



OSTIS-2013

(Open Semantic Technologies for Intelligent Systems)

УДК 004.934.5

КАМПАНЕНТЫ ДЛЯ РОЗНЫХ ПЛАТФОРМАЎ СІНТЭЗАТАРА МАЎЛЕННЯ ПА ТЭКСЦЕ ДЛЯ ІНТЭЛЕКТУАЛЬНЫХ СІСТЭМ

Гецэвіч Ю.С.* , Пакладок Д.А.* , Брэк Д.В.**

* *Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі, г. Мінск, Рэспубліка Беларусь*

yuras.hetsevich@newman.bas-net.by

dima.pokladok@gmail.com

** *Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі,*

г. Мінск, Рэспубліка Беларусь

bartwader@gmail.com

Разгледжаны пытанні актуальнасці і практычнай пабудовы кампанентаў двухмоўнага сінтэзатара беларускага і рускага маўлення па тэксце для мабільных і інтэрнэт-платформаў. Вынікі тэставання паказваюць, што сінтэзатар маўлення па тэксце для Інтэрнэта можа цалкам задавальняць запыты па агучванні старонак сайтаў, а таксама ён можа быць выкарыстаны ў якасці завершанага кампанента маўленчага інтэрфейсу любым знешнім распрацоўшчыкам праз Інтэрнэт для пабудовы ўласных інтэлектуальных сістэм.

Ключавыя словы: кампанент; сінтэз маўлення па тэксце; мабільны сінтэзатар маўлення па тэксце; пабудова Інтэрнэт-версіі сінтэзатара маўлення па тэксце.

Уводзіны

Сінтэзатары маўлення па тэксце (СМТ) на сённяшні момант дасягнулі высокай якасці і сталі складанымі шматкампанентнымі сістэмамі. Кожны з кампанентаў СМТ паэтапна працуе над вынікам папярэдняга і сам з'яўляецца завершаным рашэннем. Гэтая асаблівасць СМТ дазваляе выкарыстоўваць розныя яе кампаненты ў розных інтэлектуальных сістэмах.

Сістэмы сінтэзу маўлення па тэксце на сённяшні момант дасягнулі высокай якасці як па крытэрыях выразнасці і натуральнасці сінтэзаванага голасу, так і па тэхнічных характарыстыках, што спрыяе іх шырокаму ўжыванню ў цэнтрах апрацоўкі запытаў, пры агучванні станаў працы складаных аб'ектаў, пры стварэнні аўдыёкніг і г.д. [Лобанов и др., 2008]. У той жа час застаецца важнай задачай адаптацыя сістэм сінтэзу маўлення па тэксце (іх алгарытмаў і рэсурсаў) для розных вылічальных платформаў, якія маюць розныя сістэмныя абмежаванні дыскавай і/ці апэратыўнай памяці, што далей накладвае абмежаванні на выкарыстанне ўсіх распрацаваных алгарытмаў і лінгвістычных рэсурсаў СМТ. Гэтая задача ўзнікае з-за таго, што звычайна распрацоўшчыкі прымаюць за аксіому, што рэсурсы вылічальнай сістэмы *неабмежаваныя: тэарэтычна,*

калі прыняць вылічальную машыну за машыну Т'юрынга [Barker-Plummer, 1995], ці *практычна,* калі яны нашмат перавышаюць максімальныя паказчыкі выкарыстання памяці і тактавага працэсара алгарытмамі і рэсурсамі СМТ.

Аўтаматычныя маўленчыя тэхналогіі для кожнай канкрэтнай мовы без адаптацыі СМТ для мабільных і інтэрнэт-платформаў застаюцца недаступнымі для шырокага кола карыстальнікаў ва ўсім свеце [Internet World Stats, 2012a], у тым ліку ў Беларусі, дзе кожны другі чалавек карыстаецца Інтэрнэтам [Internet World Stats, 2012b], а колькасць абанентаў сотавай сувязі на 2012 г. перакročыла 10 млн [IT.TUT.BY, 2012].

Шырокае распаўсюджанне можа атрымаць выкарыстанне сістэм сінтэзу маўлення па тэксце на мабільных прыладах, такіх як кішэнныя персанальныя камп'ютары, смартфоны, мабільныя тэлефоны. Узорами прыстасаванняў з СМТ могуць быць такія, якія чытаюць электронныя кнігі, указанні аўтамабільнай навігацыйнай сістэмы; агучваюць SMS-паведамленні ці лісты электроннай пошты.

Пабудова сінтэзатара маўлення па тэксце для Інтэрнэта абумоўліваецца наступнымі перспектывамі:

- сляпыя ці слабабачыя атрымаюць добрую мажлівасць, каб азнаямлівацца з сайтамі ў Інтэрнэце, у той час як уладальнікі сайтаў змогуць павысіць камфортнасць для карыстальнікаў: камусьці можа быць зручна не чытаць сайты, а слухаць;

- Інтэрнэт-контэнт утрымлівае самыя дынамічныя і разнастайныя па напісанню тэксты. У іх ёсць шмат новых слоў, асаблівых тэкставых выказаў, абдруковак, што з'яўляецца добрым матэрыялам для тэставання і паляпшэння працы СМТ. У выніку СМТ змога сканаваць вялікую колькасць тэкстаў, аналізаваць і апрацоўваць іх, паказваць складаныя (ці неадназначныя) месцы экспертам;

- распрацоўшчыкі СМТ змогуць дэманстраваць сінтэзатар у розных рэжымах выкарыстання (поўны ці абмежаваны доступ) вялікай колькасці карыстальнікаў і мэтавых груп (для навуковых даследаванняў, для школьнікаў і студэнтаў, для прадпрыемстваў і патэнцыяльных замоўцаў, якія хочуць пабудаваць камерцыйныя сэрвісы агучкі тэкстаў).

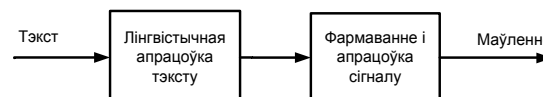
Большасць сучасных сінтэзатараў рускага маўлення па тэксце для мабільных прыладаў, напрыклад, Acapela TTS for Windows Mobile, Nuance TALKS, Mobile Speak і інш., працуюць на смартфонах пад кіраваннем аперацыйных сістэм Windows Mobile ці Symbian. Аднак смартфоны, на якіх магчыма праца гэтых СМТ, складаюць, па розных падліках, толькі 10–15 % рынку мабільных тэлефонаў. Для беларускай мовы такія сістэмы толькі пачынаюць стварацца ў выглядзе асобных блокаў СМТ [Гецэвіч і інш., 2011]. Пры гэтым асноўная маса мабільных апаратаў, якія выкарыстоўваюцца ў цяперашні час, характарызуецца нізкай хуткадзейнасцю і адносна невялікім легальна дазволена для выкарыстання аб'ёмам апэратыўнай памяці (да 24 Мб) [Mihai Preda, 2010]. У той жа час сучасныя СМТ патрабуюць вялікага аб'ёма памяці для захоўвання і выкарыстання лінгвістычных і акустычных рэсурсаў, што не дазваляе наўпрост «перанесці» існуючыя сістэмы на мабільныя платформы [Цирульник и др., 2012].

Інтэрнэт-платформы звычайна характарызуюцца аперацыйнай сістэмай (UNIX, Windows), архітэктурай выканання (C, ASP.NET, Java, Python, Pike, Ruby, Erlang), мовамі праграмавання (ASP, C, ColdFusion Markup Language, Java, JavaScript, Lua, Perl CGI, PHP, Python, Ruby, SMX, Lasso, WebDNA, Progress WebSpeed), фізічнай памяццю і выдзеленай апэратыўнай памяццю для скрыпта (звычайна каля 128 МБ), які выконваецца. На дадзены момант існуе некалькі інтэрнэт-версій СМТ для рускай мовы. Камерцыйныя кампаніі Acapela Group [Acapela Group, 2012] і Nuance на сваіх сайтах дазваляюць выкарыстоўваць дэма-вокны для сінтэзавання любога тэксту [Acapela Group, 2012], кампанія ЦРТ дэманструе інтэрнэт-сэрвісы на аснове СМТ для стварэння аўдыёпапаштовак і агучкі rss-

каналаў [Центр Речевых Технологий, 2001]. Гэтыя СМТ патрабуюць асобнага сервера, дзе могуць працаваць скампіляваныя коды, якія выконваюцца, пад канкрэтную аперацыйную сістэму. Таму яны з'яўляюцца стацыянарнымі СМТ з інтэрнэт-доступам. Яны складаны пераносіцца на іншыя серверы, бо патрэбна пастаянна дамаўляцца з правайдэрам хостынга аб асаблівых умовах па размяшчэнні СМТ у якасці аднаго з галоўных працэсаў сервера. Гэта каштуе нашмат больш, чым выкананне скрыптоў на мовах праграмавання, які ўстаноўлены на гэтым серверы па змоўчванні. У літаратуры таксама не сустракаецца апісання СМТ для крос-платформавых моваў праграмавання, напрыклад, для Perl CGI, PHP, Python і інш. Таму пабудова СМТ на інтэрнэт-платформе на скрыптавай мове праграмавання з'яўляецца новай навуковай задачай.

1. Падыход да распрацоўкі шматкампанентнага сінтэзатара маўлення па тэксце для мабільных і інтэрнэт-платформаў

СМТ – гэта сістэма, здольная генераваць маўленне па тэксце. На працягу апошніх дзесяцігоддзяў яна замацавалася як сістэма, якая складаецца з двух агульных кампанентаў (малюнак 1). Гэта кампанент лінгвістычнай апрацоўкі тэксту да фанемнага выгляду з пазнакамі націску слоў, інтанацыі (прасодыі) і рытма, а таксама кампанент апрацоўкі маўленчага сігналу, які пераўтварае раней атрыманы фанемны выгляд у гукавы сігнал маўлення [Dutoit, 1997].



Малюнак 1 – Агульная функцыянальная схема сінтэзу маўлення па тэксце

Такая архітэктурал дазваляе распрацоўшчыкам СМТ будаваць іх «з канца»: спачатку распрацаваць кампанент атрымання маўлення па фанетычным кодзе, уласцівым для абранай мовы, а пасля распрацоўваць алгарытмы лінгвістычнай апрацоўкі тэкстаў для абранай мовы ў фанемны выгляд. Такі падыход з'яўляецца зручным, бо падчас стварэння спачатку другога кампаненту распрацоўшчык атрымлівае адказ на пытанне, ці магчыма ўвогуле пабудаваць сінтэзатар маўлення па тэксце для дадзенай платформы. Калі акустычны кампанент пабудаваць удаецца, то і астатнія кампаненты можна будзе пабудаваць у будучым, бо звычайна большасць платформаў дазваляе апрацоўваць тэкставыя дадзеныя.

Такім чынам, паводле вышэй прадстаўленай архітэктурал ў наступных раздзелах артыкула будзе апісаны агульны падыход пераносу атрыманых алгарытмаў і лінгвістычных рэсурсаў пры распрацоўцы акустычнага працэсара (АП) мабільнага СМТ [Цирульник и др., 2010][Гецэвіч і

інш., 2011] для пабудовы АП СМТ для Інтэрнэт-платформы Linux (Apache) пад мову праграмавання РНР [Гецэвіч і інш., 2012]. А далей апісваюцца новыя кампаненты фанетычнага, прасадычнага і тэкставага працэсараў СМТ для Інтэрнэт-платформы. У выніку прадстаўляецца эксперыментальны абразец шматкампанентнага Інтэрнэт СМТ, які можа быць выкарыстаны ў якасці завершанага кампанента маўленчага інтэрфейсу любым знешнім распрацоўшчыкам праз Інтэрнэт для пабудовы ўласных інтэлектуальных сістэм.

2. Алгарытм працы і рэсурсы шматмоўнага акустычнага працэсара

Падабенства беларускай (БМ) і рускай (РМ) моў дазваляе стварыць універсальную мадэль акустычнага працэсара, прыдатную для абедзвюх моў адразу. Гэтаму спрыяе падабенства фанетыкі, утварэння гукаў [Lobanov et al., 2006]. Алгарытмы мадыфікацыі сігналу працуюць з алафонамі па агульнай схеме, абстрагуючыся ад пэўнай мовы. Уваходнымі дадзенымі для акустычнага працэсара з'яўляюцца алафонныя паслядоўнасці кожнай сінтагмы тэксту, значэнні даўжыні кожнага алафона і кожнага перыяду асноўнага тону. Выходныя дадзеныя – файл ў фармаце WAV. У якасці рэсурсаў выступаюць адпаведныя для кожнай мовы фанетычныя групы і галасавыя базы, якія створаныя экспертамі. Заўважым, што для дзвюх моў выкарыстоўваюцца розныя па напаўненні, але агульныя па сэнсе фанетычныя групы (паняці) – галосныя, звонкія і глухія выбухныя і фрыкатыўныя зычныя, афрыкаты, санорныя зычныя (табліца 1).

Табліца 1 – Фанетычныя групы для беларускай і рускай моў

Фанетычныя групы	Склад груп	
	Беларуская мова	Руская мова
Галосныя	А,Е,І,О,У,У	А,Е,І,О,У,У
Выбухныя звонкія зычныя	В,В',Г,Г',Д	В,В',Г,Г',Д,Д'
Выбухныя глухія зычныя	К,К',Р,Р',Т	К,К',Р,Р',Т,Т'
Афрыкаты	ДЗ,ДЗ',ДЗН,С,С',СН	С,СН',СН'
Фрыкатыўныя звонкія зычныя	В,В',ГН,ГН',ЗН,З,З'	В,В',ЗН,З,З'
Фрыкатыўныя глухія зычныя	С,С',Ф,Ф',Н,Н',Ш	С,С',Ф,Ф',Н,Н',Ш
Санорныя зычныя	Л,Л',М,М',Н,Н',Р,Р',Ж,Ж'	Л,Л',М,М',Н,Н',Р,Р',Ж

Алгарытм працы акустычнага працэсара для канкрэтнай мовы наступны [Цирульник и др., 2010]:

- 1) Атрыманне чарговага алафона з сінтагмы.
- 2) Вызначэнне групы алафона з дапамогай лінгвістычных рэсурсаў.
- 3) Змена даўжыні і, пры неабходнасці, кожнага перыяду асноўнага тону алафона для дадзенай фанетычнай групы з дапамогай алгарытму плаўнай сшыўкі.

4) Запіс чарговага алафона ў выніковы гукавы паток.

5) Выхад: гукавы паток – WAV-файл.

Калі разгледзець склад галасавых баз БМ і РМ для мабільных і інтэрнэт-платформаў (табліца 2), то колькасць алафонаў для беларускай і рускай моў адрозніваецца нязначна. Гэта абумоўлена адной групай моў – усходнеславянскай. З іншага боку, ёсць істотнае адрозненне па колькасці (і, адпаведна, па агульным памеры) для розных платформаў – галасавая база для мабільнай платформы менш на 56 % (92,6 % па памеры) для беларускай мовы і на 56 % (92,8 % па памеры) для рускай.

Табліца 2 – Параўнальная характарыстыка галасавых баз беларускай (БМ) і рускай (РМ) моў для мабільных і інтэрнэт-платформаў

Мова	Інтэрнэт-платформа		Мабільная платформа	
	Колькасць алафонаў	Памер, КБ	Колькасць алафонаў	Памер, КБ
БМ	1889	11822	832	869
РМ	1779	10699	784	773

3. Акустычны працэсар для мабільнай платформы

Памяншэнне галасавых баз для мабільнай платформы дасягалася шляхам памяншэння частаты дыскрэтызацыі і разраднасці зыходнай базы (22 кГц 16 біт у 8 кГц 8 біт) і выдаленнем дыалафонаў. Дадзеная аперацыя дазволіла запрацаваць акустычнаму працэсару на большасці сучасных мабільных тэлефонах. Так, распрацаваная праграма была паспяхова пратэставана на тэлефонах Motorola, Sony-Ericsson, LG, якія характарызуюцца тактавай частотой ARM-працэсара ад 68 да 115 МГц, аб'ёмам памяці ад 3 500 да 4 200 Кб, падтрымкай канфігурацыі CLDC 1.0 і профіля MIDP 2.0 [Гецэвіч і інш., 2011]. Сістэма дазваляе сінтэзаваць маўленчы сігнал у рэальным часе на мабільных тэлефонах з ARM-працэсарамі сёмага пакалення.

Галоўнае меню сістэмы (малюнак 2а) складаецца з пунктаў-выклікаў:

- тэкставага файла для прайгравання;
- наладак;
- прайгравання маўленчага сігналу;
- справачнай інфармацыі па сістэме.

Для сінтэзу маўлення карыстальнік павінен указаць тэкставы файл з алафонным тэкстам, а затым выбраць пункт меню «Воспроизведение». Меню наладак складаецца з выбара галасавой базы і выбара прасадычнага стылю для сінтэзу маўлення па тэксце (малюнак 2б). Для кіравання агучваннем тэксту ў сістэме рэалізаваны функцыі паўзы/працягу, а таксама поўнага спынення сінтэзу маўлення ў рэальным часе (малюнак 2в).



а) б) в)

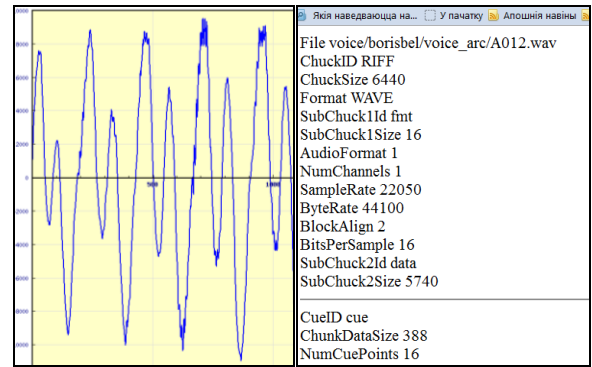
Малюнак 2 – Інтэрфейсы сінтэзатара маўлення па тэксце для мабільных платформаў: а – галоўнае меню; б – меню наладак; в – прайграванне.

4. Кампаненты сінтэзатара маўлення па тэксце для Інтэрнэт-платформы

У адрозненне ад мабільнай платформы інтэрнэт-платформа дазваляе выкарыстоўваць амаль неабмежаваныя аб'ёмы дыскавай памяці, таму галасавая база для яе не мадыфікавалася. Асноўная ўвага была звернута на аптымальнае выкарыстанне аператыўнай памяці, бо на яе абмежаванні застаюцца ў выглядзе наладак ад адміністратара сервера для кожнага скрыпта-працэса, які выконваецца ў канкрэтную адзінку часу.

Акустычны працэсар (АП) для Інтэрнэта будзеца на скрыптавай мове праграмавання PHP. Гэта робіцца для таго, каб любыя напісаны скрыпт мог лёгка пераносіцца на любую іншую платформу (ці хостынг), дзе ёсць PHP. У літаратуры не сустракаецца апісанняў працы з гукавымі файламі на PHP, тым больш няма апісанняў рэалізацыі СМТ для гэтай мовы праграмавання, акрамя ранейшай публікацыі аўтараў гэтага артыкула [Гецэвіч і інш., 2012].

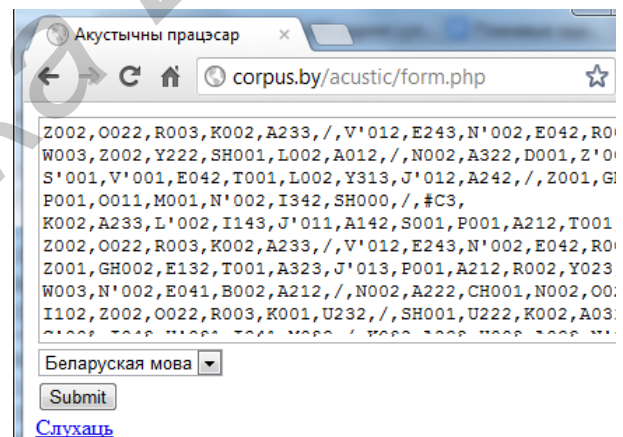
Спачатку для прагляду асцыляграм гукавых файлаў галасавой базы быў пабудаваны спецыяльны інтэрфейс, які даступны па спасылцы ў рэальным часе – <http://corpus.by/acoustic/wave.php>. У выніку зараз любы алафон магчыма «правізуалізаваць» і прагледзець яго амплітудна-часавую хвалю ў Інтэрнэце. Напрыклад, была пабудаваная асцыляграма алафона А212 (малюнак 3а). Далей з дапамогай стандартных функцый чытання файлаў на PHP была пабудаваная эксперыментальная праграма для прагляду зборнай інфармацыі (асноўныя характарыстыкі запісанага сігналу, размечаныя рэгіёны і маркеры) любога абранага WAV-файла з галасавой базы (www.corpus.by/acoustic), напрыклад, для алафона А012 з беларускай галасавой базы (малюнак 3б).



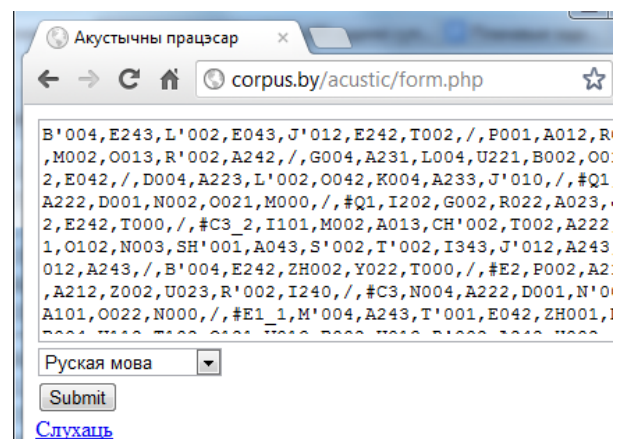
а) б)

Малюнак 3 – Вынікі працы дапаможных эксперыментальных мадэляў, якія рэалізаваныя на PHP для працы з любым файлам гукавой базы: а) вывад асцыляграмы; б) вывад асноўных характарыстык

Пабудаваныя эксперыментальныя мадэлі пацвердзілі магчымасць працы з гукавымі файламі ў WAV-фармаце праз інтэрпрэтуемую мову праграмавання PHP. Таму далей была рэалізаваная эксперыментальная праграма мадэль базавых алгарытмаў акустычнага працэсара на PHP. На малюнках 4 і 5 прадстаўлены інтэрфейсы для правэркі працы распрацаванага акустычнага працэсара СМТ у Інтэрнэце для беларускай ці рускай гукавой базы.



Малюнак 4 – Эксперыментальны акустычны кампанент сінтэзатара беларускага маўлення па тэксце ў выглядзе інтэрнэт-сайта



Малюнак 5 – Эксперыментальны акустычны кампанент сінтэзатара рускага маўлення па тэксце ў выглядзе інтэрнэт-сайта

Двухмоўны АП даступны па спасылцы www.corpus.by/acoustic/form.php. У яго інтэрфейсе карыстальнік можа напісаць фанетычны тэкст ці выкарыстаць уведзены, пазначыць мову і падаць запыт праз Submit на сервер. Праграмны скрыпт акустычнага працэсара прымае запыт, апрацоўвае яго па абзначаным вышэй алгарытме ў п. 2. У выніку будзе згенераваны гукавы WAV-файл з пазнакай выдаткаванага часу на выкананне скрыпта. Па спасылцы з надпісам «Слухаць» (www.corpus.by/acoustic/test.wav) файл можна праслухаць ці захаваць на лакальным камп'ютары. Заўважым, што для генерацыі маўлення абедзвюх моў выкарыстоўваецца адна і тая ж рэалізацыя алгарытмаў акустычнага працэсара на PHP, змяняюцца толькі галасавыя базы.

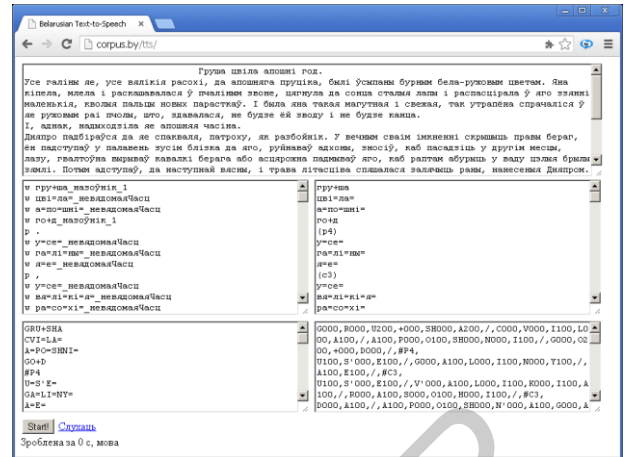
У табліцы 3 прадстаўлены вынікі экспрэс-тэставання акустычнага працэсара на абранай інтэрнэт-платформе (UNIX, Apache, PHP, 64 Мб аператыўнай памяці) для абедзвюх моў. Для кожнай мовы тэсты праводзіліся на адным і тым жа адпаведным тэксце, які быў шмат разоў раскапіраваны для павелічэння тэставай колькасці сінтагмаў. Вынікі паказваюць, што сістэма на такой платформе за адзін запыт здольная выканаць сінтэз маўлення алафоннага тэкста, аб'ём якога не большы за 780 Кб для БМ і не большы за 420 Кб для РМ. Для параўнання, сярэдняя велічыня старонкі ў Інтэрнэце не перавышае 320 Кб па дадзеных Google на 2010 г. [Ramachandran, 2010].

Табліца 3 – Экспрэс-тэставанне акустычнага працэсара ў Інтэрнэце

БМ			РМ		
Кольк асць сінтагмаў	Аб'ём тэксту, Кб	Час адказу сервера, с	Кольк асць сінтагмаў	Аб'ём тэксту, Кб	Час адказу сервера, с
5 400	520	6	3800	280	3
8 100	780	12	5700	420	7
10 800	1040	няма адказу	7600	560	няма адказу

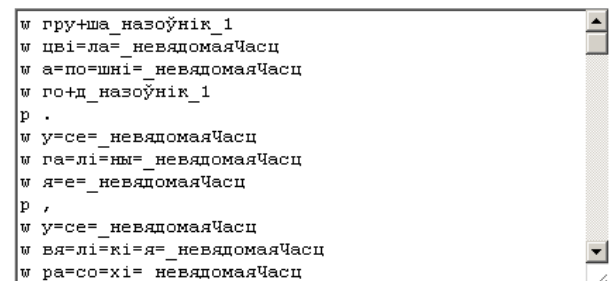
Такім чынам, пабудаваныя эксперыментальныя праграмныя мадэлі і экспрэс-тэсты пацвярджаюць, што магчыма рэалізаваць сінтэзатар маўлення па тэксце для беларускай і рускай моў для інтэрнэт-платформаў з дапамогай скрыптавай мовы праграмавання PHP.

На малюнку 6 паказаны вэб-інтэрфейс сінтэзатара беларускага маўлення па тэксце, які даступны па спасылцы www.corpus.by/tts. Дыялогавае акно складаецца з пяці тэкставых акон і кнопкі «Start!» для запуску сінтэзатара. У першае акно ўводзіцца тэкст для сінтэзу. Выніковы файл з сінтэзаваным маўленчым сігналам можна праслухаць ці спампаваць па згенераванай спасылцы «Слухаць».



Малюнак 6 – Вэб-інтэрфейс сінтэзатара маўлення па тэксце (спасылка www.corpus.by/tts)

Вынік працы тэкставага кампанента сінтэзатара можна прагледзець на малюнку 7. На дадзеным этапе ўваходны тэкст быў тыпізаваны на словы (w) і знакі прыпынку (p). У кожным слове быў прастаўлены націск праз знакі «+» і «<» (ex. гру+ша), прыпісаны маркер катэгорыі мовы (ex. назоўнік) і прыкрытэту слова (ex. 1) паводле базы дадзеных (унутранага слоўніка S). У тым выпадку, калі слова не было знойдзена ў слоўніку S, тады выстаўляўся частковы націск на кожны склад (ex. цві=ла=) і ставіўся маркер, што часціна мовы невядомая (ex. цві=ла=_невядомаяЧасц). Дадаваны кампанент можа выкарыстоўвацца незалежна ад Інтэрнэт СМТ інтэлектуальнымі сістэмамі для вызначэння тыпу паслядоўнасці сімвалаў сказу, для знаходжання націску і лексіка-граматычнай інфармацыі любога слова сказаў.



Малюнак 7 – Вынік працы тэкставага кампанента Інтэрнэт СМТ

Вынік працы тэкставага працэсара быў прааналізаваны прасадыхным працэсарам і ў выніку былі вылучаны сінтагмы і пазначаны іх інтанацыйныя тыпы маўлення (С, Р, Q, E). На малюнку 8 прадстаўлены вынік працы прасадыхнага кампанента сінтэзатара маўлення. Напрыклад, С3 – інтанацыя незавершанасці, Р4 – інтанацыя завершанасці.

```

гру+ша
цві=ла=
а=по=шні=
го+д
{р4}
у=се=
га=лі=ны=
я=е=
{с3}
у=се=
вя=лі=кі=я=
ра=со=хі=

```

Малюнак 8 – Вынік працы прасадыхнага кампанента Інтэрнэт СМТ

Наступны крок – фанетычная апрацоўка сінтагмаў. Усе літары кожнай сінтагмы па правілах пераўтварэння літара-фанема замяняюцца на адпаведныя фанемы (малюнак 9).

```

GRU+SHA
CVI=LA=
A=PO=SHNI=
GO+D
#P4
U=S'E=
GA=LI=NY=
A=E=

```

Малюнак 9 – Вынік працы фанетычнага кампанента Інтэрнэт СМТ

Правілы пераўтварэння фанетычнага кампаненту запісваюцца ў выглядзе фармальных замен кожнай літары на адпаведную фанему. У гэтым кампаненце правілы амаль без ніякіх дадатковых апрацовак выкарыстоўваюцца ў якасці аргументаў для рэгулярных выразаў пошуку-замены адных сімвалаў на іншыя:

Агульныя правілы

- A-A
- B-B
- V-V
- G-G
- D-D
- E-E

...

Выключэнні з агульных правілаў

- (O)[+=]-O
- [ЖШЦ](И)-Y
- [^АЕЁИОУЫЎЮЯ][_#](И)-Y
- [Ь](O)[+=]-J',O
- [Ь](И)-J',I
- [СШЖ](Ч)-SH'
- (T)[Ь][C]-C
- [EO](Г)[O][_#]-V
- (Б)\$-P
- (В)\$-F
- (З)\$-S
- (Д)\$-T
- (Г)\$-K
- (З)[КПСТФХЦШЩ]-S
- (З)[_#][КПСТФШ]-S
- (Л)[Н][Ц]-O

...

Спіс фанем змягчэння

- E

- Ё
- Ю
- Я
- И
- Б
- ...

Агульныя правілы змягчэння

- ([ХМНЛРБПВФДТЗСГК])[ЕЁЮЯИЬ]
- (Н)[НТСЗДЛЧ]
- (Л)[Л]
- (Р)[Р]
- ([МБПВФ])[МБПВФ]

...

Алафонны кампанент Інтэрнэт СМТ вызначае алафон фанемы X_i паводле яе натуральнага фанетычнага асяроддзя – фанемаў X_{i-1} і X_{i+1} (малюнак 10). Гэта ажыццяўляецца з дапамогай правілаў пераўтварэння фанема-алафон і знаходжаннем пазіцыі ў слове націскай галоснай фанемы.

```

G000,R000,U200,+000,SH000,A200,/,C000,V000,I100,LO
00,A100,/,A100,P000,O100,SH000,N000,I100,/,G000,O2
00,+000,D000,/,#P4,
U100,S'000,E100,/,G000,A100,L000,I100,N000,Y100,/,
A100,E100,/,#C3,
U100,S'000,E100,/,V'000,A100,L000,I100,K000,I100,A
100,/,R000,A100,S000,O100,H000,I100,/,#C3,
D000,A100,/,A100,P000,O100,SH000,N'000,A100,G000,A

```

Малюнак 10 – Вынік працы алафоннага кампанента Інтэрнэт СМТ

Правілы фанема-алафон уяўляюць сабой вынік эксперыментальнага даследавання розных фанетычных груп па іх фанетычным асяроддзі [Лобанов и др., 2008]. Адметна, што дадзеныя правілы могуць выкарыстоўвацца як для беларускай, так і для рускай моў. Яны маюць наступны выгляд:

Алфавіт фанем

- A0,A1,A2,A3,E0,E1,E2,E3,I0,I1,I2,I3,O0,O1,Y0,Y1,
- Y2,Y3,U0,U1,U2,U3,B,B',D,D',G,G',Z,Z',ZH,L,L',M,M',
- N,N',P,P',T,T',K,K',C,CH',F,F',S,S',SH,SH',H,H',V,V',J',
- R,R',#,

Цэнтральныя групы фанем

- C0:A0,A1,A2,A3,E0,E1,E2,E3,I0,I1,I2,I3,Y0,Y1,Y2,
- Y3,
- C1:U0,U1,U2,U3,O0,O1,
- C2:P,P',T,T',K',C,CH',F,F',S,S',SH,SH',H',CH*,C*,
- C3:K,G,H,GH*,

...

Групы фанем левага кантэксту

- L0:#,
- L1:P,B,F,V,M,U0,U1,U2,U3,O0,O1,W*,
- L2:SH,ZH,R,T,C,S,D,Z,N,L,A0,A1,A2,A3,E0,E1,E2,
- E3,Y0,Y1,Y2,Y3,DZ*,DZH*,CH*,
- L3:K,G,H,GH*,

...

Групы фанем правага кантэксту

- R0:#,
- R1:P,B,F,V,M,O0,O1,U0,U1,U2,U3,W*,
- R2:SH,ZH,R,T,C,S,D,Z,N,L,K,G,H,A0,A1,A2,A3,E0,

E1,E2,E3,Y0,Y1,Y2,Y3,DZ*,DZH*,CH*,

...

Правілы пераўтварэння для фанем з кожнай цэнтральнай групы

C0

L0R0-00

L1R0-10

L2R0-20

...

C2

L10R0-00

L10R4-01

L10R5-02

...

C3

L10R0-00

L10R6-01

L10R7-02

...

...

Такім чынам, былі распрацаваны чатыры кампаненты апрацоўкі дадзеных для сінтэзу маўлення па тэксце для Інтэрнэт-платформаў на аснове сумесных алгарытмаў для стацыянарных і мабільных платформаў. У спалучэнні з пятым кампанентам – АП, яны ўтвараюць паўнаватасны СМТ для Інтэрнэта.

5. Заключэнне

Пастаўлена і вырашана задача пераносу атрыманых алгарытмаў і лінгвістычных рэсурсаў мабільнага і стацыянарнага СМТ для распрацоўкі тэкставага, прасадыхнага, фанетычнага, алафоннага і акустычнага працэсараў СМТ для Інтэрнэт-платформы.

Зараз даступны карыстальнікам для тэставання акустычны кампанент у Інтэрнэце па спасылцы www.corpus.by/acoustic/form.php.

Выніковы эксперыментальны абразец шматкампанентнага Інтэрнэт СМТ, можа быць выкарыстаны ў якасці завершанага кампанента маўленчага інтэрфейсу любым знешнім распрацоўшчыкам праз Інтэрнэт для пабудовы ўласных інтэлектуальных сістэм. Ён даступны па спасылцы www.corpus.by/tts.

Значым таксама, што рэалізацыя акустычнага кампаненты сінтэзатара маўлення па тэксце на мабільных платформах дазваляе стварыць гібридную архітэктuru сінтэзу маўлення, пры якой кампанент апрацоўкі тэксту размешчаны на серверы ў Інтэрнэце, у той час як кампанент апрацоўкі маўленчага сігналу знаходзіцца на мабільным тэлефоне [Цирульник и др., 2010]. Дадатковай добрай якасцю такой архітэктury з'яўляецца магчымасць сінтэзу маўлення па адным і тым жа размечаным тэксце, які паступае з сервера, з выкарыстаннем розных галасоў і розных

прасадыхных стыляў, змешчаных на мабільным тэлефоне.

Такім чынам, на дадзены момант эксперыментальна распрацаваны ўсе кампаненты сінтэзатара маўлення па тэксце для інтэрнэт-платформ: тэкставы, прасадыхны, фанетычны, алафонны і акустычны. У хуткім часе стане магчымым, каб дадзены кампаненты маглі выкарыстоўвацца як паасобку, так і быць інтэграванымі ў любой камбінацыі ў любую інтэлектуальную сістэму. Таксама важна зазначыць, што распрацаваны сінтэзатар маўлення па тэксце для Інтэрнэту на мове праграмавання PHP характарызуецца крос-платформавасцю. Менавіта таму шматлікія сайты-будаўнікі змогуць ужыць гэты сінтэзатар без выкарыстання дадатковых тэхналогій, адрозных ад мовы напісання іх сайтаў.

6. Падзякі

Аўтары выказваюць падзяку Алене Скопінай (skelena777@gmail.com) за філалагічныя праўкі тэкста даклада.

БІБЛІАГРАФІЧНЫ СПІС

[Лобанов и др., 2008] Лобанов, Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник // Минск : Белорусская наука, 2008. – 344 с.

[Barker-Plummer, 1995] Barker-Plummer, D Turing Machines [Electronic resource]. – 1995. – Mode of access : <http://plato.stanford.edu/entries/turing-machine/>. – Date of access : 01.07.2012.

[Internet World Stats, 2012a] Top Internet Usage [Electronic resource]. – 2012. – Mode of access : <http://www.internetworldstats.com/top20.htm>. – Date of access : 01.07.2012.

[Internet World Stats, 2012b] Europe Internet Usage Stats Facebook and Population Statistics [Electronic resource]. – 2012. – Mode of access : <http://www.internetworldstats.com/stats4.htm#europe>. – Date of access : 01.07.2012.

[IT.TUT.BY, 2012] В Беларуси более 10 млн абонентов GSM-сетей [Электронный ресурс]. – 2010. – Режим доступа : <http://it.tut.by/202331>. – Дата доступа : 01.07.2012.

[Acapela Group, 2012] Acapela Group [Electronic resource]. – 2012. – Mode of access : <http://www.acapela-group.com>. – Date of access : 16.03.2012.

[Nuance Communications, 2002] Nuance company [Electronic resource]. – 2002. – Mode of access : <http://www.nuance.com>. – Date of access : 12.03.2012.

[Mihai Preda, 2010] How to work around Android's 24 MB memory limit [Electronic resource]. – 2010. – Mode of access : <http://blog.java.org/how-to-work-around-androids-24-mb-memory-limit/>. – Date of access : 01.12.2012.

[Центр Речевых Технологий, 2001] Технологии [Электронный ресурс]. – 2001. – Режим доступа : <http://www.speechpro.ru/technologies>. – Дата доступа : 01.07.2012.

[Dutoit, 1997] Dutoit, T. An Introduction to text-to-speech synthesis / T. Dutoit. – Kluwer Academic Publishers, 1997. – 286 p.

[Гецэвіч і інш., 2011] Гецэвіч, Ю.С. Сінтэзатар беларускага і рускага маўлення па тэксце для стацыянарных і мабільных платформ / Ю.С. Гецэвіч, Д.А. Пакладок // Сб. матэрыялаў I Молодежного инновационного форума «Наука и бизнес – 2011» (Минск, 14–18 нояб. 2011 г.). – Минск : Беларуская навука, 2011. – С. 14–15.

[Lobanov et al., 2006] Lobanov, B. Development of multi-voice and multi-language TTS synthesizer (languages: Belarussian, Polish, Russian) / B. Lobanov, L. Tsurulnik // Speech and Computer: proceedings of the 11-th International conference SPECOM'2006, St. Petersburg, Russia, 25-29 June, 2006 / Institute of Informatics and

Automation of RAS, Speech Informatics Group. – St.-Petersburg: Anatolia, 2006. – P. 274–283.

[Цирульник и др., 2010] Цирульник, Л.И. Система синтеза речи по тексту для мобильных телефонов / Л.И. Цирульник, Д.А. Покладок // Речевые технологии – 2010 – № 1 – С. 81–90.

[Ramachandran, 2010] Sreeram Ramachandran Web metrics: Size and number of resources / Google developers [Electronic resource]. – 2010. – Mode of access : <https://developers.google.com/speed/articles/web-metrics>. – Date of access : 14.07.2012.

[Цирульник и др., 2012] Цирульник, Л.И. Грамматический словарь и правила определения словесного ударения для синтеза речи по тексту на мобильных устройствах / Л.И. Цирульник, Д.А. Покладок // Информатика. – 2012. – № 2. – С. 61–68.

[Гецэвіч і інш., 2012] Гецэвіч, Ю.С. Распрацоўка сінтэзатара беларускага і рускага маўленьняў па тэксьце для мабільных і інтэрнэт-платформаў / Ю.С. Гецэвіч, Д.А. Пакладок, Д.В. Брэк // Развитие информатизации и государственной системы научнотехнической информации (РИНТИ-2012) : доклады XI Международной конференции (Минск, 15 ноября 2012 г.). – Минск : ОИПИ НАН Беларуси, 2012. – С. 254–259.

COMPONENTS FOR DIFFERENT PLATFORMS OF TEXT-TO-SPEECH SYNTHESIZERS FOR INTELLECTUAL SYSTEMS

Hetsevich Y.S.*, Pakladok D.A.*, Brek D.V.**

** United Institute of Informatics Problems,
National Academy of Sciences,
Minsk, Republic of Belarus*

**yuras.hetsevich@newman.bas-net.by
dima.pokladok@gmail.com**

*** The Belarusian State University of Informatics
and Radioelectronics,
Minsk, Republic of Belarus*

bartwader@gmail.com

Questions concerning urgency and practical construction of components for the bilingual (Belarusian and Russian) text-to-speech synthesizer for mobile and internet platforms are discussed here. According to the test results the speech synthesizer for the Internet can completely meet the demands for voicing of website pages. It can also be used as a complete component of a speech interface by any external developer by means of the Internet in order to build their own intelligent systems.

INTRODUCTION

Nowadays there is a problem of absence of cross-platform text-to-speech synthesizers which would be able to voice texts from Internet resources. The solution to this problem will help Internet websites to attain a higher level of interactivity. By means of such a speech synthesizer it is possible to expand audience of Internet sites. Blind or visually impaired people will have a possibility to familiarize themselves with the contents of websites. One can do something they need or like, for example, tide up in house, drive a car, and still listen to private messages, articles, news, etc. from the Internet.

The Internet TTS developers will be able to demonstrate the synthesizer in different modes of use

(full or limited access) to a large number of users and target groups (for research works, pupils and students, entrepreneurs and potential clients who want to construct commercial services of voicing texts).

MAIN PART

In the beginning experimental models were built. They confirmed the ability to work with audio files in WAV-format through an interpreted programming language PHP. After that an experimental programming model of basic algorithms of the acoustic processor for the text-to-speech synthesizer in PHP was obtained. The bilingual acoustic processor is available online at the link www.corpus.by/acoustic/form.php. By means of its interface a user can enter a phonetic text (or use already entered), identify a language and process an query to the server. The software script of the acoustic processor receives the query, processes it, and in the end generates a WAV-file. The resulting file can be listened or saved to one's local computer.

Further, all the other parts of the speech synthesizer were developed for the Internet. As a result, any written text consistently passes all stages of processing through text, prosodic, phonetic, allophonic and acoustic processors. On the output a WAV-file with a synthesized speech is created. The result of each stage is visually presented on the Internet page (www.corpus.by/tts).

CONCLUSION

Thus, at present all the components (text, prosodic, phonetic, allophonic and acoustic) of the text-to-speech synthesizer for online platforms have been experimentally developed. Soon it will become possible to use these components separately and to integrate them in any combinations into any intelligent system.

Another result of the work on development of the text-to-speech synthesizer for the Internet has been the emergence of the Internet service, which shows all the stages of converting a written text into an audio file. By means of this service it is possible to improve the quality of speech synthesis by finding and correcting errors at any stage of the process. Obviously, further work will enable to voice web-pages, articles, news and emails at a high level.