

Bírálat
dr. Németh Géza „Célorientált gépi beszédkeltés
interakciós rendszerekben”
című MTA doktori értekezéséről

dr. Németh Géza dolgozatának témája elsősorban a magyar nyelvű beszéd-szintézis (beszédgenerálás). A téma időszerűségéhez nem férhet kétség, emellett a dolgozat a kifejlesztett eljárások számos gyakorlati alkalmazását is ismerteti. Külön pozitívum, hogy a dolgozat a magyar nyelvre fókuszál (márpedig magyar kutatókon kívül senki nem fog magyar nyelvű beszéd-szintézist készíteni), így kutatásainak van egyfajta „nemzeti” karaktere is (politikai felhangoktól mentes értelemben).

A dolgozat struktúrája számomra átgondoltnak tűnik, az alapvetően könnyen olvasható, követhető. A felölelt időszak már csak a dolgozat jellege (MTA doktori értekezés) miatt is elég nagy (az I.1. tézispont egyik alátámasztó cikke még 2000-ben jelent meg), ami még körülményesebbé tehetné a bemutatott technikák követését, de fontosnak tartom a pályázó azon gondolatát (4.o.), hogy a technológiák alapvetően nem ugyanazon alapötlet tovább- és továbbfejlesztéseként születtek meg, hanem a hardver- és szoftverváltozásokat lekövetve, sokszor alapvetően más elven nyugszanak. (Ennek ellenére a dolgozat követhető, és sikerült egységesen fölfűzni az egyes technikákat.) Az időtávnak tudható be az is, hogy a bemutatott megoldások nem azonos tesztelési körülmények között, és nem is azonos adatbázisokon készültek el, lettek kipróbálva (mely tovább növelte volna a dolgozat olvashatóságát), így sajnos a legtöbb esetben nem is hasonlíthatóak össze direktben. Ahol viszont mégis, ott a pályázó ezt meg is teszi (pl. 11. ábra (32. oldal), 12. ábra (33. oldal)).

Ha már adatbázis: a beszéd-szintézisnek, a beszédtechnológia többi ágához hasonlóan, speciális adatbázisokra van szüksége; ráadásul ezek a korpuszok (az elmúlt néhány év néhány „nagyon mély” tanulási eljárásától részben eltekintve) erősen nyelvfüggőek: nem várható, hogy érthető és természetes magyar nyelvű beszédhang álljon elő angol (német, orosz, litván stb.) hangfelvételek részleteinek összefűzésével. Amellett, hogy egy adatbázis sokszor a kutatómunka előfeltétele, elkészítése szerintem (bár valóban speciális, de mégis) tudományos teljesítmény is, melynek megtervezéséhez, elkészítéséhez speciális szakértelemre és nagyfokú precizitásra van szükség. Bár Németh Géza dolgozata nem emelte ki az adatbázisokat mint saját tudományos eredményeit (vagy legalábbis nem külön tézispontokban), én azért szeretném ezt megtenni (még ha az elkészült korpuszok nem is feltétlenül nyilvánosan elérhetőek).

Az első három téziscsoport (5-7. fejezetek) tárgyalja a pályázó azon eredményeit, melyek a beszédgenerálás metodikáját érintik. Az eljárások esetében jól követhető, hogyan bővültek a hardverlehetőségek, melyek egyre összetettebb, egyúttal egyre absztraktabb algoritmusokat tettek lehetővé. (Kicsit hiányolom a legújabb trend, a mélytanulás-alapú rendszerek eredményeit, de ezek a 8. oldal tanúsága szerint szándékosan maradtak ki a dolgozatból.) Különösen érdekes, hogyan jöttek vissza a már meghaladottnak tűnő, „egyszerűbb” rendszerek a mobiltelefonos platformokon. A tézispontok jól átgondoltak, jól kifejtettek, valamint a kísérleti igazolásaik is megfelelőek (az alátámasztó publikációkat külön érintem).

A IV. téziscsoport témája (multimodális beszédinformációs rendszerek) eltér a korábbiaktól, ezek a tézispontok inkább a rendszerek felhasználói felületével kapcsolatosak. Ezeket a téziseket gyengébbnek érzem, mint az első három téziscsoportban vázoltakat – ugyanakkor ez a terület igen messze esik a saját szakterületemtől, így ezen pontok relevanciáját, újdonságértékét, vagy akár az alátámasztó publikációk konferenciáinak átlagos szintjét nehezebben is tudom megítélni.

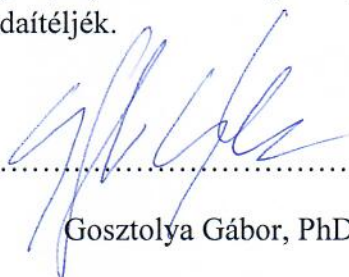
A 9. fejezet az eredmények gyakorlati alkalmazásait, a pályázóhoz kapcsolódó (megkockáztatom: fontosabb) műszaki alkotásokat mutatja be. Ezt a részt személy szerint szintén nem érzem annyira erősnek, mint a korábbi fejezeteket, illetve némelyik bemutatott megoldás (pl. a NAO robot kórházi alkalmazása) nem is igazán kapcsolódik a korábban bemutatott tudományos eredményekhez (azaz a beszédgeneráláshoz). (Ráadásul fölkérésem is a pályázó tudományos teljesítményének értékelésére vonatkozik, így nem tudom, a véleményem ezekről a fejezetekről mennyire releváns.) Azonban szerintem azt is érdemes figyelembe venni, hogy a beszédgenerálás (és általában a beszédtechnológia) egy *mérnöki* tudomány, ahol az alkalmazások nem valamiféle luxust vagy technikai érdekességet jelentenek, hanem ezek adják az egész terület létjogosultságát. Ebből persze nem következik, hogy feltétlenül ugyanannak a csoportnak kell élen járnia a konkrét alkalmazások kifejlesztésében, mint amelyik a tudományos (alap)kutatást végzi, hiszen egy informatikai alkalmazásnak sok olyan tulajdonsággal is rendelkeznie kell (megbízhatóság, a kód karbantarthatósága, alacsony futási idő és memóriaigény), amely egy eljárás tudományos publikálása szempontjából irreleváns vagy legalábbis másodrendű. Ugyanakkor a gyakorlati alkalmazások (és a MÁV hangosbemondója egy elég komoly gyakorlati alkalmazás) véleményem szerint növelik a korábbi fejezetekben ismertetett tudományos eredmények értékét, mert megmutatják, hogy azok éles körülmények között is hasznosíthatóak. (Ráadásul az egyes konkrét gyakorlati esetekben az alkalmazott eljárással szemben támasztott követelmények fölismerése és ezek alapján a megfelelő eljárás kiválasztása is tanúskodik a pályázó tudományos kvalitásairól.)

A pályázó tudományos publikációit illetően, az I-III. téziscsoportot alátámasztó cikkek alapvetően hozzájárulnak, vagy túl is teljesítik az adott időszak (2000-2016) magyarországi kutatóinak teljesítményét; bár azért a rangos nemzetközi konferenciák (pl. Interspeech [58, 66]) mellett szerepelnek magyar nyelvű (pl. MSZNY) vagy szervezésű (pl. CogInfoCom) konferenciák, valamint magyar folyóiratok (Híradástechnika, illetve Infocommunications Journal) is. Az utóbbi években elvárás lett vezető nemzetközi szakmai folyóiratokban is publikálni, ezért is volt meglepő számomra, hogy bár a pályázónak vannak ilyen lapokban megjelent cikkei (pl. IEEE Journal on Selected Topics in Signal Processing, Computer Speech & Language), a tézispontok között csak egyetlen ilyen alátámasztó hivatkozás szerepel (Speech Communication [86]). Ennek oka lehet persze az is, hogy az ilyen cikkek főleg a dolgozat benyújtása (a borító tanúsága szerint 2019) után lettek elfogadva. A pályázó (legalábbis az Irodalomjegyzékben) szerkesztőként jegyzi A magyar beszéd c. könyvet, azonban ez több tézispont (I.1., I.2., II.1., II.2.) esetén is alátámasztó publikációként van jelölve – itt véleményem szerint érdemes lett volna a konkrét vonatkozó fejezeteket föltüntetni. Összességében a pályázó publikációs teljesítményét elegendőnek tartom az MTA doktora címhez.

Formai megjegyzéseim: a megfogalmazások kellően pontosak (bár a magyar terminológia ezt nem mindig teszi egyszerűvé). Sajnos elég sok az egybe-különírási hiba. Az ábrák is esztétikusak és informatívak, bár a feliratok időnként angolul szerepelnek (pl. 17. ábra (40. oldal), 24. ábra (50. oldal), 37. és 38. ábrák (69. oldal)), ami egy magyar nyelven írt dolgozat esetén nem a legszerencsésebb. A tézispontok összefoglalása (10. fejezet) jó ötlet, viszont hiányoltam az alátámasztó publikációk felsorolását, melyek pedig a tézispontok eredeti leírásainál szerepeltek. Az irodalomjegyzékben előfordulnak hiányos / túlrövidített szerzőnevek (pl. [40]-nél „O. W. a. S. King” valószínűleg két szerző lenne (én Simon Kingre tippelek), [97]-nél „J.S. J. T. és O. E.”, [102]-ben „G.G.”, [114]-ben „G. T. B. Sárosi”, [112]-ben pedig „D. G. f. R. P. European Commission” gyanús. A [45]-ös számú cikk megjelenési helye is elég szokatlan.

Összefoglalva: véleményem szerint a tézispontok jól átgondoltak, jól megfogalmazottak, kellő mennyiségű tudományos újdonságot tartalmaznak, saját cikkekkel megfelelően alátámasztottak. A pályázó publikációs teljesítménye túlteljesíti az időszak hazai normáját, véleményem szerint meghaladja az MTA doktora címhez szükséges mennyiséget. A dolgozat jól követhető, jól olvasható, bár kisebb formai hibák akadnak benne. Külön szeretném kihangsúlyozni, hogy a pályázó nem csak a szorosan vett tudományos kutatási eredményeit foglalta össze, hanem bizonyosságot tett arról, hogy az az „előtti” és „utáni” lépésekben is jártas: aktívan részt vett a kutatás előfeltételül szolgáló akusztikus adatbázisok megtervezésében és rögzítésében, valamint a kutatási eredmények működő gyakorlati alkalmazásainak kifejlesztésében is. Mindezek fényében úgy vélem, hogy a pályázó tudományos teljesítménye megfelelő ahhoz, hogy részére az MTA doktora címet odaítéljék.

Szeged, 2021. április 27.



.....

Gosztolya Gábor, PhD
tudományos főmunkatárs

ELKH MTA-SZTE

Mesterséges Intelligencia Kutatócsoport