

Válasz Dr. Levendovszky János professzor úr bírálatára

Köszönöm Dr. Levendovszky János professzor úr alapos bírálatát, javaslatait és támogató összegzését. Válaszaimat a bírálat szerkezetéhez illeszkedve fogalmazom meg. A bíráló megállapításait *félkövér betűtípussal* idézem meg. A saját válaszaimat pedig normál betűtípussal adom meg.

1. Általános megállapítások a dolgozat témaválasztásával és struktúrájával kapcsolatban

1.1 „ A szerzőnek a PhD fokozata megszerzése óta a tématerületen végzett kutatásait és ezek eredményét foglalja össze. ... A fókuszban hangsúlyosan a magyar nyelvre vonatkozó fejlesztések állnak, elsősorban látáskorlátozottak számára elérhető alkalmazások kidolgozásával.”

A disszertáció témájának meghatározásánál (2015 körül) az MTA doktori szabályzata figyelembe vételével természetes volt, hogy a PhD fokozat megszerzése (1997) óta elért eredményeket törekedtem számba venni. Ez alatt az idő alatt egyrészt a gépi beszédkeltés három technológiai generációjának (diád/triád, korpusz, HMM) kifejlesztése és ipari bevezetése valósult meg, másrészt a 2010-es évek elején indult meg a mély neurális hálózatokon (Deep Neural Network, DNN) alapuló megoldások robbanásszerű elterjedése. A disszertáció terjedelmi korlátainak figyelembevételével azt a döntést hoztam, hogy a már megállapodottnak tekinthető technológiákat tárgyalom, mivel a DNN témakörben néhány hónap alatt is alapvető változások történhetnek, így a disszertáció megírásától a védésig terjedő időszakban az erről szóló tanulmányok meghaladottá válhatnak. A DNN témakörben a rövidebb átfutási idejű konferencia és folyóirat cikkekben publikáltam eredményeimet.

Az utóbbi években kezdtek el a nemzetközi nagyvállalatok magyar nyelvű általános célú gépi beszédkeltési megoldásokat elérhetővé tenni (pl. Google, Microsoft), ezért saját kutatásaink és az ezek alapján előálló magyar nyelvű technológiák sokáig hiánypótlóak voltak. A célorientált megoldások pedig a mai napig csak a mi megoldásainkkal valósíthatók meg versenyképes minőségben.

A látássérültek természetszerűleg minden nyelven az első felhasználók közé tartoznak, hiszen a gépi beszédkeltés teszi lehetővé fogyatékoságuk részbeni kompenzálását. Mivel hallásukra kell hagyatkozniuk, ezért tesztelőként, az esetleges hibák pontos jelölőjeként is kutatási partnereinknek tekinthetők. A gépi beszédkeltés alkalmazási lehetőségei azonban lényegesen szélesebb területet ölelnek fel. Ezt jelzi, hogy a disszertációban említett 13 alkalmazás közül csak 4 kötődik a látássérült emberekhez.

1.2 „A fejlesztési eredmények impresszívek, azonban ezeket sok esetben nem kíséri semmiféle analitikus tárgyalás. Ez egyrészt a témakör sajátossága is, azonban ezen tárgyalás hiánya olyan esetekben is előfordul, ahol lehetséges lett volna egzakt megközelítéseket ismertetni és azokon demonstrálni a módszerek működését. Az értekezés szinte kizárólag szöveg alapú, némely esetben elnagyolt és nemönmagyarázó, az eredmények feldolgozásához a háttérirodalom tanulmányozására is szükség volt.”

Bírálom jogosan mutat rá az impresszív eredmények mellett az analitikus tárgyalás hiányára. Ennek oka a kiinduló témaválasztásban keresendő. A mintegy 20 év kutatási eredményeinek (4 téziscsoport, 11 altézis) tárgyalása 18 alátámasztó publikáció alapján történik 47 oldalon. Az eredmények alkalmazásait és a kapcsolódó műszaki alkotásokat ismertető 9. fejezet 27 oldal terjedelmű. Az adott korlátok között törekedtem közérthető, tömör fogalmazásra, a részleteket ábrák és táblázatok segítségével kifejtetni. A tématerülettől távolabb álló szakértőknek valóban szükséges lehet a hivatkozott irodalom tanulmányozása. Erre a legalkalmasabb A magyar beszéd című könyv [1], melyet szerzőként és szerkesztőként is jegyeztek.

1.3 „Ugyanakkor az értekezésben főleg a már bevezetett módszerek magyar nyelvre történő adaptálása kapott részletezést. Ez nagy munkát és innovatív megközelítéseket igényelt, amelyek fontos – az életminőséget is javító - alkalmazásokkal bírnak. Azonban a munka jellege inkább fejlesztési természetű. Nem csoda, hogy a dolgozat majdnem fele azalkalmazásokat és ezek technikai paramétereinek a leírását tartalmazza.”

Köszönöm, hogy bírálóm elismeri kutatásaink innovatív jellegét és fontosságát. Sajnos, mint oly sok más területen, a beszédtechnológiában sem Magyarország a technológiai trendek meghatározója. A rendelkezésünkre álló erőforrások mellett a disszertáció címében is említett „cél-orientált” megközelítéssel konkrét funkciók megvalósítását próbáltam elérni (jellemzően Magyarországon, de időnként a világon először). Ennek kapcsán jellemzően egyszerre merülnek fel, összefonódó kutatási (pl. adatbázis elméleti megtervezése és a létrehozása, magyar nyelv-specifikus algoritmusok elméleti kutatása és megvalósítása) és fejlesztési kérdések. A teljes disszertáció terjedelmének közel a felét teszi ki a kutatási eredményeket összefoglaló tézisek tárgyalása. Ezekre 70 (nagyreszt külföldi) független hivatkozás érkezett. A kutatási eredményeket felhasználó alkalmazások / mérnöki alkotások pedig mintegy 27%-ot foglalnak el.

2. A dolgozat tartalma és struktúrája

2.1 „A „Kutatási célkitűzések” című harmadik fejezet, több a munka minőségét értékelő állításokat tartalmaz, pl. a Jaws for Windows esetén egy amerikai fejlesztő elismerő e-mail-jére utal, nem ártott volna ezt az e-mailt a függelékben csatolni az illető affiliációjával. Hasonlóan, mikor a szerző arról nyilatkozik, hogy a II: téziscsoport eredményeihez kapcsolódó előadásuk „felkeltette a hasonló témán francia nyelven dolgozó kutatók figyelmét és érdeklődtek a részletek iránt” szintén dokumentálni. Ugyanakkor fontos elismerését jelenti a munkának a két támogatott H2020-as projekt (PAELIFE és VUK AAL). A rövidítések mellett itt a címük és a konzorcium megadása is informatív lett volna.”

Köszönöm, hogy bírálóm felhívta a figyelmemet arra, hogy minden értékelő jellegű állítást pontosan dokumentálnom kell. Ezeket az állításokat csak illusztrációnak szántam. Szerencsére Szuhaj Mihály, az Informatika a Látássérültekért Alapítvány vezetője a 2000-es évek elejétől megőrizte a releváns levelezést a Freedom Scientific cég Jaws for Windows képernyőolvasó alkalmazásnak és a ProfiVox rendszerünknek az integrációjával kapcsolatban és volt szíves a vonatkozó levelezést rendelkezésemre bocsátani. A jelen válasz mellékletében a pontos neveket és e-leveél címeket adatvédelmi okokból anonimizáltam, de szükség esetén az eredeti levelek is megtekinthetők. A releváns állítások a következők, melyeket a Freedom Scientific cég munkatársai tettek 2005-ben:

1. „This Profivox synthesizer is a true winner--it is absolutely fantastic! It is fast, clear and responsive, and a true pleasure to listen to.”

2. „I am told by Stephen that Profivox is the best speech synth he has heard so far.”

A III. téziscsoport eredményeivel kapcsolatban az Acapela Group (<https://www.acapela-group.com/>) munkatársai érdeklődtek a 2015-ös Interspeech konferencián, Drezdában. A PAELIFE és a VUK projekt kódját karakterszám kímélés céljából csak a köszönetnyilvánításnál adtam meg. A teljesség kedvéért az alábbiakban a releváns adatok elérhetők:

PAELIFE: Personal Assistant to Enhance the Social Life of the Seniors,
NKFIH szerződés szám: AAL_08-1-2011-0001 (<http://www.aal-europe.eu/projects/paelife/>)

VUK: Visionless sUpporting frameworK
NKFIH szerződés szám: AAL-2014-1-183 (<http://www.aal-europe.eu/projects/vuk/>)

3. Részletes megjegyzések és kérdések a dolgozat állításaival kapcsolatosan

3.1 5. Fejezet 16. old. „A rendszerterv és a korpuszkészítés nagyon fontos feladat, de nem triviális, hogy a blokkvázlaton látható rendszerterv miért tudományos eredmény. Mennyire kellett ennek megalkotásához tudományos eszközöket használni és mennyiben nagy az újdonságtartalma.

A szubjektív értékelésnek ennél az eredménynél nincsenek megadva a paraméterei (hány ember hallgatta, milyen kor/nem eloszlásban, mik a statisztikai eredmények)”

Köszönöm bírálómnak, hogy felhívta a figyelmemet a pontosabb fogalmazásra. A 6. ábra a rendszertervnek egy egyszerűsített változata. A részletes megoldás létrehozásához egyrészt irodalomkutatást kellett végezni arról, hogy milyen elvek alapján, milyen szempontok szerint alakítottak ki hasonló technológiákat más nyelvekre (pl. [6] [7] [46] [47]), másrészt ki kellett dolgoznom a magyar nyelvre alkalmazandó megoldásokat. Kísérleteket kellett végezni arra vonatkozóan, hogy mikor lehet megfelelő megoldással tömöríteni és mikor, milyen feltételek mellett kell a jó minőség érdekében nagyobb adatmennyiséget választani. A tömörítésre példa az, hogy a magyar hosszú mássalhangzókat megfelelő minőségben elő lehet állítani a rövid változatokból, míg a magánhangzók ilyen kezelése már túlzott minőségromlást okoz. Ezek új tudományos eredmények és eredetiek.

A szubjektív értékeléseket mindig az adott időszak elvárásai szerint végeztük. A ProfiVox diád technológia első alkalmazása a Westel MAILMONDÓ e-leveél szolgáltatás volt, ahol az összehasonlító értékelést a cég végezte és a részletes tesztelési adatokat sajnos nem kaptuk meg. Viszont a későbbi kutatások során a rendszert referenciaként használva a szubjektív értékelés részletes adatait is publikáltuk (II. altézis, 11. ábra és II.3 altézis, 18. ábra).

3.2 „20. oldal A rendszert a MAILMONDÓ szolgáltatás [48] és [49] fejlesztése és alkalmazása során széles körben teszteltük és megállapítottuk, hogy jobb minőséget nyújt, mint a korábbi magyar nyelvű gépi szövegfelolvasó megoldások (ld. 11. ábra, 32.o). A német nyelvű változatot kutatási együttműködés keretében a TU Kaiserslautern és a Fraunhofer IESE anyanyelvű munkatársaival validáltuk [50]. Ennek az állításnak validációja nehéz, a jelzett cikk, illetve MATÁV 2000-ben kiadott jelentése (21 éves) nem letölthető.”

A hivatkozott cikk [50] könyvfejezetként jelent meg és az alábbi hivatkozáson elektronikus formában a Springer honlapjáról letölthető/megvásárolható:

https://link.springer.com/chapter/10.1007/978-3-540-76729-9_24

A Matáv 2000-ben kiadott, az 1999-re vonatkozó éves jelentése az alábbi hivatkozáson érhető el: https://www.telekom.hu/static-tr/sw/file/1999_eves_jelentes.pdf

A Mailmondó szolgáltatásra vonatkozó bekezdésrészlet az alábbi:

„Bővülő GSM-szolgáltatáskínálat

...Mailmondó néven bemutatkozott a hangos e-mail, valamint a jövőt idéző, kísérleti WAP (Wireless Application Protocol) szolgáltatás, a mobilkommunikáció és az IP-technológia összekapcsolódása.”

A 2004-ben kiadott, a 2003-ra vonatkozó jelentés az alábbi hivatkozáson érhető el:

<https://www.telekom.hu/static-tr/sw/file/EvesJelentes2003.pdf>

Az SMSmondó szolgáltatásra vonatkozó mondatok a következők:

„Hazai fejlesztés eredményeként indította el egyedülálló SMSmondó szolgáltatását 2003 októberében a Westel. Felhasználói maguk dönthetik el, hogy elolvassák, vagy inkább meghallgatják üzeneteiket.”

3.3 „6. Fejezet 24. oldal A szerző megközelítésében a hosszabb elemek kiválasztása növeli a minőséget. Ideális esetben lehetséges/szöveg, vagy mondat szereplejen az adatbázisban. Nem kerül konkrét kifejtése az adatbázis komplexitása és a hosszú elemekből építkező beszédszintézis minősége közötti összefüggés ? Erre vonatkozóan semmilyen becslés nincs a dolgozatban. „Természetesen ez a gyakorlatban kivitelezhetetlen, ezért olyan egységeket rögzítenek az adatbázisba, hogy a szintetizálendő mondat nagy valószínűséggel hosszú elemekből legyen összefűzhető.” Milyen statisztikai módszer alapján lettek ezek kiválasztva, hogy a fenti kritériumot teljesítsék ?”

Köszönöm bírálóm felvetését. Az időjárás jelentés (IDO1 adatbázis), ill. a fonetikailag kiegyenlített (FON1-10 adatbázisok) korpusz építéssel kapcsolatos adatok alapján kívánom a kérdését megválaszolni. A 2000 mondatból épített fonetikailag kiegyenlített adatbázis hossza átlagosan kb. 2 óra, amit természetesen befolyásol a felolvasó személy artikulációs és beszédsebessége. Tapasztalataink szerint még professzionális bemondóval is legfeljebb napi négy órát lehet stúdióban tölteni, mert egy idő után a figyelme és a hangképző szervei is fáradni kezdenek. A 2000 mondatot bemondónként átlagosan három stúdió napon, mintegy 12 óra stúdió idő felhasználásával tudtuk felvenni, tehát a „tisza” adatbázis idő mintegy hatszorosát kell stúdióban tölteni.

Az időjárás jelentés témakörében a reprezentatív szöveges adatbázis mintegy 56.000 mondatot tartalmaz. A fent említetteket arányosítva ennek a felvételéhez $28 \cdot 12 = 336$ órát kellene stúdióban tölteni, mintegy 84 napon keresztül. Ekkor tökéletes emberi minőségben lehetne előállítani az időjárás jelentés jellemző szövegeit. Nyilvánvaló, hogy a gyakorlatban ez nehezen megvalósítható. Ezért azt a peremfeltételt határoztam meg, hogy legfeljebb 6000 mondatot tartalmazhat az adatbázis, ami 9-10 stúdió nap alatt felvehető. Ehhez ún. mohó algoritmust

alkalmaztam¹, aminek a lényege az, hogy a rendelkezésre álló mondatokból minden lépésben azt választja ki, ami a lehető legtöbb, még nem kiválasztott szót lefedi a szó-gyakorisági listából. Az először kiválasztott mondatok jellemzően meglehetősen hosszúak, majd az iteráció előre haladtával a kiválasztott mondatok hossza is csökken. Az így lecsökkentett adatbázis (IDO1, 630 perc) hangfelvétele is közel két hónapos munkát igényelt. A megvalósított időjárás felolvasó rendszer évekig üzemelt a metnet.hu honlapon és az Időjárás Mindenkinek Windows8 alkalmazásban pozitív felhasználói visszacsatolással.

3.4 „A vonatkozó tézisben is említett 26. oldal 9. ábrája magától értetődőnek tűnik, szükséges lenne annak alátámasztása, hogy ezen ábra megalkotásához milyen tudományos kihívások megoldására volt szükség.”

A hagyományos eljárások jellemzően előállítják az adott hangsorozathoz kapcsolódó kívánatos prozódiai jellemzőket (alapfrekvencia, időtartam és intenzitás menet) és az ezekkel a jellemzőkkel bíró akusztikai elemeket választják ki az adatbázisból. Ezt tekintjük direkt prozódiai modellnek. A vonatkozó eredeti ötlet az, hogy ha megcímkézzük az adatbázis elemeket (jellemzően szó szinten) a mondatban elfoglalt helyük alapján, akkor jó eséllyel kizárólag a szöveges pozíció elemzés segítségével is helyes prozódia áll elő. Ezt hívom indirekt prozódiai modellnek. A megoldás helyességét számos szubjektív teszt és gyakorlati alkalmazás igazolja. Mint a legtöbb elvi modell, ez sem tökéletes, ezért a durva hibák elkerülése érdekében szükséges néhány szabály alapján ellenőrizni az előállított elem sorozatot. Például a kijelentő mondat végén mindig csökken az alapfrekvencia. Ha a keresési algoritmus itt hibát eredményez, jelfeldolgozó algoritmus segítségével korrigálunk.

3.5 „A prozódiai címkézés utáni, a címke illeszkedést kereső „válogató függvényt” nem részletezi a szerző. Nincs vizsgálat arra vonatkozólag, hogy például hogyan lehet a legyorsabb keresést alkalmazni.”

Köszönöm bírálóm felvetését. Ez első lépésben egy szó-szintű keresés, ami gyorsan lefut. Viszont a hang szintű illesztésnél érdekes lehet a gyorsítás lehetőségét vizsgálni. A jövőben erre figyelmet fogok fordítani.

3.6 „A 27. és 28. oldalon két költségfüggvényt $C(n)$ és $P(n)$ definiál a szerző, majd ezek összegét optimalizálja. A $C(n)$ függvény súlyértékeit „500 mondat többszöri szintézisével határoztuk meg”. Konkrétan hogyan? Másrészt miért állítja elő a teljes célfüggvényt additív formában $C(n)$ -ből és $P(n)$ -ből, hiszen az összegfüggvény optimuma távol eshet mind $C(n)$ és $P(n)$ optimumától és nem biztos, hogy jó kompromisszumos megoldást szolgáltat. Miért nem kényszeres optimalizálásként tekint a problémára, ahol az egyik függvény adott minőségű kényszere mellett keresi a másik optimumát. Az sem kerül részletezésre, hogy milyen statisztika alapján generált korpuszban találhatunk nagy valószínűséggel szófüzéreket (nemcsak szavakat)?

Köszönöm bírálóm megjegyzéseit. Sajnálom, hogy az áttekintő ismertetés nehezen követhető. A költségfüggvények súlyértékeinek kiinduló értékeit a 28-30. oldalon ismertetett elvek alapján állítottuk be. Ezután a beszédkorpuszban nem szereplő, de a témakörbe illeszkedő 500 mondatot szintetizáltunk. Ha a hangsorban hibát észleltünk, akkor megvizsgáltuk, hogy az

¹ L. R. Cormen, „Chapter 17 "Greedy Algorithms",” in Introduction to Algorithms, McGraw-Hill, 1990, p. 768.

optimális szó- ill. hangsorozat állt-e elő vagy szuboptimális eredményt hozott-e létre az algoritmus? Ehhez egy olyan fejlesztői eszközt készítettünk, amivel rugalmasan lehetett az egy-egy pozícióban felmerülő adatbázis elemeket (jellemzően n -best = 5) megtekinteni és az alternatívákat meghallgatással tesztelni. A megfelelő elemek kiválasztásához iteratív módosítást végeztünk a súlyokon. Ez is jelentős időt vett igényben, de a kutatás időpontjában nem létezett objektív mérési módszer. Napjainkban került előtérbe a mélytanuló módszerek felhasználásával az ilyen feladatokhoz felhasználható objektív minőség mértékek kutatása.

A célfüggvény összeg formában való előállítását indokolja, hogy bármely részfüggvény magas értéke rossz minőségű végeredményt szolgáltat. Ezen kívül a részfüggvények egymásra is hatnak. Ha az egyik szempont szerint másik elemet választok, akkor az kihat a másik értékeire is. Ráadásul az objektív költségfüggvény számérték nem feltétlenül korrelál a szubjektív érzékeléssel, noha éppen ezt a célt próbáljuk elérni a fonetikai szempontok szerint kialakított költségekkel.

A cél-orientált megközelítésnek megfelelően az adott alkalmazási területre reprezentatív szövegtörzset alapján állítjuk elő a felolvasandó mondatok listáját. Ennek mérete függ az alkalmazás követelményeitől is. Például az időjárás-jelentés témaköre nagy változatosságú, de bizonyos kifejezések (pl. Holnap napos idő lesz.) az átlagosnál gyakrabban fordulnak elő. Ha ezt tudjuk, és bevesszük az adatbázisba, akkor a rendszernek ezt meg kell találni és a komplett mondatot kell visszaadni. Hasonlóan az utastájékoztató esetén a „Keleti pályaudvar” és annak ragozott formái is gyakran előfordulnak, ezért jó eséllyel be kell venni az adatbázisba és a kereső algoritmusnak jó működés esetén meg kell találni azt. A tervezésnél azt is figyelembe kell venni, hogy pl. az időjárás-jelentés alapvetően kényelmi szolgáltatás, ezért a hiba jobban elfogadható, mint az üzemeltetésben alapvető pályaudvari utastájékoztatónál.

3.7 „7. Fejezet, 5 A HMM beszédkeltésre való használatánál nem kerül kifejtésre a tanulási algoritmusok kérdésköre (pl. Baum-Welch algoritmus), hogy ezek milyen gyorsan érik el az optimumot.”

Köszönöm bírálóm megjegyzését. Sajnálom, hogy nem jeleztem elég egyértelműen, hogy a magyar nyelvű HMM TTS rendszer megalkotásakor a nagoyai egyetemen fejlesztett HTS keretrendszer eszközként használtam, beleértve a tanulási algoritmust² is.

4. Összefoglaló és javaslat a nyilvános vitára bocsátásról

„ ... az eredmények minősége megfelel az MTA doktora címmel szemben támasztott követelményeknek és a disszertációnak a nyilvános vitára bocsátását mindenképpen javaslom.”

Köszönöm dr. Levendovszky János professzor úr támogató összegző sorait.

Budafok, 2021. július 8.

dr. Németh Géza

² [39] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi és T. Kitamura, „Speech parameter generation algorithms for HMM-based speech synthesis,” in Proc. of ICASSP, Istanbul, Turkey, 2000

Melléklet

1. A ProfiVox rendszer minőségét dicsérő amerikai e-levél váltás

From: Zsolt
Sent: Tuesday, May 10, 2005 5:37 PM
To: Mihály
Subject: Fw: Speech synthesizer information

Szia Misi!

Továbbítom neked azt a levélváltást, melyben a Profivox-ot dicsérik.

--

Üdv:

..... Zsolt

A JAWS képernyőolvasó program honosítási projektjének vezetője.

-----Eredeti üzenet-----

Feladó: [Oleg](#)

Címzett: [.... Zsolt](#)

Elküldve: 2005. május 9. 21:19

Tárgy: Re: Speech synthesizer information

Hello Zsolt,

Well, I got persuaded of a couple of things today.

1. This Profivox synthesizer is a true winner--it is absolutely fantastic! It is fast, clear and responsive, and a true pleasure to listen to.

2. It must take eternity to learn Hungarian. With other European language, you often find familiar roots, so understanding is fairly easy. Not so with Hungarian! I've experienced this during my trips to Budapest in the past; I've also heard this from a good friend of mine who is now the pastor at Danube International Church in Budapest. But here is yet another chance to experience this phenomenon, when you try to catch at least a single familiar word--and you fail miserably! :)

Oleg

----- Original Message -----

From: [.... Zsolt](#)

To: [Oleg](#)

Sent: Monday, May 09, 2005 9:43 PM

Subject: Re: Speech synthesizer information

:~))

I'm waiting for your opinion.

--

Üdv:

..... Zsolt

A JAWS képernyőolvasó program honosítási projektjének vezetője.

-----Eredeti üzenet-----

Feladó: [Oleg](#)

Címzett: [.... Zsolt](#)

Elküldve: 2005. május 9. 19:24

Tárgy: Re: Speech synthesizer information

Thanks, I am downloading it right now!

Oleg

----- Original Message -----

From: [.... Zsolt](#)

To: [Oleg](#)

Sent: Monday, May 09, 2005 9:05 PM

Subject: Re: Speech synthesizer information

Hi Oleg,

Here is the download link for JAWS 5.10 Hungarian version:

[.....](#)

Just download the program and start it!

The Run-time JAWS uses Profivox to speak the setup screens.

--

Regards:

Zsolt

Manager of the Hungarian JAWS localization project.

-----Eredeti üzenet-----

Feladó: [Oleg](#)

Címzett: [.... Zsolt](#)

Elküldve: 2005. május 9. 18:55

Tárgy: Re: Speech synthesizer information

Hello Szolt,

Thank you for the info on Profivox. This is quite helpful! I am told by Stephen that Profivox is the best speech synth he has heard so far. Unfortunately, I haven't had a chance to hear it in action, but Stephen's words are quite trustworthy.

Oleg

----- Original Message -----

From: [Torma Zsolt](#)

To: [..... @freedomscientific.com](#)

Cc: [Kiss Géza](#) ; [Németh Géza](#)

Sent: Monday, May 09, 2005 8:31 PM

Subject: Speech synthesizer information

Hi Oleg,

Here are some information about our speech synthesizer. In Hungary, We using only this product with JAWS.

1. Product name.
Profivox

2. Manufacturer's name and URL.
Budapest University of Technology and Economics, Department of Telecommunications and Informatics, Laboratory of Speech Technology

3. Is this synthesizer SAPI4 or SAPI5?
The TTS is integrated into Jaws using Jaws proprietary interface, but the SAPI4 interface is implemented, and implementation of SAPI5 is underway.

4. Is it available for download and purchase to end-users, or is it for resellers only?
At present it is available to end-users as part of the Jaws program. Besides this, it is available to companies based on a contract established with the specific company.

5. Any other remarks you consider relevant.
The TTS is available for other European languages also, namely German, Polish, Spanish.

6. JAWS implementation:

We have two voices: a male and a female.

Both "person" can speak text as it is written or by using the Hungarian pronunciation rules.

The following speech parameters can be set: rate, pitch, volume and punctuation.

If you have any questions, please contact me or the developers (their address can be found in the CC field of this message).

--

Regards:

Zsolt

Manager of the Hungarian JAWS localization project