

**РАЗДЕЛ 6  
АКТУАЛЬНЫЕ НАПРАВЛЕНИЯ  
ПРИКЛАДНОЙ ЛИНГВИСТИКИ**

**SECTION 6  
TOPICAL ISSUES OF APPLIED LINGUISTICS**

**ПЕРСПЕКТИВЫ РАЗВИТТЯ АЎТАМАТЫЧНАЙ АПРАЦОЎКІ  
ВУСНАГА МАЎЛЕННЯ**

**PROSPECTS FOR AUTOMATIC PROCESSING  
OF ORAL SPEECH**

**Я.С. Зяноўка<sup>1)</sup>, М.В. Супрунчук<sup>2)</sup>, Ю.С. Гецевіч<sup>3)</sup>  
Ya. Zianouka<sup>1)</sup>, M. Suprunchuk<sup>2)</sup>, Yu. Hetsevich<sup>3)</sup>**

АІП НАН Беларусі  
Мінск, Рэспубліка Беларусь

UIIP of NASB  
Minsk, Belarus

*e-mails:* <sup>1)</sup>evgeniakacan@gmail.com, <sup>2)</sup>ms@philology.by,  
<sup>3)</sup>yuras.hetsevich@gmail.com

Дадзены артыкул прысвечаны праблеме аўтаматычнай апрацоўкі інтанацыйных асаблівасцей маўлення. Апісана мэтазгоднасць правядзення тэарэтычных і прыкладных даследаванняў у галіне інтаналогіі. Прыведзены асноўныя канферэнцыі, асацыяцыі і арганізацыі, праца якіх накіравана на пошук і рашэнне актуальных задач прасадычнай апрацоўкі маўлення. Вызначаны сістэмы сінтэзу маўлення (у тым ліку і для беларускай мовы), якасць агучвання тэкставай інфармацыі якіх дасягнула высокага ўзроўню

This article is devoted to the problem of automatic intonation processing. The expediency of conducting theoretical and applied research in the field of intonology is described. Main conferences, associations and organizations whose work is aimed at finding and solving urgent tasks of prosodic processing are presented. Speech synthesis systems (including for the Belarusian language) with high quality voicing are identified.

*Ключавыя словы:* інтанацыя; даследаванне; аўтаматычная апрацоўка; сістэмы маўленчых зносін; сістэмы сінтэзу маўлення.

*Keywords:* intonation; research; automatic processing; speech communication systems; text-to-speech systems.

*Уводзіны.* Развіццё вылічальнай тэхнікі і яе паўсюднае ўкараненне ў чалавека-машынных сістэмах кіравання актуалізуюць арганізацыю зносін чалавека і камп'ютара ў адной з найбольш зручных для чалавека

формаў – у форме маўленчага дыялогу на натуральнай мове. Зручнасць, натуральнасць і прастата зносін не патрабуе спецыяльнай падрыхтоўкі, што істотна пашырае кола патэнцыйных карыстальнікаў вылічальных сістэм і павышае эфектыўнасць іх прымянення.

Стварэнне сістэм маўленчых зносін з электронна-вылічальнымі машынамі (ЭВМ) патрабуе вырашэння дзвюх асноўных праблем: распрацоўкі якасных сістэм аўтаматычнага сінтэзу і сістэм аўтаматычнага аналізу і распазнання маўлення. Гэтыя задачы яшчэ не вырашаны ні для адной з моў на высокім узроўні, нягледзячы на намаганні многіх вучоных розных краін. Даследчыкі сутыкаюцца з прынцыповымі цяжкасцямі, звязанымі са стварэннем інтэлектуальнай сістэмы, якая мадэлюе чытанне чалавекам адвольных тэкстаў. Варта адзначыць, што далёка не ўсе вынікі, атрыманыя для адной мовы, падыходзяць пры распрацоўцы падобных механізмаў на іншых мовах. Сістэмы кожнай мовы ўнікальныя. Неабходна канкрэтна-моўная распрацоўка фанетычных і прасадыхных правілаў арганізацыі маўлення для наладжвання дыялогу паміж чалавекам і ЭВМ і штучнага маўлення ў цэлым.

*Асноўная частка.* На сённяшні дзень праблема прыкладнога даследавання інтанацыі разглядаецца з розных бакоў: гэта тэарытычныя і практычныя аспекты даследавання маўлення. Часцей за ўсё тэарытычныя веды і напрацоўкі, якія патрабуюць грунтоўных расшэнняў, асвятляюцца на розных міжнародна-навуковых канферэнцыях, напрыклад *Speech Prosody*, якая праходзіць раз у два гады сярод зацікаўленых спецыялістаў (*SProSIG*) Міжнароднай асацыяцыі маўленчых зносін (*ISCA*). Мэтай асацыяцыі з'яўляецца садзейнічанне ў міжнародным кантэксце ва ўсім свеце дзейнасці і абмену ведамі ва ўсіх галінах, звязаных з навукай і тэхналогіяй маўленчай камунікацыі. Асацыяцыя прызначана для ўсіх асоб і ўстаноў, зацікаўленых у фундаментальных даследаваннях і тэхналагічных распрацоўках. Іх мэтай з'яўляецца апісанне, тлумачэнне і прайграванне розных аспектаў чалавечых зносін з дапамогай маўлення, а менавіта лінгвістыка, у прыватнасці, фанетыка, распазнаванне і сінтэз маўлення, дапамога ў медыцынскай дыягностыцы паталогій голасу. Таксама трэба адзначыць міжнародную арганізацыю *IEEE Signal Processing Society*, мэтай якой з'яўляецца папулярызацыя і распаўсюджванне сучаснай навуковай інфармацыі і рэсурсаў; інфармаванне грамадства па праблемах апрацоўкі сігналаў; прадастаўленне людзям магчымасці ўзаемадзейнічаць і абменьвацца ідэямі ў рамках апрацоўкі сігналаў. Такія канферэнцыі як *Дыялог*, *Computational Linguistics and Intellectual Technologies*, *ICAFE*, *ICA*, *Eurospeech*, *Interspeech* і інш. прадстаўляюць

найбуйнейшыя ў свеце дыскусійныя пляцоўкі, прысвечаныя апрацоўцы маўлення, з удзелам больш за 1000 удзельнікаў і больш за 600 дакладаў штогод. На канферэнцыях асабліва ўвага надаецца міждысцыплінарным падходам. Кожная канферэнцыя забяспечвае шырокі падыход да пытанняў апрацоўкі маўлення праз вусныя і стэндавыя выступы, пленарныя даклады сусветна вядомых экспертаў, навучальныя дапаможнікі, спецыяльныя сесіі і выставы.

У прыкладным аспекце інтанацыя разглядаецца як адзін з кампанентаў розных сістэм: інтанацыйны блок у складзе камп'ютарна-навучальных сістэм (*PROSODIA, MOMEL, ProZed, INTSINT, a prosody editor for linguists* і інш.); інтанацыйны і прасадычны працэсары ў сістэмах распазнавання і сінтэзу маўлення; інтанацыя ў дадатках для стварэння, візуалізацыі і пошуку па маўленчых фрагментах (*SPPAS*).

Існуе даволі шмат камерцыйных і бясплатных сінтэзаваных галасоў у сістэмах сінтэзу маўлення (CCM), якія працуюць з рознымі аперацыйнымі сістэмамі. У іх складзе прасадычны працэсар адказвае за аўтаматычную працэдуру фарміравання бягучых контураў мелодыі, сілы гуку, фанемнай працягласці і працягласці паўзаў на аснове аналізу пэўных уласцівасцяў уваходнага тэксту і яго прасадычнай разметкі. Найбольш вядомымі з'яўляюцца камерцыйныя сістэмы, у якіх для праслухоўвання і ацэнкі сінтэзу маўлення даступныя толькі анлайн дэма-версіі з абмежаваннямі. Гэта *Acapela Group* (Бельгія, Францыя, Швецыя), *Apple* (ЗША), *Google Translate* (ЗША) *iSpeech* (ЗША), *Ivona TTS* (Польшча, ЗША), *Mary TTS* (Германія), *Microsoft* (ЗША), *Nuance Loquendo* (Італія, ЗША), *Nuance Vocalizer* (ЗША), *ReadSpeaker, Svox* (Швейцарыя, ЗША), *Tingwo* (Швейцарыя), *VitalVoice TTS* (Расія, Санкт-Пецярбург). У большасці сістэм праблема аўтаматычнай сегментацыі маўлення вырашаецца на розным узроўні праз асобныя прыватныя механізмы і алгарытмы, якія знаходзяцца ў закрытым доступе.

Акрамя таго, распаўсюджаны даследаванні тэкставай дэлімітацыі ў параўнальным аналізе розных прасадычных аспектаў. З развіццём лінгвістыкі тэксту і тэорыі дыскурсу рашэнне прасадычных пытанняў звязана з вывучэннем дыскурсіўных і тэкставых асаблівасцяў выкарыстання інтанацыі. На дадзены момант ствараюцца разнастайныя нацыянальныя і прыватныя корпусы тэкстаў розных моў, на аснове якіх праводзяцца аналітычныя і статыстычныя эксперыменты (*Livonian prosody corpus, Erzya prosody corpus, Moksha prosody corpus, Meadow Mari prosody corpus, COEC, Bàsaa - oral spontané, DIRNDL, Aix-MARSEC database, ACSYNT* і інш.).

Што тычыцца даследаванняў для вызначэння інтанацыйных межаў тэкстаў на рускай мове, трэба адзначыць прыватную кампанію *Цэнтр*

маўленчых тэхналогій, распрацоўшчыка інавацыйных сістэм у сферы тэхналогій сінтэзу і распазнання маўлення, аналізу аўдыя- і відэаінфармацыі, распазнання асоб, галасавой і мультымадальнай біямэтрыі. Многія распрацоўкі і рашэнні кампаніі не маюць аналагаў і выкарыстоўваюцца па ўсім свеце, уключаючы ЗША, Лацінскую Амерыку, Блізкі Усход, Еўропу. Маштаб укараненняў дазваляе казаць аб рэальным тэхналагічным лідэрстве ў галіне маўленчых тэхналогій і бімадальнай біямэтрыі. Іх воблака *MPT (SpeechPro)* – гэта інтэрнэт-сэрвіс, які дазваляе агучваць сінтэзаваным голасам любую тэкставую інфармацыю. На сайце [cp.speechpro.com/service/tts](http://cp.speechpro.com/service/tts) даступны дэма-сінтэз маўлення рознымі галасамі, дзе можна ацаніць якасць сінтэзу. На сённяшні дзень даступна 8 галасоў (жаночых і мужчынскіх), якія размаўляюць на трох мовах (руская, англійская, казахская). Сістэма гарантуе пісьменную расстаноўку націскаў, правільнае чытанне скарачэнняў, лікаў, абрэвіятур. Прасадыхны аспект рэалізуецца на аснове пунктуацыі, праз якую і адбываецца дэлімітацыя тэксту на сінтагмы. Падобны прынцып рэалізаваны ў праграме рускамоўнага сінтэзу 'Мультифон' [1], у сістэме сінтэзу маўлення па тэксце 'Aramar' і 'Vital Voice', 'Агафон', 'Агафія' і інш.

Таксама варта адзначыць кампанію *RHVoice Lab*, якая распрацавала *CCM RHVoice*. Гэта адзіны расійскі сінтэзатар маўлення з адкрытым зыходным кодам, выкарыстанне якога даступна на камп'ютары ці тэлефоне любога карыстальніка з абмежаванымі магчымасцямі зроку. Стварэнне новых галасоў дазволіць выбіраць той ці іншы голас для асобнай задачы і для кожнага чалавека. *RHVoice* выкарыстоўвае статыстычныя параметрычны сінтэз. Ён абапіраецца на існуючыя маўленчыя тэхналогіі з адкрытым зыходным кодам (перш за ўсё HTS і звязаныя кампаненты). Галасы ствараюцца на аснове запісаў натуральнай чалавечай гаворкі. Яны кампактныя, паколькі на камп'ютары карыстальнікаў захоўваюцца толькі статыстычныя мадэлі. І хоць агучыць не хапае натуральнасці тых сінтэзатараў, якія генеруюць маўленне, камбінуючы фрагменты саміх запісаў, прадстаўленыя галасы ўсё ж вельмі разборлівыя і нагадваюць дыктараў, якія запісалі зыходны матэрыял. На дадзены момант *RHVoice* акрамя рускай падтрымлівае амерыканскую англійскую, бразільскую, партугальскую, нямецкую, грузінскую, украінскую і беларускую мовы. Тэарэтычна падтрымка іншых моў можа быць рэалізавана, калі атрымаецца знайсці або стварыць усе неабходныя рэсурсы.

На жаль, у адкрытым доступе адсутнічаюць крыніцы адносна *CCM* для беларускай мовы, акрамя *RHVoice* і распрацовак лабараторыі распазнання і сінтэзу маўлення АПП НАН Беларусі [2]. На сённяшні

дзень супрацоўнікі лабараторыі стварылі тры рэалізацыі сінтэзатар маўлення па тэксце для беларускай мовы. Гэта сістэма сінтэзу маўлення на славянскіх мовах *Мультыфон* [1], рэалізаваная на мове C++, сістэма сінтэзу маўлення для мабільных прыстасаванняў, рэалізаваная на мове Java [3], і інтэрнэт-версія на мове PHP [4]. Акрамя таго, распрацавана і пастаянна дапаўняецца платформа для апрацоўкі тэкставай і гукавой інфармацыі для розных даменаў [www.corpus.by](http://www.corpus.by) [5], якая ўяўляе сабой набор інструментальных сродкаў (сэрвісаў) па апрацоўцы тэксту, маўлення і іншых даных. Платформа прадастаўляе сэрвісы такенізацыі, марфалагічнага аналізу, агучанага электроннага граматычнага слоўніка, пазнакі часціны мовы, лічылніка частот, праверкі арфаграфіі, пераўтварэння тэксту ў маўленне і многіх іншых. Дадзены рэсурс пастаянна абнаўляецца, каб забяспечыць карыстальнікаў наборам інструментаў (паслуг) для апрацоўкі тэксту, голасу і іншых даных (на сённяшні дзень – больш за 70 сэрвісаў). Распрацаваныя сэрвісы групуюцца ў тэматычныя дамены для больш зручнага выкарыстання ў пэўных галінах прымянення. Таксама на платформе размешчаны інструменты па апрацоўцы гукавой інфармацыі: *праверка сегментацыі маўлення, прагназаванне працягласці прамовы, распазнаванне эмоцый, трэніроўка распазнавання маўлення, тэматычнае распазнаванне маўлення* і інш.

Важна адзначыць аналізатар і трэнажор інтанацыі *IntonTrainer* для даследавання і вывучэння інтанацыйных асаблівасцей рускай, кітайскай, нямецкай, беларускай моў, брытанскага і амерыканскага варыянтаў англійскай мовы [6]. *IntonTrainer* – гэта праграма, якая прызначана для вывучэння правільнага вымаўлення і аналізу інтанацыйных фраз. Сістэма заснавана на інтэрактыўных слыхавых і візуальных кантактах з навучэнцам, падчас чаго адбываецца аналіз інтанацыйных фраз, параўнанне меладычных партрэтаў эталона і фразы, вымаўленай вучнем, і іх супастаўленне паміж сабой. Акрамя таго, даецца ацэнка інтанацыйнага падабенства эталоннай і вымаўленай вучнем фраз. Дадзены комплекс таксама распрацаваны супрацоўнікамі лабараторыі распазнавання і сінтэзу маўлення АПІ НАН Беларусі як дапаможнік у засваенні інтанацыі роднай і замежнай моў. Акрамя шасці модуляў (рускай, беларускай, кітайскай, нямецкай моў, брытанскага і амерыканскага варыянтаў англійскай мовы) распрацаваны два модулі па вывучэнні інтанацыі эмацыйных выказванняў рускай і англійскай моў і інтанацыі спеваў рускай мовы. Трэнажор знаходзіцца ў адкрытым доступе і мае бясплатны кантэнт, з якім можна пазнаёміцца на афіцыйным сайце праграмы. На сённяшні дзень гэта адзіная бясплатная праграма, якая прапаноўвае вывучэнне адразу шасці моў.

*Заклучэнне.* Такім чынам, на сённяшні дзень застаецца актуальнай задача вывучэння і даследавання інтанацыйных асаблівасцей выказвання ў розных мовах і, як вынік, аўтаматычнай апацоўкі прасадыхных характарыстык маўлення. Вывучэнне праблемы якаснага агучвання інфармацыі сістэмамі сінтэзу маўлення актуальнае як у практычным плане стварэння ўніверсальных сістэм маўленчага ўводу і вываду інфармацыі з камп'ютара, так і ў тэарэтычным плане стварэння інтэлектуальных персаналізаваных мадэляў чытання тэкстаў розных стыляў і зместу. Вышэйразгледжаныя напрацоўкі і рэсурсы з'яўляюцца сведчаннем зацікаўленасці навуковага кола ў вырашэнні прыкладных задач і падставай для працягу тэарэтычных і практычных даследаванняў у дадзеным накірунку.

### **Бібліяграфічныя спасылкі**

1. Лобанов Б.М. «Мультифон» – система персонализированного синтеза речи по тексту на славянских языках // *Лингвистическая полифония: сборник статей в честь юбилея профессора Р.К. Потаповой* / Отв. ред. чл.-корр. РАН В.А. Виноградов. М.: Языки славянских культур, 2007.
2. Лабораторыя распазнавання і сінтэзу маўлення [Электронны рэсурс]. 2022. URL: <http://ssrlab.by/> (дата звароту: 11.07.2017).
3. Цирульник Л.И., Покладок Д.А. Система синтеза речи по тексту для мобильных телефонов // *Речевые технологии*. 2010. № 1.
4. Сінтэзатар маўлення па тэксце [Электронны рэсурс]. 2022. URL: <http://corpus.by/TextToSpeechSynthesizer/?lang=be>. (дата звароту: 29.03.2021).
5. Платформа для апрацоўкі тэкставай і гукавой інфармацыі для розных тэматычных даменаў corpus.by // [Электронны рэсурс]. 2022. URL: <http://corpus.by/> (дата звароту: 12.02.2022).
6. IntonTrainer // IntonTrainer speech analyzer and trainer [Электронны рэсурс]. 2011. URL: <https://intontrainer.by/> (дата звароту: 29.01.2022).