^t$ mátrix (az ún. Kalman-gain) az $\hat{\mathbf{x}}^t$ *a priori* és a *posteriori* kovariancia mátrixainak segítségével számolható. Az $e^t = \mathbf{y}^t - \mathbf{H}\hat{\mathbf{x}}^{(t|t-1)}$ kifejezés a rekonstrukciós hiba, ugyanis zajmentes esetben $\mathbf{H}\hat{\mathbf{x}}^{(t|t-1)}$ tökéletesen visszaállítaná az inputot. $\mathbf{K}$ szerepe a predikciós tag és a rekonstrukciós hiba hatásának egyensúlyozása (nagy zaj esetében nem engedi érvényesíteni a rekonstrukciós hibát). E számolás azonban mátrixinverziót igényel, mely „globális tudást" feltételez, ezért pusztán lokális kölcsönhatásokkal (mint a neuronok esetében a Hebb-tanulás) nem végezhető el. Az általános megoldás ráadásul a zaj folyamatok kovariancia mátrixait ismertnek feltételezi és a modellparaméterek is rögzítettek. A rekurzív predikciós hiba módszer [Ljung és Soderstrom, 1993] azonban megoldást kínál ezekre a problémákra [Póczos és Lőrincz, 2003]. A predikciós egyenlet egy tetszőleges paraméterezett közelítése a következő:

$$\hat{x}_i^{t+1} = \sum_j F_{ij}\hat{x}_j^t + \theta_i^t \sum_l K_{il} e_l^t \tag{6}$$

$\mathbf{K}$ faktorizációja pedig a következő alakhoz vezet:

$$\hat{x}_m^{t+1} = \sum_j F_{mj}\hat{x}_j^t + \sum_i N_{mi}\,\theta_i^t \sum_l K_{il} e_l^t \tag{7}$$

Célunk a θ paraméter hangolása, hogy a $J_k(\theta_k) = \frac{1}{2}E[(\epsilon_k^t)^2]$ költségfüggvényt minimalizáljuk, ahol $\epsilon_k^t = \sum_l K_{kl} e_l^t$ a transzformált hiba. Az eredményül kapott hangolási egyenletek:

$$\theta_k^{t+1} = \theta_k^t + \alpha \sum_{lj} K_{kl} H_{lj} W_{jk} \epsilon_k^t \tag{8}$$

$$W_{ik}^{t+1} = \sum_j F_{ij} W_{jk}^t \xi_k - \theta_i^t \sum_{lj} K_{il} H_{lj} W_{jk}^t \xi_k + \delta_{ik} \epsilon_k^t \tag{9}$$

Ahol $W_{ik}^t = \frac{\partial \hat{x}_i^t}{\partial \theta_k}$ egy segédmátrix, δ a Kronecker-delta, ξ pedig egy segédvektor, melynek segítségével hagyományos neurális formában felírható egyenlethez jutottunk, α (időfüggő) tanulási ráta. ξ tekinthető egyfajta véletlen „zajnak", melyet a rendszer maga hoz létre. Különböző matematikai és biológiai megszorítások figyelembevételével 5 egyszerűsített modellt vezettem le, melyek közül a legegyszerűbb a következő forma:

$$W_{ii}^{t+1} \approx W_{ii}^t + \gamma\{-\theta_i^t W_{ii}^t \xi_i + \epsilon_i^t\} \tag{10}$$

$$\theta_k^{t+1} \approx \theta_k^t + \alpha W_{kk} \epsilon_k^t \tag{11}$$

Érdekes módon, a **W** hangolásának egyenlete (10) megegyezik a gátló szinapszisok kísérleti úton megfigyelt változását leíró modellel [Komatsu és Iwakiri, 1993, Komatsu, 1996], ennek alapján a leképezésben feltettük, hogy e kapcsolatok zömében gátló jellegűek, így a posztszinaptikus sejtek természetétől függően inhibiciót vagy dizinhibiciót fejtenek ki.

A 8.ábra összehasonlítja az eredeti Kálmán-szűrő és a PRE módszer segítségével levezetett közelítéseinek teljesítményét. Látható, hogy lényegében mindegyik modell konvergál az optimális megoldáshoz, hirtelen paraméterváltozás esetében pedig az online modellek még felül is múlják a (Gauss-hibát feltételező) Kálmán-szűrőt. Azt is megmutattuk, hogy –a matematikai eredmények megőrzése mellett–a paraméterezett predikciós egyenlet (6.egyenlet) minimális változtatásával (7.egyenlet) olyan grafikus reprezentációhoz juthatunk, mely természetesen illeszkedik a Lőrincz-Buzsáki modellhez. Az 9. ábra mutatja az EC-HC hurok anatómiai kapcsolatait, az eredményül kapott neurális (lokális), online Kálmán-szűrő minimális szerkezetét és leképezését az EC-HC hurokra.

9. ábra. **(A)** A hippokampusz és környéke releváns kapcsolatai (az EC III→CA1 direkt út' nélkül). Az EC II és III az entorhinális kéreg felszíni rétegeit jelöli, az EC V-VI pedig a mély rétegeket. A fekete nyilak elsősorban gejesztő, a fehér nyilak főleg gátló kapcsolatokat jelölnek. Lokális „mikro-körök' feltehetően minden rétegben találhatóak. **(B)** A neurális Kálmán-szűrő minimális modellje. **(C)** A modell leképezése a EC-HC hurokra. A transzformációk a kapcsolati rendszerekben valósulnak meg, egy adott réteg neurális aktivitásai pedig a vektoroknak feleltethetőek meg. CA3-ban történik a fehérítés (ICA első lépése), maga a jel szeparálás pedig a CA1-ben. A θ paraméter moduláló hatását feltehetőleg a CA1 funkcionális „mikro-körei" fejtik ki, **W** pedig elsősorban gátló kapcsolatok révén biztosítja a kölcsönhatást. A prediktív belső modellt pedig az EC mély rétegeinek „fenntartott aktivitásai"[Egorov és mtsai., 2002] működtetik.

A tézisben szereplő saját hivatkozások

[Lőrincz és mtsai, 2000a] Lőrincz, A., **Szirtes, G.**, and Takács, B. (2000). Is the hippocampus engaged in forming and encoding independent components of temporal sequences? In Bower, J., editor, *Conf. on Computational Neuroscience*. Elsevier.

[Lőrincz és mtsai, 2001a] Lőrincz, A., Szatmáry, B., **Szirtes, G.**, and Takács, B. (2001a). Recognition of novelty made easy: Constraints of channel capacity on generative networks. In French, R., editor, *Connectionist Models of Learning, Development and Evolution. Proceedings of the 6th Neural Computation Workshop (NCPW6)*, pages 73–82, London. Springer Verlag.

[Lőrincz és mtsai, 2001b] Lőrincz, A., **Szirtes, G.**, Takács, B. and Buzsáki, G. (2001b). Independent component analysis of temporal sequences forms place cells. *Neurocomputing*, (38-40):769–774.

[Lőrincz és mtsai, 2002a] Lőrincz, A., Póczos, B., **Szirtes, G.**, and Takács, B. (2002a). Ockham's razor at work: Modeling of the 'homunculus'. *Brain and Mind*, 3:187–220.

[Lőrincz és mtsai, 2002b] Lőrincz, A., Szatmáry, B., and **Szirtes, G**. (2002b). Mystery of structure and function of sensory processing areas of the neocortex: A resolution. *J. Comp. Neurosci.*, 13:187–205.

[Lőrincz és mtsai, 2002c] Lőrincz, A., **Szirtes, G.**, Takács, B., Biederman, I., and Vogels, R. (2002c). Relating priming and repetition suppression. *Int. J. of Neural Systems*, 12:187–202.

[Szirtes és Lőrincz, 2002d] **Szirtes, G.** and Lőrincz, A. (2002). Low level priming as a consequence of perception. In *Connectionist Models of Cognition and Perception. Proceedings of the 7th Neural Computation Workshop (NCPW7)*, pages 223–235. World Scientific.

[Kéri és mtsai, 2002e] Kéri, S., Benedek, G., Janka, Z., Aszalós, P., Szatmáry, B., **Szirtes, G.**, and Lőrincz, A. (2002). Categories, prototypes and memory systems in alzheimer's disease. *Trends in Cognitive Science*, 6:132–136.

[Szirtes és mtsai, 2003a] **Szirtes, G.**, Palotai, Z., and Lőrincz, A. (2003). Emergence of scale-free properties in hebbian networks. submitted to Neural Network.

[Szirtes és mtsai, 2003b] **Szirtes, G.**, Póczos, B. and Lőrincz, A. (2003). Neural Kalman-gain. under preparation.

Irodalomjegyzék

[Abbott és Nelson, 2000] Abbott, L. és Nelson, S. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183.

[Barabási és Albert, 1999] Barabási, A. L. és Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

[Barlow, 1972] Barlow, H. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:295–311.

[Bell és Sejnowski, 1997] Bell, A. J. és Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Res.*, 37:3327–3338.

[Burkitt és mtsai., 2003] Burkitt, A. N., Meffin, H., és Grayden, D. B. (2003). Spike timing-dependent plasticity: The relationship to rate-based learning for models with weight dynamics determined by a stable fixed-point. To appear in Neural Computation.

[Charles és Fyfe, 1998] Charles, D. és Fyfe, C. (1998). Modelling multiple cause structure using rectification constraints. *Network: Computations in Neural Systems*, 9:167–182.

[Comon, 1994] Comon, P. (1994). Independent component analysis - A new concept? *Signal Processing*, 36:287–314.

[Egorov és mtsai., 2002] Egorov, A., Hamam, B., Fransén, E., Hasselmo, M., és Alonso, A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, 420:173–178.

[Field, 1987] Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, A4:2379–2394.

[Field, 1994] Field, D. (1994). What is the goal of sensory coding? *Neural Computation*, 6:559–601.

[Graf és Schachter, 1985] Graf, P. és Schachter, D. (1985). Implicit and explicit memory for new associtions in normal subjects and amnesic patients. *J. Exp. Psychology: Learning, Memory and Cognition*, 11:501–518.

[Grastyán és mtsai., 1959] Grastyán, E., Lissák, K., Madarász, I., és Donhoffer, H. (1959). The hippocampal electrical activity during the development of conditioned reflexes. *Electroencephal. Clin. Neurophysiol.*, 11:409–430.

[Grossberg, 1982] Grossberg, S. (1982). *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control.* Boston: Reidel.

[Henze és mtsai., 1996] Henze, D., WE, W. C., és Barrionuevo, G. (1996). Dendritic morphology and its effects on the amplitude and rise-time of synaptic signals in hippocampal ca3 pyramidal cells. *J. Comp. Neurology*, 369:331–344.

[Horn, 1977] Horn, B. (1977). Understanding image intensities. *Artificial Intelligence*, 8:201–231.

[Jaffe és Carnevale, 1999] Jaffe, D. B. és Carnevale, N. T. (1999). Passive normalization of synaptic integration influenced by dendritic architecture. *J. Neurophysiol*, 82:3268–3285.

[Jutten és Herault, 1991] Jutten, C. és Herault, J. (1991). Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

[Knowlton és Squire, 1993] Knowlton, B. és Squire, L. (1993). The learning of natural categories: parallel memory systems for item memory and category-level knowledge. *Science*, 262:147–149.

[Komatsu, 1996] Komatsu, Y. (1996). Gabab receptors, monoamine receptors, and post-synaptic inositol trisphosphate-induced ca2+ release are involved in the induction of long-term potentiation at visual cortical inhibitory synapses. *J. Neurosci.*, 16:6342–6352.

[Komatsu és Iwakiri, 1993] Komatsu, Y. és Iwakiri, M. (1993). Long-term modification of inhibitory synaptic transmission in developing visual cortex. *NeuroReport*, 4:907–910.

[Levy, 1996] Levy, W. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, 6:579–590.

[Lisman, 1999] Lisman, J. (1999). Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-ca3 interactions. *Neuron*, 22:233–242.

[Ljung és Soderstrom, 1993] Ljung, L. és Soderstrom, T. (1993). *Theory and practice of recursive identification*. MIT Press, Cambridge, MA.

[Lőrincz és Buzsáki, 2000] Lőrincz, A. és Buzsáki, G. (2000). Two–phase computational model training long–term memories in the entorhinal–hippocampal region. In Scharfman, H., Witter, M., és Schwarz, R., editors, *The parahippocampal region: Implications for neurological and psychiatric diseases*, volume 911, pages 83–111. NYAS, New York.

[McCloskey és Cohen, 1989] McCloskey, M. és Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Bower, G. H., editor, *The Psychology of Learning and Motivation:*, volume 24, pages 109–164. NY: Academic Press.

[Olshausen és Field, 1996] Olshausen, B. és Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

[Olshausen és Field, 1997] Olshausen, B. és Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325.

[Póczos és Lőrincz, 2003] Póczos, B. és Lőrincz, A. (2003). Kalman-filter using local interaction. http://arxiv.org/abs/cs.AI/0302039.

[Rieke és mtsai., 1996] Rieke, F. D. W., de Ruyter van Steveninck, R., és Bialek, W. (1996). *Spikes - Exploring the neural code.* MIT Press, Cambridge, MA.

[Schachter, 1987] Schachter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13:501–518.

[Searle, 1992] Searle, J. (1992). *The rediscovery of mind.* Bradford Books, MIT Press, Cambridge, MA.

[Squire, 1992] Squire, L. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *J. Cog. Neurosci.*, 4:232–243.

[Treves és mtsai., 1999] Treves, A., Panzeri, S., Rolls, E., Booth, M., és Wakeman, E. (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Computation*, 11:601–631.