

hosszu.gabor.laszlo\_231\_24

# **Speciális mintarendszerek és magzati szívhangok informatikai modellezése**

MTA doktori értekezés tézisei

Készítette Hosszú Gábor  
a műszaki tudomány kandidátusa

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Elektronikus Eszközök Tanszéke

Budapest, 2024

## Tartalomjegyzék

|  |    |
|--|----|
| Bevezetés  | 2  |
| Kutatási célok   | 3  |
| Az előzmények rövid áttekintése  | 3  |
| Mintarendszerek evolúciójának informatikai leírása és modellezésének kérdése | 3  |
| Akusztikus magzati szívhangvizsgálat   | 5  |
| Új tudományos eredmények   | 6  |
| Eredmények hasznosíthatósága   | 20 |
| Mintaevolúció informatikai modellezése                                       | 20 |
| Magzati szívzörejek identifikációja  | 20 |
| Rövidítések és jelölések   | 21 |
| Legfontosabb saját közlemények   | 22 |
| Hivatkozott nem saját irodalom   | 24 |

### Bevezetés

Kutatásaim két pillérré támaszkodnak, az *adatelemzésre* és a *szimbolikus kommunikációra*; kutatási módszereim alapja pedig különböző mintázatok vizsgálata matematikai modellezés és informatikai eszközök alkalmazásával. Az emberi *írás (script)* általánosításával bevezettem a *mintarendszer (pattern system)* fogalmát, amely szimbólumok, valamint ezek használatát szabályozó szintaktikai szabályok (szintaxis) és elrendezési szabályok által meghatározott szimbolikus kommunikációs forma. A mintarendszerek közül az időbeli evolúcióval rendelkezőket vizsgáltam azzal a céllal, hogy a biológiai evolúció matematikai és informatikai rekonstrukciós módszereinek alkalmazásával modellezem egy utód-mintarendszer kialakulását elődjeiből. Feltételeztem továbbá, hogy a mintarendszerek evolúciójának kell, hogy matematikailag leírható alapjai legyenek. Az általam kidolgozott vizsgálati módszereket összefoglalóan *mintaevolúciónak* neveztem el.

Az időbeli fejlődést mutató mintarendszerek példái az informatikai értelemben vett karaktereket kódoló különböző kódrendszerek, amelyek evolúciója a távírászatban a Baudot-kódtól a Baudot–Murray-kódon át, az USA-ban használt ASCII kódtáblázaton és a különböző nemzeti kódtáblázatokon keresztül az önmagában is rendszeresen változó Unicode kódolásig terjednek. Az evolúciót mutató mintarendszerek másik példája a morzekód, amelynek fejlődése a Morse által kitalált rendszertől a Vail által módosított kódon (amerikai morzekódon) át a több lépésben kialakult nemzetközi morzekódig tart. Az evolúciós tulajdonságokkal rendelkező mintarendszerek további példája a zenei kotta, a matematikai jelölések és a vonalkódok létrehozásához szükséges szimbólumkészlet a használati szabályokkal. A mintarendszerek kiterjeszthetők több dimenzióra is, így mintarendszernek tekinthető az integrált áramköri tervezés grafikus szimbólumkészlete az áramköri konstrukciós és tervezési szabályokkal, valamint a QR-kódok (*quick-response code*) kialakításához szükséges szimbólumkészlet, szintaxis és elrendezési szabályok együttese. Ebben a témakörben számomra legérdekesebbnek az írások fejlődése tűnt, két okból is: *(i)* analógiát fedeztem fel a biológiai rendszerek és az írások között, *(ii)* felismertem, hogy az írások filogenetikájának meghatározására a legjobb eredmény az evolúciós modellezés matematikai eszközeivel várható.

Kutatásaim másik területe a *magzati szívhangvizsgálat*, amelyben azzal foglalkoztam, hogy miként lehet egy elosztott hash táblázat alapú adatbázist egyes ritkán megjelenő orvosi adatok gyűjtésére és értékelésére használni. Az elvégzett vizsgálatok célja a mérnöki gyakorlatban is ismert informatikai és matematikai modellezési módszerek alkalmazása volt, a kérdéses mért orvosi adatok rendszerszerű leírása és a megfigyelések közötti összefüggések feltárása érdekében.

## Kutatási célok

### *Az előzmények rövid áttekintése*

A mintarendszerek speciális területe, az írások filogenetikájával foglalkozik pl. Skelton [1]; továbbá a mintaevolúcióval rokon terület a számítógépes nyelvészet és beszédtechnológia. Ezekbe tartozik a nyelvek evolúciójának filogenetikai modellezése (pl. Gutkin és Sproat [2]), a nyelv absztrakt rendszerként való leírása (pl. Prószycki [3]), valamint a beszédfelismerés és a beszéd-szintézis (pl. Németh [4]). A mintaevolúciós kutatások eredményei szerepet kaphatnak a jövő kognitív infokommunikációs rendszereiben is (pl. Sallai [5]).

Az orvosi informatika területén az akusztikus magzati szívhangok vizsgálatának hazánkban komoly hagyományai vannak, ezek közé tartozik Kovács et al. [6]. Az orvosi diagnosztikába tartozó ezen kutatásba bekapcsolódva a fentebb jelzett irodalomban tárgyalt szívhangmonitor készülék segítségével a világon egyedülálló minőségben rendelkezésre álló, akusztikus úton nyert magzati szívhangok elemzésével foglalkoztam.

### *Mintarendszerek evolúciójának informatikai leírása és modellezésének kérdése*

Az első tézisemben a mintarendszerekkel kapcsolatosan bevezettem a *minta (pattern)* fogalmát, amely egyrészt szimbólumszekvenciát, másrészt jelszekvenciát jelent. A *szimbólumszekvencia (symbol sequence)* a mintarendszer szintaktikai szabályai szerint szerkesztett adatsorozat, ahol a *szintaktikai szabály (syntactic rule)* az az összefüggés, amely meghatározza egy adott üzenetből a szimbólumszekvencia létrehozását. A szimbólumszekvenciára példa a morzekódolt üzenet, egy integrált áramkör layoutjának tervrajza vagy akár egy adott írással létrehozott szöveg – mindegyik egy-egy meghatározott adatsorozattal megadva. A szintaktikai szabályoknak az integrált áramköri layoutok mint 3D mintarendszerek esetén az áramköri konstrukciós szabályok felelnek meg.

A *jelszekvencia (graph sequence)* egy szimbólumszekvencia megjelenítése valamilyen technológiával, a mintarendszerhez tartozó elrendezési szabályok szerint, ahol az *elrendezési szabály (layout rule)* egy összefüggés, amely meghatározza egy szimbólumszekvencia megvalósítását jelentő jelszekvencia jellemzőit. Jelszekvencia pl. egy fizikailag mért 1D jelsorozat, amely morzekódolt üzenetként (szimbólumszekvenciaként) értelmezhető; egy chipen kialakított 2D vagy 3D mintázat, amely integrált áramköri layoutként értelmezhető; vagy egy hordozó anyagon valamilyen technológiával megvalósított 1D vagy 2D felirat. Írások esetén a jelszekvenciának az alkalmazott technológiától függően a felirat vagy a dokumentum fogalmát feleltetem meg. Az elrendezési szabályoknak az integrált áramköri layoutok esetén a tervezési szabályok felelnek meg.

A kutatásom egyik fontos állomása volt, amikor felismertem, hogy a biológiai evolúció fogalmai és vizsgálati módszerei különösen jól alkalmazhatók az időben korábban használt írások fejlődésének informatikai modellezésére, és az ezekkel valamikor létrehozott, mára csak a biológiai őskövéletekhez hasonlóan fossziliaként fennmaradt feliratok (általánosan jelszekvenciák) analizálására. A mintarendszer (és ezen belül az írás) fogalmát megfeleltettem a biológiai taxonnak; a *taxon* ugyanis a biológiában általában egy faj vagy fajta, a mintaevolúcióban pedig egy mintarendszer (pl. egy írás) vagy ennek egy változata. A mintarendszereket leíró jellemzők a jellegek, amelyek szimbólumok, szintaxis vagy elrendezés szabályok. A szimbólum egyik tulajdonsága a *glif (glyph)*, amely egy topológiai jellegű rajzolat; egyetlen szimbólumhoz több glif is tartozhat. A szimbólum valamilyen technológiával való megvalósítása a szimbólumhoz tartozó egyik glif fizikai értelemben való létrejötte, ami egy alaki értelemben vett *jel (graph)*. Az alaki értelemben vett jel egy önálló szerepű, vizuálisan vagy másképpen észlelhető alaki egység.

A Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karán évek óta folytatok mintaevolúciós kutatást, amelyhez gépi tanulási és informatikai módszereket alkalmazok. A mintaevolúciós kutatás célja a vizsgált mintarendszerekkel különböző korokban készült és máig fennmaradt jelszekvenciák megfejtéséhez szükséges ismeretek felderítése. A vizsgálódásaim mintegy nyolcvan afrikai és eurázsiai írásra terjedtek ki, ilyenek pl. a különböző sémi, óitáliai, óhispán, óberber és kis-ázsiai alfabetikus és belső-ázsiai írások. Az általam vizsgált problematika elsődlegesen a steppei (sztyeppei) népek által használt írások evolúciós modellezése esetén kritikus. Ennek egyik oka, hogy Belső-Ázsia, a Kárpát-medence és a kettőt összekötő eurázsiai steppe évezred óta egyrészt különböző kultúrák találkozási helye volt, másrészt az itt élő lovas-nomád népesség időnként lakhelyet változtatva, nagy mozgékonyssággal haladt, elsősorban keletről nyugat felé. A változatos földrajzi, történelmi viszonyokból és technológiai ismeretekbeli különbségekből adódóan az eurázsiai steppén sokféle írás született, amelyek egy részéről igen hiányosak a szakirodalomban jelenleg fellelhető ismeretek. Előzetes adatokból ismert tény, hogy ezek Belső-Ázsiában alakultak ki, de az evolúciójuk a paleográfiai szakirodalom alapján homályba vész. Ezeket a körülményeket figyelembe kellett vennem a kutatási célok kijelölésénél. Számos tudományág vizsgálja a fejlődésüket, ezek eredményeit kiegészítve az általam kidolgozott informatikai módszerekkel megvalósítottam a régészeti lelőhelyekről előkerülő, nagyszámú, de gyakran töredékes és meghatározhatatlan korú jelszekvenciának (feliratnak) evolúciós rendbe történő elhelyezését, és ezzel megfejtésük elősegítését.

A vizsgált írások mint speciális mintarendszerek sajátosságait figyelembe véve különböző szintű informatikai és matematikai modellezési eljárásokat dolgoztam ki, amelyek eltérő feltételek között használhatók. Ezek közül a legegyszerűbb a *szimbólumleszármazási lánc* kialakítása, amely evolúciósan összefüggő szimbólumok közötti leszármazási sor kialakítása páronként vett topológiai eltérések minimalizálásával egymástól jelentősen eltérő mintarendszerek evolúciójának modellezésére. A szimbólumleszármazási lánc olyan esetekben alakítható ki, amikor a leszármazásban egymás után következő mintarendszerek szimbólumai elegendően különböznek egymástól. Ezek a kérdések képezik az első tézisem megállapításait.

A második tézisemben a mintarendszerek mint taxonok hasonlósági viszonyait vizsgáló *fenetikai* modellezéssel foglalkozom. A mintarendszerek vizsgálatának első lépése a *jellegetervezés (feature engineering)* – más néven *jellegválasztás (feature selection)* – amely a bemeneti nyers adatok feldolgozásának és rendszerezésének folyamata a gépi tanulási algoritmusok működéséhez hasznos jellegek meghatározása érdekében. Felismertem, hogy azokban az esetekben, amikor az egymáshoz közeli mintarendszerek (mint taxonok) evolúciósan rokon szimbólumaihoz tartozó glifváltakat átfednek egymással, pusztán az egyes szimbólumok mint jellegek vizsgálata alapján nem állapítható meg a jellegek változásának iránya, vagyis a *jellegpolaritás*. Ennek megoldására létrehoztam a mintarendszerek rokon jellegeinek *hasonlósági jellegcsoportjait*, amelyek szöveges adatbázisok. A hasonlósági jellegcsoportok alkalmazásának előnye, hogy a különböző taxonokba tartozó, evolúciósan rokon (*homológ*) és egymással sok esetben megegyező jellegek között nem kell leszármazási sort kialakítani. A hasonlósági jellegcsoportokra épülve a biológiai fenetika modellezési módszereit felhasználva kifejlesztettem a mintarendszerek *egyszerű fenetikai modellezését*.

Kidolgoztam a *kiterjesztett fenetikai modellt*, mint az egyszerű fenetikai modell kiegészítését eredetmodellekkel, amelyek technológiai és egyéb körülményektől függő adatok. A kiterjesztett fenetikai modellezés keretében csak akkor kerülhetnek különböző mintarendszerbe tartozó jellegek azonos hasonlósági jellegcsoportba, ha az ezekhez a jellegekhez hozzárendelt lehetséges eredetmodellek mindegyike alapján az egyes mintarendszerekhez tartozó, egymásnak megfeleltethető jellegek homológok (közös leszármazásból eredőnek) tekinthetők. Ezzel az eljárással csökkentettem annak az esélyét, hogy evolúciósan függetlenül kialakult, de egymáshoz hasonló jellegek (*homoplázia* esete) tévedésből azonos hasonlósági jellegcsoportba kerüljenek.

Kidolgoztam a jelszekvenciáknak az egyes írásokba történő hagyományos besorolásának igazolására egy klaszterezésre és főkomponens elemzésre épülő összetett algoritmust. Bevezettem különböző *holofiletikus indexeket* annak számszerűsítésére, hogy egy filogenetikai fában a jelszekvenciák milyen mértékben alkotnak holofiletikus csoportot, vagyis az általuk alkotott részfa leveleinek hányad része tartozik azonos mintarendszerbe.

A harmadik tézisemben a kiterjesztett fenetikai modellezésben közvetetten jelen lévő evolúciós információk meghatározása céljából bevezettem egy új fogalmat, a *szukcesszív eliminációt*, amelynek lényege a kiterjesztett fenetikai modellből adódó taxon–jelleg adatmátrixon végzett mátrixműveletekkel a vizsgált taxonok eredeti jellegkészletéből egy olyan szűrt jellegkészlet adatvezérelten történő meghatározása, amely ezen taxonok fejlődésének kezdeti állapotát írja le. Ennek alapján létrehoztam a vizsgált mintarendszerek kladisztikai modelljét; továbbá a szűrt jellegkészlet kezdeti állapotot leíró tulajdonságának validálása érdekében kidolgoztam egy többváltozós analízisen alapuló eljárást, amely a főkoordináta elemzés és a *k*-közép klaszterezés kombinációjára épül. Mindezekkel olyan összetett eljárást sikerült kifejlesztenem, amely lehetővé teszi az egyes mintarendszerek esetében rendelkezésre álló igen kevés adat alapján is egy közelítő evolúciós modell felállítását.

### ***Akusztikus magzati szívhangvizsgálat***

Bevezetem a zajszint fogalmát a magzati szívhangfelvételek előzetes zajosság szerinti osztályozása céljából. A mért adatok elemzéséből sikerült feltárnom egy diagnosztikai szempontból különleges esetet, amely két eltérő eredetű magzati szívzöreje jelentkezéséből áll, amelyek egyike egy jelenleg még azonosíthatatlan fiziológiai elváltozásból adódhat. Ez a felismerés egyben alátámasztotta azt, hogy a hosszú idejű fonokardiográfiai vizsgálatok azért is szükségesek, mert a magzat a mérés alatt helyét változtathatja, és az elhelyezkedés függvényében egyes esetekben ez lehetőséget ad arra, hogy többféle szívzöreje is detektálható legyen ugyanazon a felvételen.

A rendelkezésre álló nagy mennyiségű szívhangfelvételtől nyerhető információ növelése érdekében javaslatot tettem egy elosztott hash táblázat alapú adatbázisra, amely szakértői rendszerként működve képes lehet a rendkívül ritkán jelentkező események detektálására és az ennek kapcsán gyűjtött adatok rendszerszerű felhasználására.

Az akusztikus magzati szívhang kutatásáról érdemes megjegyezni, hogy ez a 2000-es évek kezdetén teljesen új irány volt, aminek eredményeképpen számos színvonalas közlemény mellett egy hazai IEEE-cikk is megjelent [D20]. Abban az időben a világon lényegében csak két hely volt, ahol ezt vizsgálták, az egyik a Budapesti Műszaki és Gazdaságtudományi Egyetemen, a másik Mexikóban. Ma már világszerte legalább kétszáz magas színvonalú kutatóhelyen foglalkoznak akusztikus magzati szívhangkutatással, bár a gyakorlatban még mindig az ultrahangot használják annak ellenére, hogy annak teljesen ártalmatlan volta nincs egyértelműen bizonyítva. Véleményem szerint ezek a fonokardiográfiai mérések is részt fognak kapni a magzatvizsgálatokban annak bizonyos korlátait kiszélesítve. Jó példa erre, hogy a Gottsegen György Országos Kardiovaszkuláris Intézetben sem tudtak kimutatni olyan magzati szívzörejt, amely az általunk használt módszerrel nem található meg.

## Új tudományos eredmények

### 1. tézis: Kidolgoztam az időben változó mintarendszerek jellegei evolúciójának matematikai modellezéssel történő vizsgálatát.

*1.1. altézis: Meghatároztam a mintarendszerek evolúciója modellezéséhez szükséges fogalomrendszert, továbbá kidolgoztam a minimális entrópia elvet, amely a parszimónia elv mintaevolúciós változata [D1], [D2], [D27], [D33], [D44].*

Meghatároztam az általam alkalmazott jel és szimbólum fogalmakat a jelelmélet (szemiotika) szerint: a szemiotikai értelemben vett *jel (sign)* bármilyen jelentéshordozó dolog, van egy jelentése, amely más mint saját maga, és amely ezért információt tud közölni valakivel, aki a jelet értelmezi vagy dekódolja; a *szimbólum (symbol)* a szemiotikai értelemben vett jel egyik típusa; olyan jel, amelyben a helyettesítő tárgyhoz egy, az értelmező tudatában lévő eszme révén kapcsolódik. A szimbólum tulajdonságai a hozzá tartozó *glif*, a jelentés vagy hangérték, a használati időszak, a használat térbeli kiterjedése stb. Különbséget tettem a szemiotikai értelemben vett jel és az alakzati értelemben vett *jel (graph)* fogalma között, az utóbbi a szimbólum egyik glifjének valamilyen technológiával történő megvalósítása.

A mintaevolúció általam vizsgált eseteiben a biológiai evolúcióhoz képest sokkal gyakoribb a homoplázia, amelynek oka, hogy a mintarendszerek jellegeinek változatossága kisebb a biológiai fajok örökítő anyagának sokféleségénél. Az időbeli evolúcióval rendelkező írások – mint speciális mintarendszerek – alkalmas modellezési tárgyak a mintaevolúciós eljárások ellenőrzésére, így erre a problematikára alkalmaztam a kifejlesztett módszereket. A mintarendszert alkotó szimbólum felfogható mint egy jelszekvenciabeli *jel (graph)* elvonatkoztatása; a szimbólum típusai az írások esetében a *(i)* hangértékkel vagy jelentéssel bíró graféma, a *(ii)* tamga vagy *(iii)* jelentés nélküli díszjel.

Meghatároztam a szimbólum egyik mintarendszerből a másikba történő átvételének szükséges feltételeit, amelyek: *(i)* alakzati hasonlóság; *(ii)* hangértékbeli vagy jelentésbeli kapcsolat; *(iii)* technológiai, időbeli és térbeli kapcsolat és *(iv)* motiváció. Ezen feltételek együttesen rendszerint nem teljesülnek, ezért bevezettem a *minimális entrópia elvet* a legvalószínűbb előd megtalálásához: eszerint a lehetséges leszármazási kapcsolatok közül azt kell kiválasztani, amelynél a legkevesebb feltételezéssel teljesülnek az előbbi feltételek. A minimális entrópia elv a *parszimónia elv* (takarékosági elv) mintaevolúciós alkalmazása.

*1.2. altézis: Kidolgoztam a szimbólumok alakzatának karakterisztikusságát jellemző glif-összetettségi mutatót, amellyel numerikusan figyelembe vehető a glifek által hordozott változó bonyolultságú topológiai információ [D2], [D33], [D34], [D38], [D46].*

Bevezettem a glifek jellegzetességét leíró *glif-összetettségi mutató* (GCP) fogalmát a vizsgált mintegy nyolcvan írás szimbólumaihoz tartozó glifek topológiájának statisztikai elemzése alapján. Ezzel lehetővé vált ennek az eddig csak szubjektíven leírható jellemzőnek a számszerűsítése:

$$GCP = (\text{sgn } N) \left( \sum_{i=1}^N e_i^{(N)} - 1 \right) + 2(\text{sgn } L) \left( \sum_{j=1}^L e_j^{(L)} + 1 \right) + A + \sum_{k=1}^R e_k^{(R)} + (\text{sgn}(S - 1))S^2, \quad (1)$$

ahol  $N$  a csomópontok (*node*) száma;  $e_i^{(N)}$  az  $i$ -edik csomóponthoz tartozó élek (*edge*) száma;  $L$  az át nem lapoló hurkok (*loop*) száma;  $e_j^{(L)}$  a  $j$ -edik hurok csomópontjai közötti élek száma;  $A$  az egyik hurokba sem tartozó ívek (*arch*) száma;  $R$  a glifben tükrözött vagy párhuzamos transzformációs (*reflexion*) kapcsolatok száma;  $e_k^{(R)}$  a glif vonalainak  $k$ -edik tükrözési vagy eltolási transzformációs kapcsolatába tartozó élek száma, továbbá  $S$  a glif topológiáját alkotó egymástól elkülönült (*separated*) alakzatok száma.

**1.3. altézis:** Létrehoztam a szimbólumok glifjeinek leírására szolgáló alakzati jellemzők rendszerét, és kidolgoztam a glifek hasonlósága számszerűsítésének módszerét [D33], [D34], [D38], [D46].

Összefüggést állítottam fel két glif topológiai tulajdonságait leíró *alakzati jellemzői* és a különbözőségük között, amivel a glifek összehasonlítását számszerűsítettem. Egy adott mintarendszerbe tartozó glif alakzati jellemzőit  $\mathbf{g}$  vektorral jelölöm, amelynek  $(g_1, g_2, \dots)$  elemei a glif alakzati jellemzőinek számszerűsített értékei. A két glif ( $\mathbf{g}^1$  és  $\mathbf{g}^2$ ) különbözőségét (2) szerint határozom meg.

$$d(\mathbf{g}^1, \mathbf{g}^2) = \frac{\sqrt{\sum_{i=1}^{N_G} w_i (g_i^2 - g_i^1)^2}}{\sqrt{\sum_{i=1}^{N_G} w_i (\max(g_i^1, g_i^2))^2}} \quad (2)$$

ahol  $N_G$  az alakzati jellemzők száma (vizsgálataimban  $N_G = 13$ ), a  $(w_1, w_2, \dots, w_{N_G})$  az alakzati jellemzők fontosságát kifejező súlytényezők, amelyek értékeit a vizsgált mintarendszereknek a szakirodalomból ismert leszármazási kapcsolatai alapján heurisztikus úton állapítottam meg.

**1.4. altézis:** Felismertem, hogy a mintarendszerek jellegeinek változásai felfoghatók elemi jellegtranszformációk sorozataként, továbbá a jellegtranszformációk különböző jellegfejlődési elvek szerint érvényesülnek [D1], [D2], [D33], [D38], [D44], [D46].

Az általam vizsgált mintarendszerekbeli jellegek közül a szimbólumok fejlődését elemi jellegtranszformációk ( $M_1, M_2, \dots, M_i, \dots, M_{21}$ ) mint operátorok sorozataként határoztam meg, amely operátorok jelentése pl. alakváltozat, díszesítés, értelmezés alapú alakítás, helyettesítés, hurokképzés stb. Egy új  $\mathbf{g}^1$  glif létrejöttét a  $\mathbf{g}^0$  régiből a következő operátoros összefüggéssel fejeztem ki:

$$\mathbf{g}^1 = \prod_i (M_i, 1 \leq i \leq 21) \mathbf{g}^0 \quad (3)$$

Megmutattam, hogy a jellegtranszformációkat kifejező  $M_i$  operátorból tetszőleges számú alkalmazható egymás után. Két,  $\mathbf{g}^1$  és  $\mathbf{g}^0$  glif közötti rokonság annál nagyobb, minél kevesebb operátor kell a  $\mathbf{g}^1$  létrejöttéhez a  $\mathbf{g}^0$ -ból (vö. minimális entrópia elv). Meghatároztam azokat a jellegfejlődési elveket, amelyek érvényesültek a vizsgált mintarendszerek evolúciójában, pl. adaptáció, alakváltozat, átvétel, egységesülés stb.

**1.5. altézis:** Kidolgoztam a szimbólumszármaztatás módszerét, amely a szimbólum-glifek hasonlósági viszonyai alapján a szimbólumok leszármazási láncain alapul [D34], [D38], [D46].

Bevezettem a mintarendszerek közötti rokonsági kapcsolatok feltárásának az egyes szimbólumok fejlődése vizsgálatán alapuló módszerét, amelyhez szimbólumleszármazási láncokat (*Symbol Lineage Chain*, SLC) alkottam a láncban szomszédos szimbólumok jellegzetes glifjei közötti, (4)-beli  $E_{SLC}$  hibafüggvénnyel összegzett topológiai eltérések egész láncra vett négyzetösszegének minimalizálásával. Az egyes szimbólumleszármazási láncokat úgy hoztam létre, hogy a különböző származási lehetőségek közül azt választottam, amely a láncban lévő szimbólumok glifjeinek legkisebb eltérését jelentette.

$$E_{SLC} = \frac{1}{N_{SLC} - 1} \sum_{i=1}^{N_{SLC}} [d(\mathbf{g}^i, \mathbf{g}^{i+1})]^2, \quad (4)$$

ahol  $\mathbf{g}^i$  a vizsgált  $SLC = (\mathbf{g}^1, \mathbf{g}^2, \dots, \mathbf{g}^{N_{SLC}})$   $i$ -edik szimbólumának glifje, továbbá  $d(\mathbf{g}^i, \mathbf{g}^{i+1})$  az  $i$ -edik és az  $(i+1)$ -edik glif eltérése, valamint  $N_{SLC}$  a vizsgált SLC elemszáma és  $(N_{SLC} - 1)$  a glifek közötti páronkénti eltérések száma. A (4) összefüggésben a négyzetre emelés célja a nagy alakzati eltérések súlyának növelése a kisebb eltérésekkel szemben, mivel a kisebb eltérések az egyéni kézírásból eredő változatokból is adódhatnak (pl. egy glif egyik vonalának a szokásoshoz képesti túlhúzása), így kevésbé alkalmasak a leszármazási kapcsolatok leírására.

A szimbólumok rokonsági kapcsolatainak meghatározhatósága függ a glifjeik jellegzetességétől, így minél nagyobb egy SLC elemeinek átlagos GCP-je ( $\overline{GCP}_{SLC}$ ) – vagyis az (5) kifejezés értéke – annál megbízhatóbb egy adott SLC mint leszármazási modell.

$$\overline{GCP}_{SLC} = \frac{1}{N_{SLC}} \sum_{i=1}^{N_{SLC}} GCP(\mathbf{g}^i), \quad (5)$$

ahol  $GCP(\mathbf{g}^i)$  a  $\mathbf{g}^i$  glif (1) szerinti glif-összetettségi mutatója. Meghatároztam a szimbólumleszármazási láncok alkalmazási korlátját, ami abból adódik, hogy a glifek alakzati változékonysága nagyobb lehet a glifek evolúciós változásánál. Így csak kellően eltérő mintarendszerek esetén lehet a szimbólumok glifjeinek eltérése alapján megállapítani a különböző mintarendszerekbe tartozó, egymással rokon, és emiatt hasonló szimbólumok (általánosan: jellegek) leszármazási irányát, vagyis a jellegpolaritást.

**2. tézis: Kifejlesztettem a mintaevolúció fenetikai modellezésének módszereit a fenetikában ismert eljárások módosításával, és kidolgoztam a jelszekvenciák előzetes osztályozását minősítő összetett többváltozós adatelemző eljárást.**

**2.1. altézis:** *Bevezettem a hasonlósági jellegcsoport fogalmát a mintarendszerek vizsgálatához szükséges jellegtervezés végrehajtására, amely az egyes jellegek taxonbeli előfordulását határozza meg. Ennek alapján létrehoztam a vizsgált mintarendszerek egyszerű fenetikai modelljét [D33], [D44].*

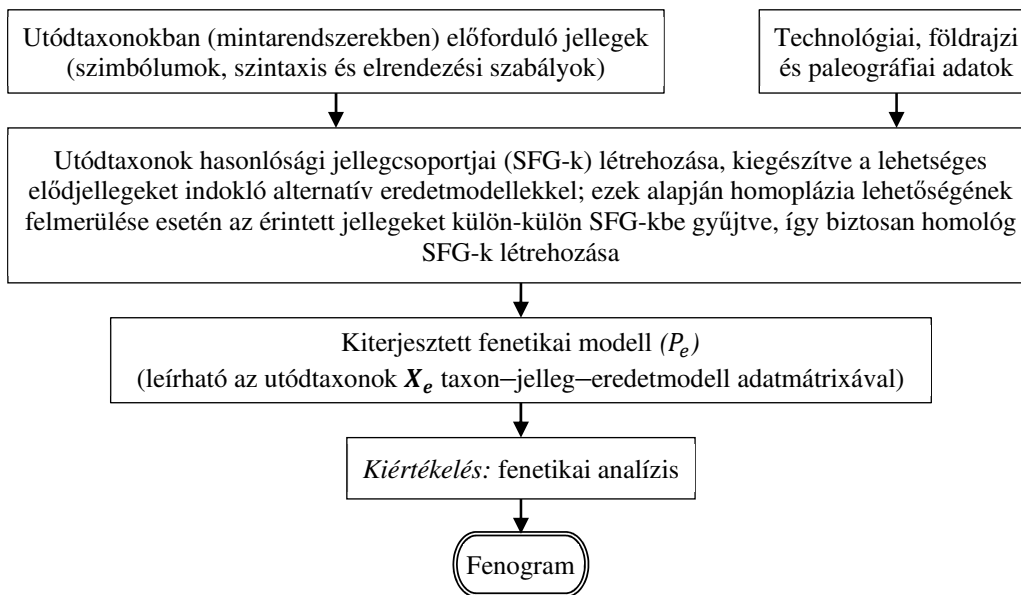
A szimbólumleszármazási láncoknak a hasonló mintarendszerek esetére a fentiekben leírt módszertani korlátján túlléptem azzal, hogy az egyes jellegekhez létrehoztam egy-egy adatbázist, az ún. *hasonlósági jellegcsoportot* (SFG). Az SFG-k alkalmazásának előnye, hogy a feltételezhetően rokon jellegek közötti jellegpolaritást nem vizsgálja, hanem egy taxon egy jellegéhez tartozó jellegállapotot a taxonnak az adott jelleg leíró SFG-ben való részvétele jelenti. Az SFG-k halmazára bevezettem az *egyszerű fenetikai modell* fogalmát. Az ebből adódó *taxon-jelleg adatmátrixot*



hierarchikus klaszterezéssel (UPGMA, WPGMA és Ward-módszerrel), valamint többdimenziós skálázás után  $k$ -közép klaszterezéssel értékeltem ki.

**2.2. altézis:** *Kidolgoztam a mintarendszerek filogenetikai modellezésének fő akadályát jelentő nagyszámú homoplázia algoritmizált kezelésére a kiterjesztett fenetikai modellezési eljárást, amely alternatív leszármazási lehetőségeket leíró eredetmodellek rendszerére épül. Felismertem továbbá a vizsgált mintarendszerek fejlődésében egyfajta retikuláris evolúció hatását [D1], [D30].*

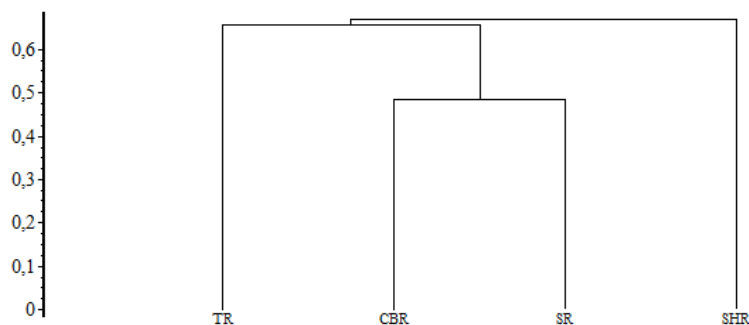
Kidolgoztam a *kiterjesztett fenetikai modellt* ( $P_e$ ), amely az egyszerű fenetikai modellt kiegészíti a szóba jöhető lehetséges elődjellegek figyelembe vevő, *eredetmodelleknek* nevezett alternatív evolúciós megfontolásokkal (1. ábra). A kiterjesztett fenetikai modell kialakításakor csak olyan jellegeket csoportosítottam egy SFG-be, amelyek esetén a rendelkezésre álló eredetmodellek alapján a homoplázia lehetőségét kizárhattam.



1. ábra: A kiterjesztett fenetikai modell előállításának és fenetikai kiértékelésének folyamata

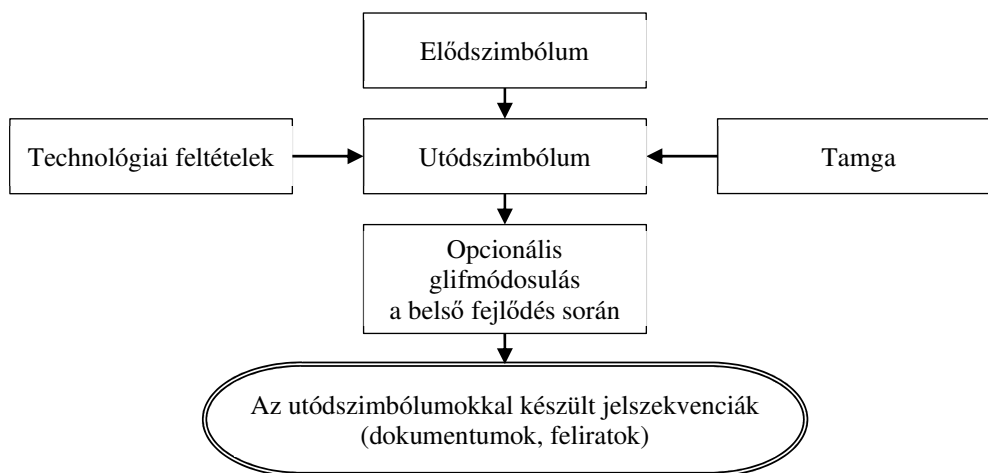
Az általam kifejlesztett  $P_e$  kiterjesztett fenetikai modell leírható egy  $N_t \times N_f \times N_o$  méretű,  $X_e$ -vel jelölt 3D bináris taxon-jelleg-eredetmodell adatmátrixszal, ahol  $N_t = 31$  a  $P_e$ -beli taxonok száma,  $N_f = 119$  a jellegek száma és  $N_o = 5$  az SFG-nkénti eredetmodellek maximális száma. A taxonok halmaza magában foglalja a vizsgált utódtaxonokat ( $N_t^d = 4$ ), a lehetséges elődtaxonokat ( $N_t^a = 26$ ) és egy további hipotetikus taxont (HTU), amely az elődtaxonok egymásra hatásából adódó utódjellegek közvetlen elődjét képviseli. A  $P_e$  az utódtaxonokra nézve átfogó, vagyis tartalmazza az összes jellegüket; így a  $P_e$  alapján az utódtaxonokra vonatkozóan végezhető fenetikai analízis.

Megmutattam, hogy a vizsgált taxonok különbözőségének leírására a jellegállapotok meglétében való egyezéseket kiemelő Sørensen–Dice-együttható a legalkalmasabb, mivel a jellegek túlnyomó része szimbólum, és egy szimbólum hiánya egy taxonban nem jellegzetes, tekintettel arra, hogy az összes jellegként figyelembe vett szimbólumnak csak kisebb része szerepel egy-egy taxonban. Igazoltam, hogy klaszterezéshez legalkalmasabb a WPGMA (2. ábra) és a klaszterek homogenitását optimalizáló Ward-módszer használata, mivel a klaszterekben eltérő számú taxon szerepel.



2. ábra: Az utódtaxonok WPGMA-val számított fenogramja

Felismertem, hogy egyes mintarendszerekben az utódjellegek egy része nem egyszerű leszármazással, hanem retikuláris evolúcióval jött létre. A vizsgált esetben az utódjelleg egy szimbólum, és tulajdonságait a glifje kivételével egy adott eredetmodellbeli elődjellegből nyeri, míg a glifje egy, az eredetmodellben nem szereplő elődtaxonba tartozó szimbólumból származik; az eddigi vizsgálatokban a glif forrása valamilyen tamga volt. Feltételezésem szerint a tamgák azért hatottak az utódtaxonokra, mert az utódtaxonokkal készült jelszekvenciákat az átvevő népszerűség az elődtaxonokétól eltérő, a tamgákhoz használt technológiával valósította meg (3. ábra).



3. ábra: A technológiai feltételek szerepe a tamgák hatásában

**2.3. altézis:** Kifejlesztettem a jelszekvenciák előzetes csoportosításának igazolására egy hierarchikus klaszterezésre és főkomponens-elemzésre épülő összetett algoritmust. Ennek keretében bevezettem a holofiletikus indexet annak számszerűsítésére, hogy egy filogenetikai fában az egy osztályba tartozó jelszekvenciák hányad részben alkotnak holofiletikus részfát; az összes jelszekvenciára vonatkozó együttes holofiletikus indexet; a súlyozott holofiletikus indexet annak figyelembe vételére, hogy az egyes jelszekvenciák hányad részben tartalmazzák az egy osztályba tartozó jelszekvenciákban előforduló jellegeket; és az összes jelszekvenciára vonatkozó súlyozott együttes holofiletikus indexet [D6], [D8], [D27].

Hierarchikus klaszterezéssel létrehoztam a vizsgált mintarendszerekkel készült, fennmaradt és megfejtett jelszekvenciák fenogramját, ahol a fa levelei a jelszekvenciák. Felhasználva azt a tényt, hogy a fenogram egy részfája holofiletikus csoportot képvisel, ha minden levele azonos mintarendszerbe tartozik, meghatároztam a fenogramban a holofiletikus csoportokhoz tartozó levélobjektumoknak az összes levélobjektumhoz képesti arányát. Megmutattam, hogy annál jobbnak minősíthető a jelszekvenciák előzetes (*a priori*) osztályozása, minél nagyobb arányban tartoznak egy

vagy több holofiletikus csoportba az adott osztályba tartozó jelszekvenciák. A filogenetikai fa  $l_i$  levelei a vizsgálatba bevont objektumok, jelölje ezek halmazát  $\mathbf{l}$ ; számuk  $|\mathbf{l}| = n$  ( $n$  az objektumok száma);  $i \in \{1, \dots, n\}$ . Legyen  $v_j$  a filogenetikai fa közbülső csúcsa ( $j = 1, 2, \dots$ ), amelyek halmazát jelölje a  $\mathbf{v}$  vektor:  $v_j \in \mathbf{v}$ . Az  $n$  objektum (jelszekvencia) előzetesen különböző  $\mathbf{c}_p$  osztályokba (mintarendszerekbe) van csoportosítva; az osztályok halmaza  $\mathbf{c}$ , az osztályok száma  $|\mathbf{c}|$ . Az osztályokba sorolás kizárólagos és teljes, vagyis  $\mathbf{c}_p \cap \mathbf{c}_q = \emptyset \forall \mathbf{c}_p, \mathbf{c}_q \in \mathbf{c}$  esetén ( $p, q \in \{1, \dots, |\mathbf{c}|\}$ );  $\mathbf{l}(\mathbf{c}_p)$  az előzetesen a  $\mathbf{c}_p$  osztályba sorolt  $l_i$  levélobjektumok halmaza;  $\mathbf{l}(\mathbf{c}_p) \subseteq \mathbf{l}$ ; és  $\sum_{p=1}^{|\mathbf{c}|} |\mathbf{l}(\mathbf{c}_p)| = |\mathbf{l}|$ . Így igaz (6).

$$\sum_{p=1}^{|\mathbf{c}|} |\mathbf{l}(\mathbf{c}_p)| = n \quad (6)$$

Az  $\mathbf{l}^r(\mathbf{c}_p)$  a  $\mathbf{c}_p$  osztályhoz tartozó egyik *holofiletikus csoport*, ha  $\forall l_i \in \mathbf{l}^r(\mathbf{c}_p)$  levél objektumra igaz, hogy  $\exists v_j \in \mathbf{v}$  közbenső csomópont, amelytől  $\forall l_i \in \mathbf{l}^r(\mathbf{c}_p)$  levél kevésbé különbözik mint bármely más  $l_j \notin \mathbf{l}^r(\mathbf{c}_p)$  levél a fán;  $i, j \in \{1, \dots, n\}$ ;  $r = 1, 2, \dots$  a  $\mathbf{c}_p$  osztályhoz tartozó különálló  $\mathbf{l}^r(\mathbf{c}_p)$  holofiletikus csoportok sorszáma; így  $\mathbf{l}^r(\mathbf{c}_p) \subseteq \mathbf{l}(\mathbf{c}_p)$ . Az  $\mathbf{l}^r(\mathbf{c}_p)$  halmaz elemeinek száma  $|\mathbf{l}^r(\mathbf{c}_p)|$ . Bevezettem egy adott  $\mathbf{c}_p$  osztályhoz (mintarendszerhez) tartozó *holofiletikus index* ( $HI_{\mathbf{c}_p}$ ) mértéket. A  $HI_{\mathbf{c}_p}$  egy filogenetikai fán lévő objektumok  $\mathbf{c}_p$  osztályának azon részén alapul, amelyek egy ( $r = 1$ ) vagy több ( $r > 1$ ) különálló  $\mathbf{l}^r(\mathbf{c}_p)$  holofiletikus csoporthoz tartoznak:

$$HI_{\mathbf{c}_p} = \frac{\sum_r |\mathbf{l}^r(\mathbf{c}_p)|}{|\mathbf{l}(\mathbf{c}_p)|}; \mathbf{l}^r(\mathbf{c}_p) \cap \mathbf{l}^s(\mathbf{c}_p) = \emptyset, \text{ ha } r \neq s; r, s = 1, 2, \dots; p \in \{1, \dots, |\mathbf{c}|\} \quad (7)$$

Az *együttes holofiletikus index* ( $JHI$ ) az  $l_i$  levélobjektumok halmazának azon részét jelöli, amelyek bármely  $\mathbf{c}_p$  osztályhoz tartozó  $\mathbf{l}^r(\mathbf{c}_p)$  holofiletikus csoportok valamelyikébe tartoznak:

$$JHI = \frac{\sum_{p=1}^{|\mathbf{c}|} \sum_r |\mathbf{l}^r(\mathbf{c}_p)|}{\sum_{p=1}^{|\mathbf{c}|} |\mathbf{l}(\mathbf{c}_p)|} = \frac{\sum_{p=1}^{|\mathbf{c}|} (HI_{\mathbf{c}_p} \cdot |\mathbf{l}(\mathbf{c}_p)|)}{n} \quad (8)$$

A  $HI_{\mathbf{c}_p}$  és a  $JHI$  a filogenetikai fát alkotó objektumoknak a  $\mathbf{c}$  osztályozáshoz képesti elkülönültségét írja le. Ezen indexek értékészletére igaz, hogy  $HI_{\mathbf{c}_p}, JHI \in [0, 1]$  ( $p \in \{1, \dots, |\mathbf{c}|\}$ ). Ha a  $HI_{\mathbf{c}_p}$  vagy a  $JHI$  értéke közel 1, akkor ez alátámasztja az  $l_i$  objektumok ( $i = 1, \dots, n$ ) előzetes osztályozását a  $\mathbf{c}_p \in \mathbf{c}$  osztályokba.

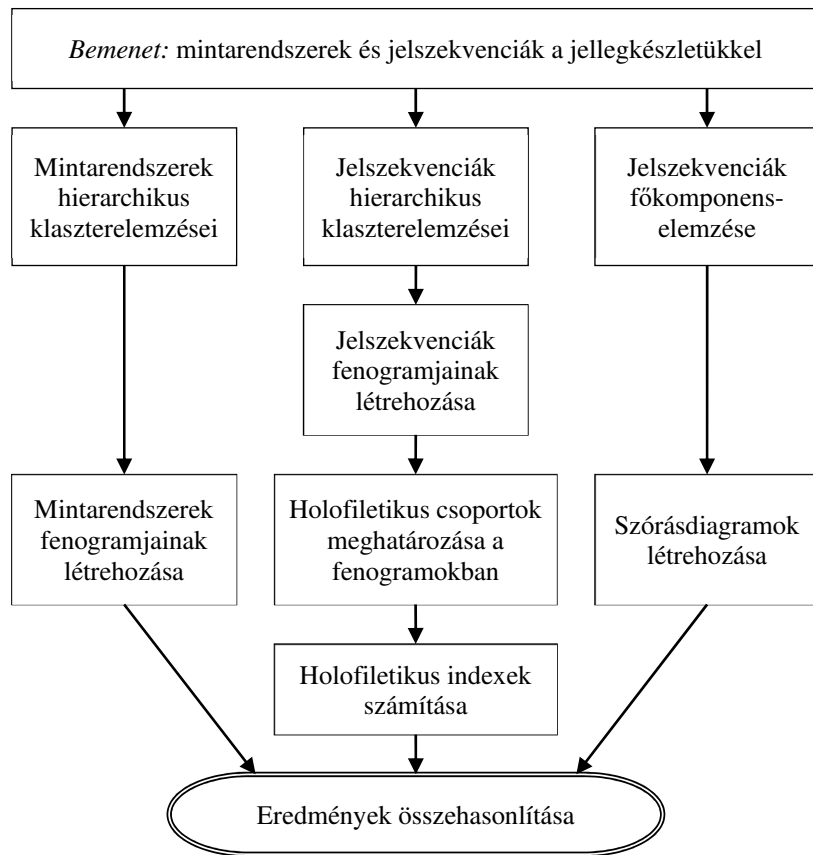
Bevezettem egy adott  $\mathbf{c}_p$  osztályhoz tartozó *súlyozott holofiletikus indexet* ( $WHI_{\mathbf{c}_p}$ ) a jelszekvenciabeli jellegek számának figyelembe vételéhez. Legyen az  $f_i$  *jellegszám* az  $l_i$  jelszekvenciabeli jellegek száma,  $i \in \{1, \dots, n\}$ . Jelölje  $\mathbf{F}^r(\mathbf{c}_p)$  az  $\mathbf{l}^r(\mathbf{c}_p)$ ,  $r = 1, 2, \dots$  holofiletikus csoportba tartozó,  $l_i \in \mathbf{l}^r(\mathbf{c}_p)$  jelszekvenciákban lévő  $f_i$  jellegszámok halmazát, vagyis  $f_i \in \mathbf{F}^r(\mathbf{c}_p)$ . Jelölje  $f_{\mathbf{c}_p}$  a  $\mathbf{c}_p$  osztályra jellemző jellegszámot, vagyis azon jellegek számát, amelyek legalább egyszer szerepelnek a  $\mathbf{c}_p$  osztályba tartozó  $l_i \in \mathbf{l}(\mathbf{c}_p)$  jelszekvenciák valamelyikében. Ezek felhasználásával bevezettem a  $WHI_{\mathbf{c}_p}$  súlyozott holofiletikus indexet:

$$WHI_{c_p} = \frac{\sum_r \sum F^r(c_p)}{f_{c_p} \cdot |I(c_p)|}; r = 1, 2, \dots; p \in \{1, \dots, |c|\}, \quad (9)$$

ahol  $\sum F^r(c_p)$  jelöli  $F^r(c_p)$  halmazok ( $r = 1, 2, \dots$ ) elemeinek összegét. Hasonlóan meghatároztam a *súlyozott együttes holofiletikus indexet (WJHI)* a következő összefüggést felállítva:

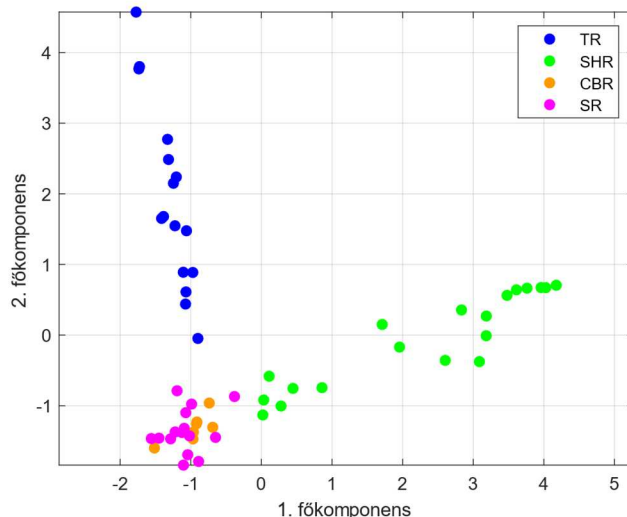
$$WJHI = \frac{\sum_{p=1}^{|c|} \sum_r \sum F^r(c_p)}{\sum_{p=1}^{|c|} (f_{c_p} \cdot |I(c_p)|)} = \frac{\sum_{p=1}^{|c|} (WHI_{c_p} \cdot f_{c_p} \cdot |I(c_p)|)}{\sum_{p=1}^{|c|} (f_{c_p} \cdot |I(c_p)|)} \quad (10)$$

A jelszekvencia-csoportosítás igazolására általam kifejlesztett többváltozós elemzés folyamatábráját bemutatja a 4. ábra.



4. ábra: A jelszekvencia-csoportosítást igazoló többváltozós elemzés fő lépései

A főkomponens-elemzésből (PCA) adódó főkomponensek az eredeti objektum–jelleg adatmátrix variancia–kovariancia mátrixának sajátvektorai; a 2D PCA eredményét bemutatja az 5. ábra.



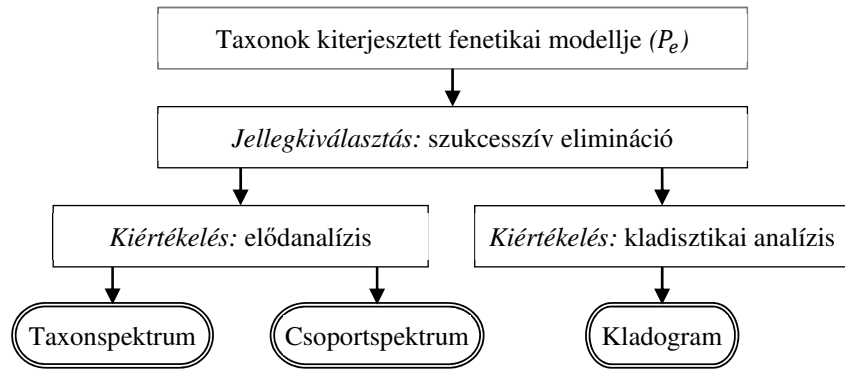
5. ábra: A jelszekvenciák mint objektumok ordinációjának 2D szórásdiagramja

A szórásdiagram egyes mintarendszerekhez (TR és SHR) tartozó objektumok (jelszekvenciák) között egyértelmű elkülönülést mutat, ami arra utal, hogy ezek tekintetében a jelszekvenciák előzetes besorolása pontos, ezzel szemben a CBR és az SR mintarendszerekbe sorolt jelszekvenciák elkülönülése bizonytalan. A holofiletikus indexekből adódó és a PCA által nyújtott eredmények összhangban állnak. Hasonló eredményeket kaptam PCA helyett PCoA alkalmazásával, az utóbbinál akár euklideszi távolságot, akár Sørensen–Dice-különbözőséget használva.

**3. tézis: Kidolgoztam a szukcesszív eliminációt, mint egy iteratív jellegkiválasztást a kiterjesztett fenetikai modell kiértékelésére, ezzel lehetővé téve kladisztikai analízis elvégzését; valamint kifejlesztettem a jellegkiválasztás helyességét igazoló jelszekvencia-elemző eljárást.**

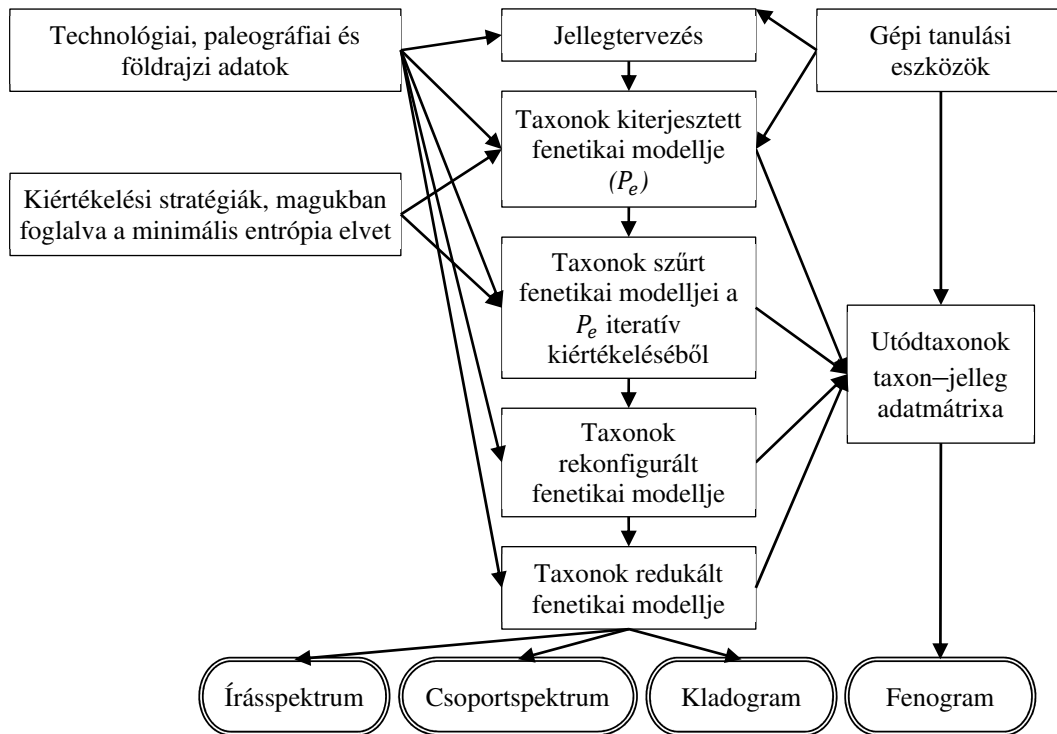
**3.1. altézis: Bevezettem egy új fogalmat, a szukcesszív eliminációt, amelynek lényege a kiterjesztett fenetikai modellben közvetetten meglévő evolúciós adatok általános elvek alapján, adatvezérelten történő kiválasztása; kidolgoztam továbbá a szűrt, a rekonfigurált és a redukált fenetikai modell fogalmát [D1], [D30].**

Kidolgoztam a  $P_e$  kiterjesztett fenetikai modellre alkalmazható, evolúciós analízis módszereket, amelyek célja leszármazási összefüggések feltárása az utódtaxonok és az elődtaxonok kapcsolatáról. Bevezettem a *szukcesszív elimináció* fogalmát, amely olyan jellegkiválasztás, amelyik általános elvek alapján kinyeri a kiterjesztett fenetikai modellben közvetetten meglévő evolúciós adatokat. A szukcesszív elimináció eredményét egyrészt elődanalízissel, másrészt kladisztikai elemzéssel értékelem ki. Az *elődanalízis* az utódtaxonok elődeit határozza meg az evolúció menetének felderítése nélkül, eredménye az általam bevezetett taxonspektrum, illetve csoportspektrum. A *taxonspektrum* és a *csoportspektrum* egy utódtaxon jellegeinek a biztos elődtaxonok, illetve elődtaxoncsoportok szerinti megoszlását ábrázolja, vagyis azt, hogy egy utódtaxon jellegei közül mennyi származik adott modell alapján valamelyik elődtaxonból, illetve elődtaxoncsoportból. A *kladisztikai analízis* az utódtaxonok evolúciós elágazási sorozatát határozza meg, eredménye a kladogram. A  $P_e$  kiterjesztett fenetikai modell evolúciós kiértékelését összefoglalja a 6. ábra.



6. ábra: A kiterjesztett fenetikai modell evolúciós elemzése

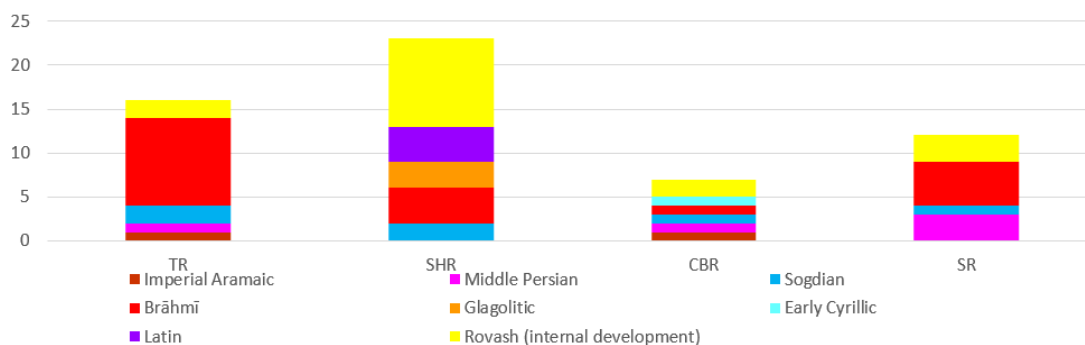
A  $P_e$ -ből a szukcesszív elimináció (7. ábra) keretében azon SFG-k értékelhetők ki, amelyek valamilyen szempontból egyértelműek; pl. csak egyetlen elődtaxont tartalmaznak. Azonban ezek az összes SFG-nek csak töredékét jelentik, így csak ezekre alapozva az elemzést, csökken a modell *adatgazdagsága*, vagyis a felhasznált SFG-k száma. Ennek növeléséhez számításba vettem, hogy az SFG-k egy részében az elődtaxonok rokonok, az elődjellegeik hasonlók, így bizonytalan, hogy közülük melyik a tényleges előd. A hasonló elődtaxonokból elődtaxoncsoportokat hozva létre az egyértelmű SFG-k száma nő azon SFG-kkel, amelyekben az elődtaxonok azonos csoportba tartoznak; így nőhet a modell *adatgazdagsága*. Megjegyzendő, hogy ezzel a modell *felbontása* csökken, hiszen elődtaxoncsoportok használatakor elvileg nem kerülhet meghatározásra az, hogy egy elődtaxoncsoportból melyik elődtaxon a tényleges előd.



7. ábra: A szukcesszív eliminációs eljárás és a kiértékelési lehetőségek

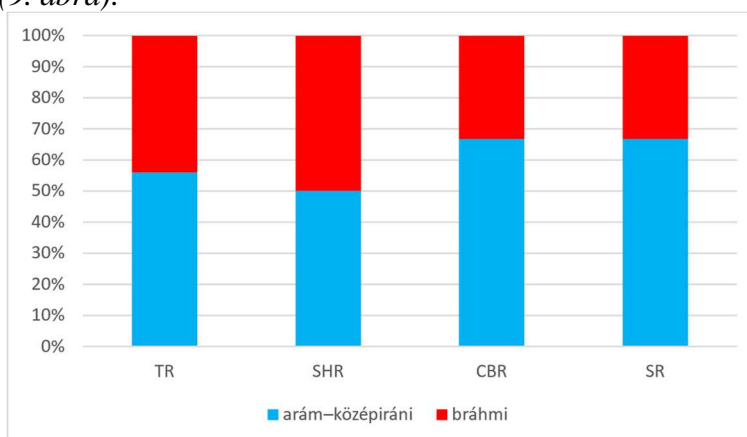
Meghatároztam a *szűrt fenetikai modell* fogalmát, ami a taxonok fenetikai modelljében szereplő SFG-k egy részét figyelmen kívül hagyva állítható elő. Ennek kiértékelésekor csak akkor hozható létre

taxonspektrum, ha az SFG-k mindegyike csak egyetlen elődtaxont tartalmaz, illetve akkor képezhető csoportspektrum, ha az egyes SFG-k mindegyike csak egyetlen elődtaxoncsoportot tartalmaz. Ha a fenetikai modell nem egyértelmű az elődtaxonokra vagy elődtaxoncsoportokra nézve, akkor valamilyen *kiértékelési stratégiát* megvalósító *feldolgozási operátorral* létrehozok egy egyértelmű szűrt fenetikai modellt. Kiértékelési stratégiára példa egy fenetikai modelltől a több elődtaxon tartalmazó SFG-k kihagyása. A szukcesszív elimináció adatvezérelt, mert a lépéseit csak a kiértékelési stratégiák és az adatszerkezet határozza meg, míg az SFG-kre vonatkozó egyedi megfontolásoknak nincs szerepük. A szukcesszív eliminációs eljárásban egy megfelelő feldolgozási operátort az eredeti, kiterjesztett fenetikai modellre ( $P_e$ ) alkalmazva az eredmény az elődtaxonokra egyértelmű szűrt fenetikai modell ( $P_{oneT}$ ) lesz, amelynek adatgazdagsága azonban kisebb a  $P_e$ -énél. A [D1] közleménybeli adatokból, a  $P_{oneT}$  alapján számított taxonspektrumokat bemutatja a 8. ábra.



8. ábra:  $P_{oneT}$  alapján az egyes utódtaxonok taxonspektrumai

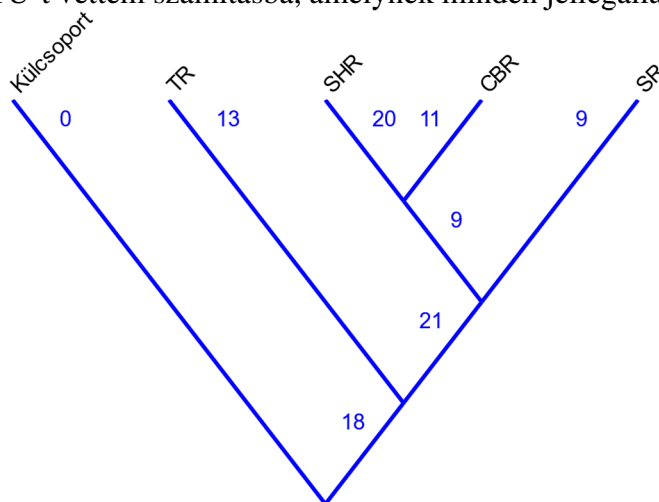
A szukcesszív elimináció egyes lépéseivel növelem a modell adatgazdagságát. Ennek egyik forrása az, hogy az egymáshoz hasonló elődtaxonok helyett elődtaxoncsoportokat veszek figyelembe, így azon SFG-k, amelyek csak azért nem egyértelműek, mert több, hasonló elődtaxon egymáshoz hasonló jellege közül nem dönthető el, hogy az aktuális utódjelleg melyikből származik, az elődtaxoncsoport bevezetésével egyértelművé válnak. A szukcesszív elimináció utolsó lépése a modell korlátozása az utódtaxonok fejlődésének adott fülkében zajlott szakaszára; ahol a mintaevolúciós *fülke (niche)* az egy mintarendszerrel készült jelszekvenciák készítőinek szokásos földrajzi, technológiai és kulturális körülményei. Az egy fülkére korlátozott modelltől képeztem az utódtaxonok csoportspektrumait, amelyeket normalizálva kiegyenlítődnek az egyes utódtaxonokhoz tartozó SFG-k eltérő számából adódó különbségek (9. ábra).



9. ábra: A kezdeti fülkére redukált fenetikai modellen alapuló normalizált csoportspektrumok

**3.3. altézis:** Megmutattam, hogy kladisztikai elemzés hajtható végre a kiterjesztett fenetikai modellből szukcesszív eliminációval meghatározott kezdeti közös jellegkészlet alapján [D9].

Felismertem, hogy ha a szukcesszív eliminációs eljárással (7. ábra) a [D1] közleménybeli utódtaxonok kiterjesztett fenetikai modelljéből sikerül olyan fenetikai modellt létrehozni, amelyben lévő SFG-k mindegyikében csak egyetlen elődtaxoncsoport található, akkor az utódtaxonokra elvégezhető a kladisztikai analízis, feltéve, hogy volt egy feltételezett közös elődjük (HTU). Az evolúció eredményeképpen az utódtaxonok a HTU-beli jellegeik egy részét fokozatosan elvesztették, így kulcsoportként a HTU-t vettem számításba, amelynek minden jellegállapota 1; lásd 10. ábra.

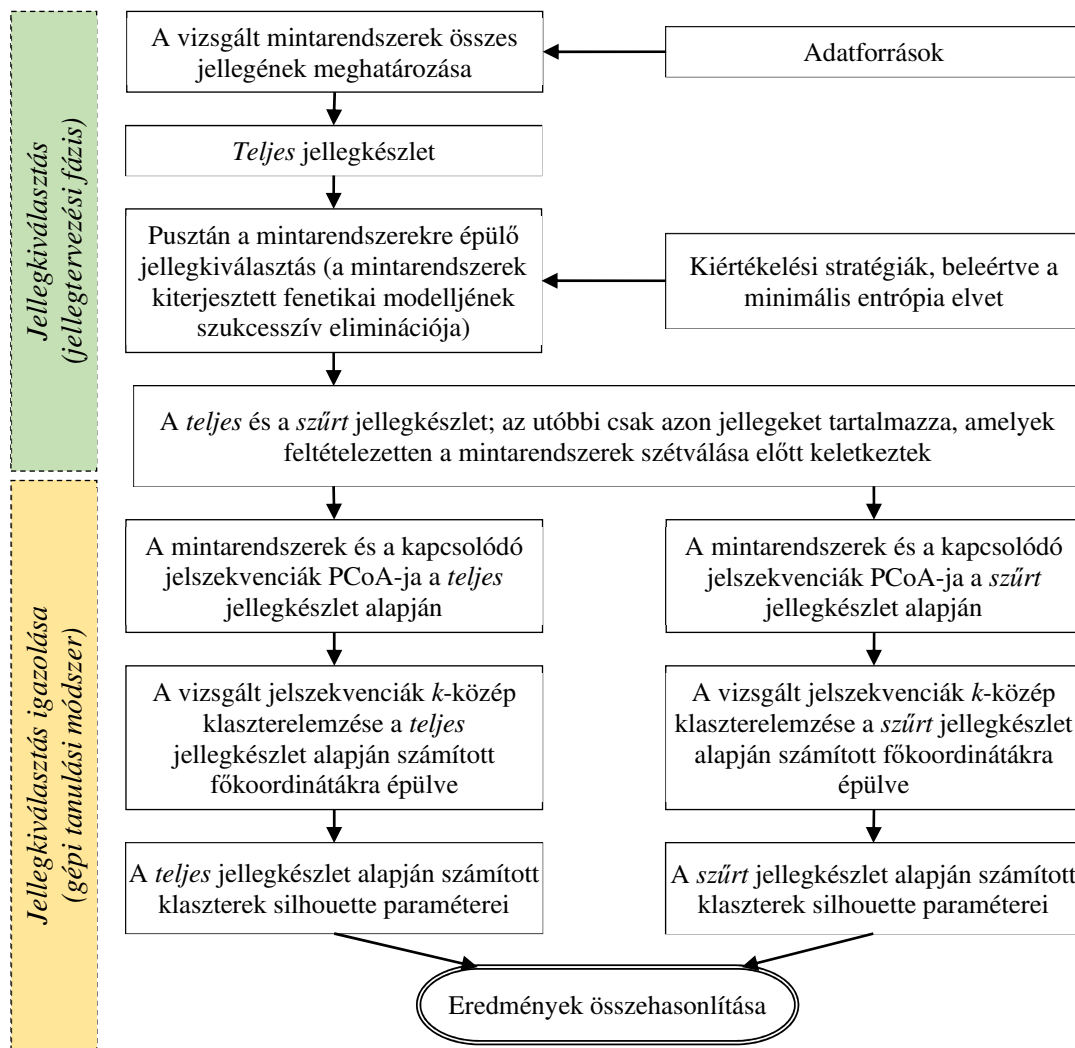


10. ábra: A kiszámított kladogram az ághosszak (jellegállapot-változások száma) feltüntetésével

**3.3. altézis:** Kidolgoztam a jelszekvenciák főkoordináta elemzésén alapuló eljárást annak igazolására, hogy a szukcesszív eliminációból nyerhető szűrt jellegkészlet a vizsgált mintarendszerek fejlődésének kezdetét jellemzi [D7].

A kiterjesztett fenetikai modell kiértékelésével megmutattam a vizsgált mintarendszerek közös elődeinek létezését, és meghatároztam, hogy a teljes, 119 jellegből álló jellegkészletből a vizsgált utódtaxonok evolúciójának kezdeti ökológiai fülkéjében 72 jelleg alakult ki, amely egy, a szukcesszív eliminációval létrehozott, legkorábbinak tekintett szűrt jellegkészlet. Kidolgoztam egy ordinációs elemzésen alapuló eljárást annak igazolására, hogy a szűrt jellegkészlet valóban a vizsgált mintarendszerek fejlődésének korai szakaszára jellemző, lásd 11. ábra.





11. ábra: A jellegkiválasztás ordinációs elemzésen alapuló igazolásának algoritmusai

A kapott eredmények azt mutatják, hogy a vizsgált mintarendszerekkel létrehozott jelszekvenciák elkülönülése a teljes jellegkészlet esetén lényegesen nagyobb, mint a szűrt jellegkészletnél. Ebből az adódik, hogy a szűrt jellegkészlet a taxonok szétválásával járó fejlődés korai szakaszára jellemző.

**4. tézis:** Vizsgálatokat folytattam a rendkívül ritkán fellépő magzati szívzörejek identifikálása és adatainak elosztott adatbázisban történő tárolása és feldolgozása céljából.

**4.1. altézis:** A magzati fonokardiográfia területén bevezettem a magzati szívzörejek kereséséhez alkalmazott zajszint fogalmát a magzati szívhangfelvételek előzetes zajosság szerinti osztályozása érdekében. A zajszinttel a keresett magzati szívzörejek, az S1 hang és a külső zavaró jelek (mint pl. az anyai hangjelek) erős ingadozása miatt az általánosan használt jel/zaj viszonytól eltérően itt a zajt mint adott időszakban mérhető zavaró jelet az előzőekben megtalált S1 hanghoz viszonyítom. Amennyiben az S1-re vonatkozó zajszint egy bizonyos mértéket meghalad, akkor a szívhangfelvétel ezen kritikus időszakában a szívzörejek keresését meg sem indítom [D20].

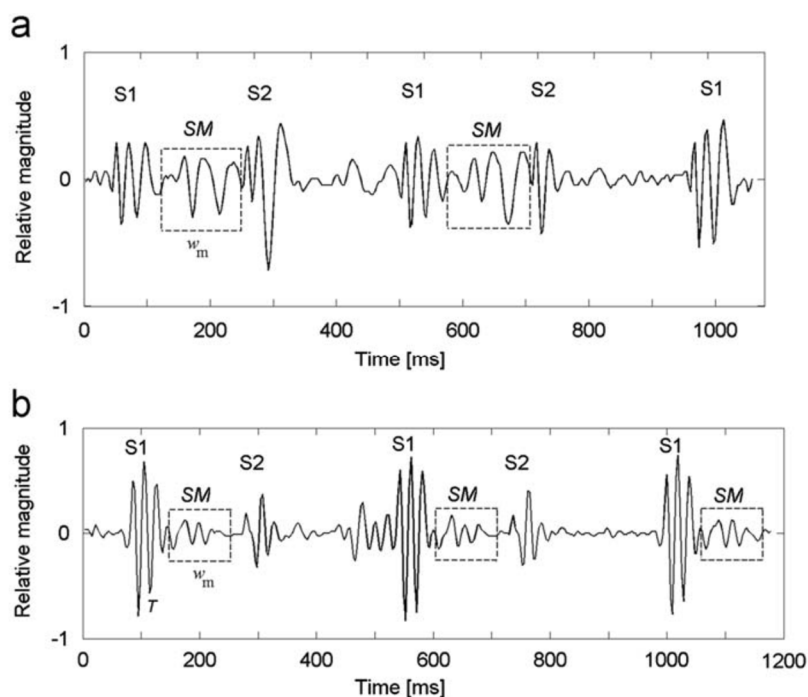
Bevezettem egy jellemzőt az akusztikus magzati szívhangfelvételeknek a jel és a zaj viszonya alapján történő minősítéséhez. Ilyen célokra általában az átlagos zaj- és jelteljesítmény arányából számítható *jel/zaj viszonyt* (SNR) szokták alkalmazni. Az SNR azonban az akusztikus magzati szívhangfelvételek leírására nem kellően pontos, mivel mind a jel, mind a zaj (pl. anyai szívhang beszűrődő része) időszakosan jelenik meg; továbbá esetünkben a zaj nem véletlenszerű, hanem a magzat és az anya különböző mozgásaiból adódó determinisztikus hangösszetevőről. Bevezettem a *zajsint* (NL) fogalmát, amely a zaj átlagos amplitúdójának ( $\hat{A}_{zaj}$ ) és az S1 hang átlagos amplitúdójának ( $\hat{A}_{S1}$ ) aránya:

$$NL = \frac{\hat{A}_{zaj}}{\hat{A}_{S1}} \quad (11)$$

Az NL zajsintben benne lehet az esetleges magzati szívzöreje is, mivel az S1 szisztolés ütőhangok helyének meghatározásakor (a felvételek feldolgozásának első fázisában) a szívzörejeiket nem lehet elkülöníteni. Az NL értéke a keresés egyes időszakaiban módosul az S1 hang előző szakaszban talált értéke alapján. A zajsintet a mérések minősítéséhez használtam: alacsony zajsintnél a hangfelvétel elemeinek azonosítására elegendő az amplitúdókorrelációs módszer [D42], magas zajsintnél viszont erőforrás-igényesebb matematikai eszközök szükségesek, ilyen pl. a wavelet transzformáció [D19].

**4.2. altézis:** *Igazoltam, hogy a hosszú idejű és többszörösen végzett fonokardiográfiai vizsgálatok abból a szempontból is szükségesek, mert ezek adnak lehetőséget arra, hogy a magzat – mivel a mérés alatt a helyét sok esetben változtatja – elhelyezkedése függvényében többféle szívzöreje is detektálható legyen a felvételeken. Felismertem kétféle szívzöreje sajátos együttállását, ezzel igazolva az akusztikus magzati szívhangvizsgálat szelektivitását a szívzörejelek detektálhatósága tekintetében [D22].*

Egy hosszú idejű magzati szívhangvizsgálat szükségszerűségének további indokát tártam fel azon túlmenően, hogy a magzat változó aktivitásából adódóan a rövid idejű mérések eltérő mérési eredményeket szolgáltatnának. A hosszú idejű vizsgálat ugyanis azt a többlet lehetőséget nyújtja, hogy ez idő alatt a magzati mozgásból adódóan olyan hangjelek is érzékelhetővé válnak, amelyek rövid idejű vizsgálatoknál nagy valószínűséggel rejtve maradnának. A 12. ábra ugyanazon hangfelvétel két szelvényét mutatja, amelyek lényegesen különböznek egymástól. A felvétel egy olyan magzathoz tartozik, amelynél a születés után echokardiográfiával kamrai sövényhiányt (VSD) diagnosztizáltak. Ennek alapján megállapítottam, hogy két teljesen különböző formájú és időzítésű szívzöreje látható rajta. Az *a*) típusú szívzöreje a 12. ábra felső részén látható. Ez egy holoszisztolés (a szisztolés szakasz egészében tartó) szívzöreje, amely kismértékben crescendo (fokozatosan erősödő) jellegű, időben csökkenő frekvenciájú és inkább a szisztolés szakasz második felében helyezkedik el. Ezt a szívzörejt a VSD tüneteként lehet értékelni. Az *a*) szívzörejt a mérés első felében rögzítették. Meglepő módon, 10 perc múlva egy másik, *b*) szívzöreje került rögzítésre (12. ábra alsó része), amely helyzetét, valamint frekvenciaspektrumát tekintve jelentősen eltér az elsőtől. A *b*) szívzöreje decrescendo (fokozatosan csökkenő intenzitás) jellegű, inkább a szisztolés szakasz első felében található és időben növekvő frekvenciájú.

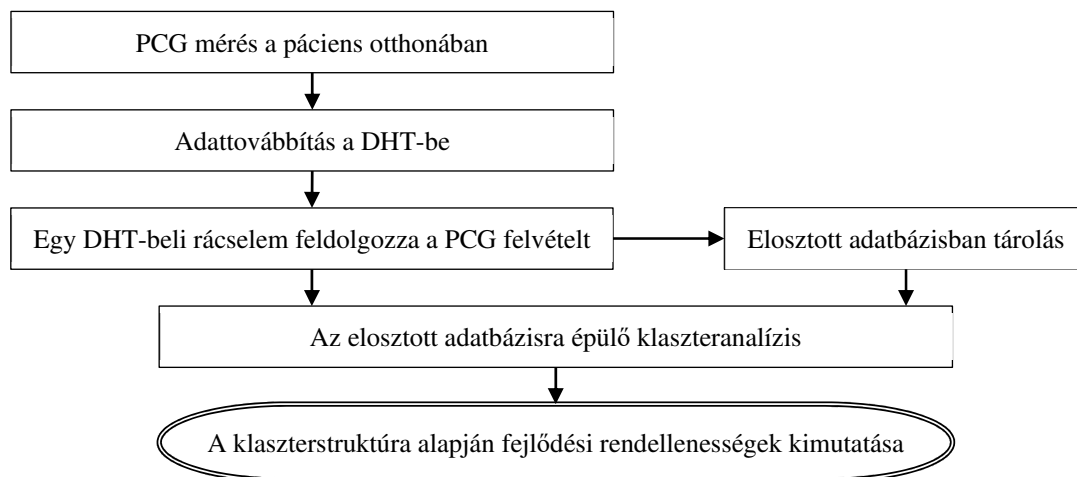


12. ábra: Kétféle magzati szívzöreje [D22]

A kardiológusokkal történt konzultációkon több javaslat született a jelenség magyarázatára. Ezek egyike az, hogy a zöreje oka a magzati légzés ideiglenes megjelenése; azonban az észlelt szívzöreje eltér a magzati légzés kaotikus jelalakjától, változó frekvenciaösszetevőitől és tízszer erősebb intenzitásától. A másik javasolt magyarázat a magzat helyzetének változása, azonban pusztán a jelátviteli közeg megváltozása nem indokolja a jelalak ilyen mérvű módosulását. A kétféle jelenség leginkább elfogadható indoka az lehet, hogy viszonylag távol egymástól két eltérő hangforrás létezik. A billentyűt rögzítő ínhúrok (*chordae tendineae*) rezgése kizárható a VSD-hez való közelsége miatt. Esetleg a hang a Still-zöreje lehet, bár ez a szívzöreje 2 éves kortól szokott fellépni. Összességében az észlelt *b*) típusú szívzöreje a magzati légzéstől vagy mozgástól eltérő fiziológiai, illetve morfológiai okból jött létre, de az érzékelés szempontjából szorosan összefügg a magzat helyzetével.

**4.3. altézis:** Kidolgoztam az egyedi fonokardiográfiai felvételeknek egy elosztott hash táblázatra épülő adatorientált informatikai rendszerbe történő foglалását, amely egy mérőhelyen kapott adatokat az ott azonosított szívzöreje alapján az adott szívzörejmintázatot tartalmazó szívhangfelvételek gyűjtésére kijelölt intézménybe küldi [D48].

Az akusztikus magzati szívhangfelvételekben ritkán fordulnak elő szívzöreje [D18], [D19], ezért ahhoz, hogy a statisztikai biztonságunk megfelelő nagyszámú esetre nézve rendelkezünk adatokkal kidolgoztam egy DHT rendszer tervét, amelyben egy mérőhelyen azonosított szívzöreje adatait az adott szívzörejmintázatot tartalmazó szívhangfelvételek gyűjtésére kijelölt intézménybe küldi. Mivel a magzati szívhang rendellenességeinek típusai (pl. szívzöreje, időbeli eltolódás [*split*]) sokfélék, ezeket eltérő módszerekkel és változó mérési körülmények között mérik, ami megnehezíti az orvosi szempontból fontos következtetések levonását. Ezért javasoltam, hogy a mérési protokoll az egyes felvételekhez metaadatokként automatikusan rendeljen hozzá markereket egy előre rögzített készletből, és a DHT hálózatban az egyes fájlok tárolása a markerek hash értékei alapján történjen. Az így kialakításra kerülő elosztott adatbázis egy veleszületett szívhibát (CHD) kutató szakértői együttműködés része lehet (13. ábra), amely a zöreje-jelenségeket rendkívül ritkán tartalmazó, és ezért az önmagában hatékonyan nem felhasználható felvételeket zörejespecifikusan összegyűjti.



13. ábra: Szívzörejmintázatok elosztott adatbázison alapuló gyűjtése

## Eredmények használhatósága

### *Mintaevolúció informatikai modellezése*

A mintaevolúció területén elért eredmények a jelenleg ismert több millió eddig megfejtetlen jelszekvencia (pl. régészeti tárgyakon talált feliratok) által hordozott információ egy részének felderítéséhez alkalmazhatók, továbbá lehetőséget teremtenek az időbeli fejlődést mutató mintarendszerek evolúciójának az eddigieknél alaposabb, algoritmikus meghatározására. A kidolgozott informatikai eljárásokat az általam vezetett kutatócsoport különböző írásokra mint speciális mintarendszerekre alkalmazta, pl. arab és arám írásokra végeztünk fenetikai és kladisztikai modellezést, valamint megfejtetlen jelszekvenciák jelentésének meghatározását végeztük el.

Munkám eredményességét abban tudom lemérni, hogy az elmúlt másfél évtizedben a Kárpát-medencében előkerült egyes jelszekvenciák megfejtését végző kutatócsoportok tagjaként szerepet kaptam az ezeket alkotó jeleknek korábbról ismert szimbólumoknak való megfeleltetésében. A feladat nehézségét az adja, hogy a különböző, evolúciósan független szimbólumoknak gyakran hasonló jelszekvenciabeli jelek felelnek meg, máskor egy szimbólumot igen eltérő módon jelenítettek meg a különböző jelszekvenciákban. A mintaevolúciós kutatásnak egy jelentős alkalmazási területe adódik abból, hogy az eurázsiai steppe hektikus történelmi és földrajzi viszonyai és az igen összetett mintaevolúciós (írásevolúciós) folyamat együttesen azzal járt, hogy a mintarendszerek sok évszázados fejlődésének csak viszonylag kevés fennmaradt emléke (jelszekvenciák) maradt, és ezek egy részét sem tudták megfejteni. Várható, hogy a jövőben az eddiginél sokkal több fosszília kerül elő, és további megtalált jelszekvenciák megfejtésében komoly segítséget adhat a jelen kutatás, ahogy ezt az irodalomjegyzékben szereplő számos publikáció mutatja.

### *Magzati szívzörejek identifikációja*

A magzati szívhangok vizsgálata körében az általam meghatározott NL zajszint több tízezer magzat vizsgálatánál került felhasználásra, a javasolt P2P alapú elosztott adatbázis pedig új lehetőséget ad a ritka magzati szívzörejek azonosítására és tulajdonságaik megismerésére.

A magzati fonokardiográfiát az orvostudomány még csak korlátozottan alkalmazza, eddig nem nyert teljes elfogadást. Ennek az oka az, hogy az orvosok ismert berendezésekkel dolgoznak, amelyekre megvannak a pontos előírások, így mással dolgozni nehezebb; de a nagy ultrahangos gyártó cégek anyagi érdekeltsége is közrejátszik ebben. Konkrét példákkal is alá tudom támasztani, hogy bizonyos

kritikus esetekben a fonokardiográfiai vizsgálat nemhogy előnyösebb, hanem kizárólag ezzel előzhető meg egy születés előtti végzetes kimenetelű válsághelyzet. Ezért várható, hogy az ultrahang mellett a magzati fonokardiográfia szerves része lesz a magzati vizsgálatoknak, azon helyzetekben, amelyek nehezen vagy pontatlanul oldhatók meg ultrahanggal. Erre jó példa – ugyan ez nem tárgya az értekezésnek – a magzati mozgás, amely hihetetlen sok, eddig feltáratlan információt tartalmaz. A magzati fonokardiográfia fontos kiegészítése lesz a magzati vizsgálatoknak, s ennek is át kell esnie azon a felfutási szakaszon, amelynek a gyógyászatban majd minden készüléknek át kellett esnie. Ebben a tekintetben az, hogy a zajszint egy elkerülhetetlen de nagyon fontos paraméter, rengeteg mérési típus esetén előjön.

## Rövidítések és jelölések

| Fogalom | Leírás   |
|---------|--|
| CBR     | Kárpát-medencei rovás ( <i>Carpathian Basin Rovash</i> ), egy vizsgált mintarendszer               |
| CHD     | Veleszületett szívbetegség (szívhiba) ( <i>Congenital Heart Defect</i> )                           |
| CTG     | Kardiotokográf ( <i>CardioTocoGraph</i> )  |
| DHT     | Elosztott hash táblázat ( <i>Distributed Hash Table</i> )  |
| DM      | Diasztolés szívzöreje ( <i>Diastolic Murmur</i> )  |
| FHR     | Magzati szívverés ( <i>Fetal Heart Rate</i> )  |
| GCP     | Glif-összetettségi mutató ( <i>Glyph Complexity Parameter</i> )                                    |
| HI      | Holofiletikus index ( <i>Holophyletic Index</i> )  |
| HTU     | Feltételezett taxon ( <i>Hypothetical Taxonomic Unit</i> )   |
| JHI     | Együttes holofiletikus index ( <i>Joint Holophyletic Index</i> )                                   |
| NL      | Zajszint ( <i>Noise Level</i> ), a zaj és az S1 hang átlagos amplitúdójának aránya                 |
| OM      | Eredetmodell ( <i>Origin Model</i> )   |
| P2P     | Egyenrangú ( <i>Peer-to-Peer</i> ) kommunikációs modell  |
| PCA     | Főkomponens elemzés ( <i>Principal Component Analysis</i> )  |
| PCG     | Fonokardiográfia ( <i>PhonoCardioGraphy</i> )  |
| PCoA    | Főkoordináta elemzés ( <i>Principal Coordinate Analysis</i> )                                      |
| S1      | Az I. szívhang ( <i>first heart sound</i> ), a szisztolé kezdetét jelző szívütés                   |
| S2      | A II. szívhang ( <i>second heart sound</i> ), a diasztolé kezdetét jelző szívütés                  |
| SFG     | Hasonlósági jellegcsoport ( <i>Similarity Features Group</i> )                                     |
| SHR     | Székely-magyar rovás ( <i>Székely-Hungarian Rovash</i> ), egy vizsgált mintarendszer               |
| SLC     | Szimbólumleszármazási lánc ( <i>Symbol Lineage Chain</i> )   |
| SM      | Szisztolés szívzöreje ( <i>Systolic Murmur</i> ), az I. és II. szívhang között hallható szívzöreje |
| SNR     | Jel/zaj viszony ( <i>Signal-to-Noise Ratio</i> ), az átlagos zaj- és jelteljesítmény aránya        |
| SR      | Steppei rovás ( <i>Steppe Rovash</i> ), egy vizsgált mintarendszer                                 |
| TR      | Türk rovás ( <i>Turkic Rovash</i> ), egy vizsgált mintarendszer                                    |
| UPGMA   | Csoportátlag eljárás ( <i>Unweighted Pair Group Method with Arithmetic mean</i> )                  |
| VSD     | Kamrai sövényhiány ( <i>Ventricular Septal Defect</i> ), a szív fejlődési rendellenessége          |
| WHI     | Súlyozott holofiletikus index ( <i>Weighted Holophyletic Index</i> )                               |
| WJHI    | Súlyozott együttes holofiletikus index ( <i>Weighted Joint Holophyletic Index</i> )                |
| WPGMA   | Egyszerű átlag módszer ( <i>Weighted Pair Group Method with Arithmetic mean</i> )                  |

## Legfontosabb saját közlemények

### Önálló könyv

- [D1] G. Hosszú, *Scriptinformatics. Extended Phenetic Approach to Script Evolution*. Budapest: Nap Kiadó, 2021. Hozzáférhető: [http://napkiado.hu/media/Hosszu\\_Gabor-Scriptinformatics.pdf](http://napkiado.hu/media/Hosszu_Gabor-Scriptinformatics.pdf).
- [D2] G. Hosszú, *Heritage of Scribes*. 3<sup>rd</sup> ed. Earlier editions: 2011, 2012. Budapest: Rovas Foundation, 2013. [Online]. Hozzáférhető: <http://google.hu/books?id=TyK8azCqC34C&pg>.
- [D3] Hosszú G. és Keresztes P., *VHDL alapú rendszertervezés*, Bicske: SZAK Kiadó, 2012.
- [D4] Hosszú G., *Az internetes kommunikáció informatikai alapjai*. Budapest: Novella Kiadó, 2005.
- [D5] Hosszú G., *Internetes médiakommunikáció*, Budapest: LSI Oktatóközpont, 2001.

### Folyóiratcikkek

- [D6] G. Hosszú, “Validation of graph sequence clusters through multivariate analysis: application to Rovash scripts,” *Herit. Sci.* (Springer Nature), vol. 12, no. 1, Apr. 2024, Art. no. 110. Hozzáférhető: <https://rdcu.be/dDBnB>
- [D7] G. Hosszú, “Ordination-based verification of feature selection in pattern evolution research,” *Intell. Data Anal.*, vol. 28, no. 4, 2024, pp. 891–907.
- [D8] G. Hosszú, “Applications of the Pattern Evolution Research,” *Acta Polytechnica Hungarica*, benyújtva.
- [D9] G. Hosszú, “Modelling the Evolution of Scripts,” *SN Comput. Sci.*, benyújtva.
- [D10] O. A. Salman, G. Hosszú, and F. Kovács, “A New Feature Selection Algorithm for Evolutionary Analysis of Aramaic and Arabic Script Variants,” *Int. J. Intell. Eng. Inform.*, vol. 10, no. 4, pp. 313–331, 2022.
- [D11] T. Puskás, and G. Hosszú, “A Cladistic Approach to the Evolution of Steppe Scripts,” *Int. J. Intell. Eng. Inform.*, vol. 10, no. 1, pp. 52–73, 2022.
- [D12] O. A. Salman, and G. Hosszú, “A Phenetic Approach to Selected Variants of Arabic and Aramaic Scripts,” *Int. J. Data Analytics*, vol. 3, no. 1, Article 19, pages 23, 2022.
- [D13] O. A. Salman, and G. Hosszú, “Cladistic Analysis of the Evolution of Some Aramaic and Arabic Script Varieties,” *Int. J. Appl. Evol. Comput.*, vol. 12, no. 4, pp. 18–38, 2021.
- [D14] Zelliger E. és Hosszú G., „A Szarvasi tütartó felirata”, *Magyar Nyelv*, vol. 117, no. 4, pp. 403–430, 2021.
- [D15] L. L. Tóth and G. Hosszú, “A New Topological Method for Examining Historical Inscriptions,” *J. Inf. Technol. Res.*, vol. 12, no. 2, pp. 1–16, 2019.
- [D16] R. E. I. Pardede, L. L. Tóth, G. A. Jeney, F. Kovács, and G. Hosszú, “Four-Layer Grapheme Model for Computational Paleography,” *J. Inf. Technol. Res.*, vol. 9, no. 4, pp. 64–82, 2016.
- [D17] L. L. Tóth, R. E. I. Pardede, G. A. Jeney, F. Kovács, and G. Hosszú, “Preprocessing Algorithm for Deciphering Historical Inscriptions Using String Metric,” *Int. J. Eng. and Technol. Innov.*, vol. 6, no. 3, pp. 202–213, 2016.
- [D18] F. Kovács, K. Kádár, G. Hosszú, Á. Balogh, T. Zsedrovits, N. Kersner, A. Nagy, and G. Jeney, “Screening of Congenital Heart Diseases with Fetal Phonocardiography,” *Int. J. Medical, Health, Biomedical, Bioeng. and Pharmaceutical Eng.*, vol. 9, no. 6, pp. 436–440, 2015.
- [D19] F. Kovács, C. Horváth, Á. T. Balogh, and G. Hosszú, “Fetal phonocardiography – Past and future possibilities,” *Comput. Meth. Programs in Biomed.* (Elsevier), vol. 104, no. 1, pp. 19–25, Oct. 2011.
- [D20] F. Kovács, C. Horváth, Á. T. Balogh, and G. Hosszú, “Extended Noninvasive Fetal Monitoring by Detailed Analysis of Data Measured With Phonocardiography,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 64–70, Jan. 2011.
- [D21] F. Kovács, M. Török, C. Horváth, Á. T. Balogh, T. Zsedrovits, and G. Hosszú, “A new, phonocardiography-based telemetric fetal home monitoring system,” *Telemedicine and e-Health*, vol. 16, no. 8, pp. 878–882, 2010.
- [D22] F. Kovács, N. Kersner, K. Kádár, and G. Hosszú, “Computer method for perinatal screening of cardiac murmur using fetal phonocardiography,” *Computers in Biology and Medicine* (Elsevier), vol. 39, pp. 1130–1136, 2009.
- [D23] Kozma R., Horváth Cs., Kovács F. és Hosszú G., „Akusztikus magzati szívhang feldolgozást végző céláramkör”, *Informatika*, vol. 6, no. 2, pp. 53–59, 2003.

- [D24] G. Richly, R. Kozma, F. Kovács, and G. Hosszú, “Optimised soundprint selection for identification in audio streams,” *IEE Proceedings-Communications*, vol. 148, no. 5, pp. 287–289, 2001.

### Konferenciaticikk

- [D25] O. A. Salman and G. Hosszú, “Using distance-based methods to calculate optimal and suboptimal parsimony trees,” in *WAIT 2024. Proc. Workshop on the Advances of Information Technology*, Dept. of Control Engineering and Information Technology (IIT) BME, 2024, pp. 79–84.
- [D26] O. A. Salman and G. Hosszú, “Phylogenetic modelling scripts for identifying script versions,” in *CENTERIS – International Conference on ENTERprise Information Systems, Porto, Portugal, 8–10 November 2023, Procedia Comput. Sci.* (Elsevier), vol. 239, 2024, pp. 1417–1424.
- [D27] G. Hosszú, “The Pattern Evolution Research,” in *Proc. 14th IEEE Int. Conf. on Cognitive Infocommunications*, IEEE, 2023, pp. 167–172.
- [D28] O. A. Salman and G. Hosszú, “Phylogenetic Inference Using Advanced Feature Selection,” in *Proc. 14th IEEE Int. Conf. on Cognitive Infocommunications*, IEEE, 2023, pp. 173–178.
- [D29] O. A. Salman and G. Hosszú, “Optimised feature dimension reduction method and its impact on the search for optimal trees,” in *WAIT 2023. Proc. Workshop on the Advances of Information Technology*, Dept. of Control Engineering and Information Technology (IIT) BME, 2023, pp. 23–28.
- [D30] G. Hosszú, “Data-Driven Phenetic Modeling of Scripts’ Evolution,” in *LISS 2020. Proc. 10th Int. Conf. Logistics, Informatics and Service Sciences*, Springer, 2021, pp. 389–403.
- [D31] L. L. Tóth, F. Kovács, and G. Hosszú, “Deciphering Historical Inscriptions Using Machine Learning Methods,” in *LISS 2020. Proc. 10th Int. Conf. Logistics, Informatics and Service Sciences*, Springer, 2021, pp. 419–435.
- [D32] F. Kovács, M. Á. Goda, G. Hosszú, and T. Telek, “A Proposed Phonography-Based Measurement of Fetal Breathing Movement Using Segmented Structures with Frequency Splitting,” in *EMBC 2020. Proc. 42<sup>nd</sup> Annu. Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, 2020, pp. 4483–4486.
- [D33] G. Hosszú and F. Kovács, “Topological Analysis of Ancient Glyphs,” in *Proc. 2016 IEEE Int. Conf. Systems, Man, and Cybernetics (SMC 2016)*, 2016, pp. 2248–2253.
- [D34] Hosszú G., „Topológiai eltérések minimalizálására visszavezetett graféma leszármazási vizsgálatok”, in *Az informatika korszerű technikái*, 2014, pp. 60–71.
- [D35] F. Kovács, G. Fodor, and G. Hosszú, “A Proposed Novel Method for CHD Screening by Fetal Heart Murmur Detection using Phonocardiography,” in *Conf. EFMI STC Data and Knowledge for Medical Decision Support, Prague*, IOS Press BV, 2013, pp.160–164. (*Stud. Health Technol. Inform.* 2013; 186: 160–164).
- [D36] G. Fodor, G. Hosszú, and F. Kovács, “Model Improvement and Cluster Analysis of the Fetal First Heart Sounds,” in *5<sup>th</sup> European Conf. of the Int. Federation for Medical and Biological Eng., Budapest, 2011*, Springer, IFMBE Proceedings, vol. 37, 2012, pp. 279–282.
- [D37] G. Fodor, Á. T. Balogh, G. Hosszú, and F. Kovács, “Screening for Congenital Heart Diseases by Murmurs Using Telemedical Phonocardiography,” in *34<sup>th</sup> Annu. Int. Conf. IEEE Eng. in Medicine and Biology Soc. (EMBC, 2012, San Diego)*, 2012, pp. 6100-6103.
- [D38] Hosszú G., „A grafémák leszármazási elemzése”, in *I. Alk. Számítógépes Paleográfiai Konf.*, Hosszú G., szerk., BME VIK, 2012, pp. 18–36.
- [D39] Tóth L. L.; R. Pardede és Hosszú G., „Grafémák kanonikus összetevőkre bontása”, in *I. Alk. Számítógépes Paleográfiai Konf.*, Hosszú G., szerk., BME VIK, 2012, pp. 112–126.
- [D40] E. Kósa, C. Horváth, N. Kersner, K. Kádár, F. Kovács, M. Török, and G. Hosszú, “Experiences with Fetal Phonocardiographic Telemonitoring and Future Possibilities,” in *30<sup>th</sup> Annu. Int. IEEE EMBS Conf., 2008, Vancouver*, pp. 5859–5862.
- [D41] F. Kovács, C. Horváth, M. Török, and G. Hosszú, “Fetal Breathing Transmission in Phonocardiographic Monitoring Telemedicine Systems,” in *Proc. 28th Annu. Int. IEEE EMBS Conf., 2006, New York*, pp. 5226–5229.
- [D42] F. Kovács, C. Horváth, M. Török, and G. Hosszú, “Long-Term Phonocardiographic Fetal Home Monitoring for Telemedicine Systems,” in *27th Annu. Int. IEEE EMBS Conf., 2005, Shanghai*, pp. 3946–3949.

- [D43] G. Richly, R. Kozma, G. Hosszú, and F. Kovács, “A Proposed Method for Improved Sound-Print Selection for Identification Purposes,” in *Proc. 5th WSES/IEEE CSCC, 2001, Rethymnon, Greece*, pp. 4961–4964.

### Könyvfejezet

- [D44] G. Hosszú, “Phenetic Approach to Script Evolution,” in *Kodikologie und Paläographie im Digitalen Zeitalter 4. Codicology and Palaeography in the Digital Age 4*, H. Busch, F. Fischer, and P. Sahle, Eds., Norderstedt: Books on Demand, 2017, pp. 179–252.
- [D45] L. L. Tóth, R. E. I. Pardede, G. A. Jeney, F. Kovács, and G. Hosszú, “Application of the Cluster Analysis in Computational Paleography,” in *Handbook of Research on Advanced Computational Techniques for Simulation-Based Engineering*, P. Samui, Ed., Hershey: Engineering Science Reference, 2016, pp. 525–543.
- [D46] G. Hosszú, “A Novel Computerized Paleographical Method for Determining the Evolution of Graphemes,” in *Encyclopedia of Information Science and Technology*, M. Khosrow-Pour, Ed., 3<sup>rd</sup> ed., Hershey: Information Science Reference, 2015, ch. 194, pp. 2017–2031.
- [D47] G. Hosszú, “Mathematical Statistical Examinations on Script Relics,” in *Data Mining and Analysis in the Engineering Field*, V. Bhatnagar, Ed., Hershey: Information Science Reference, 2014, ch. 8, pp. 142–158.
- [D48] C. Horváth, G. Fodor, F. Kovács, and G. Hosszú, “A Proposed Scalable Environment for Medical Data Processing and Evaluation,” in *Handbook of Research on Developments in e-Health and Telemedicine: Technological and Social Perspectives*, M. M. Cunha, A. Tavares, and R. Simoes, Eds., Hershey: Medical Information Science Reference, 2010, pp. 603–613.

### Vezetéssel sikerrel megvédett PhD értekezések

- Varga L., „*Kisfogyasztású áramkörök tervezési módszertanának kutatása*”, BME, Budapest, 2010.
- Czirkos Z., „*Számítógépek P2P együttműködésen alapuló védelme*”, BME, Budapest, 2012.
- Horváth P., „*VHDL alapú áramkörszintézis*”, BME, Budapest, 2016.

### Hivatkozott nem saját irodalom

- [1] C. Skelton, “Methods of using phylogenetic systematics to reconstruct the history of the Linear B script,” *Archaeometry*, vol. 50, pp. 158–176, 2008.
- [2] A. Gutkin and R. Sproat, “Areal and Phylogenetic Features for Multilingual Speech Synthesis,” *Proc. of Interspeech 2017, ISCA, August 20–24, 2017, Stockholm*, 2017, pp. 2078–2082.
- [3] G. Prószéky, “Machine Translation and the Rule-to-Rule Hypothesis,” in K. Károly and Á. Fóris, Eds., *New Trends in Translation Studies (In Honour of Kinga Klaudy)*, Budapest: Akadémiai, 2005, pp. 207–218.
- [4] G. Németh, “Gaps to Bridge in Speech Technology,” *Lecture Notes in Computer Science*, vol. 8773, pp. 15–23, 2014.
- [5] G. Sallai, “The Cradle of the Cognitive Infocommunications,” *Acta Polytechnica Hungarica*, vol. 9 no. 1, pp. 171–181, 2012.
- [6] F. Kovács, M. Török, and I. Hábermajer, “A rule-based phonocardiographic model for long-term fetal heart rate monitoring,” *IEEE Trans. Biomed. Eng.*, vol. 47, no. 1, pp. 225–230, 2000.