

# РАСПАЗНАВАННЕ БЕЛАРУСКАГА МАЎЛЕННЯ Ў АЎДЫЯФАЙЛАХ І ВІДЭАРЭСУРСАХ

*Малахоткіна Е. А., Тараева К. С.*

*Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі  
г. Мінск, Рэспубліка Беларусь*

*Пятрова Н.Я. – к.філал.н., дацэнт*

У артыкуле аналізуюцца сучасныя падыходы да распазнавання беларускага маўлення, ацэньваецца вартасць вядомых рашэнняў і прапануецца архітэктурна сістэма для апрацоўкі аўдыя- і відэапатокаў. Асаблівая ўвага надаецца выбару мэдэлей, метадам павышэння дакладнасці і практычным кейсам прымянення.

Развіццё тэхналогій аўтаматычнага распазнавання маўлення (Automatic Speech Recognition, ASR) з'яўляецца адным з ключавых напрамкаў у галіне штучнага інтэлекту і лінгвістыкі. У той час як для сусветных моў існуюць высокадакладныя камерцыйныя рашэнні, распазнаванне беларускага маўлення застаецца актуальнай навукова-практычнай задачай. Абмежаванасць маўленчых карпусоў, асаблівасці фаналогіі і нізкі камерцыйны попыт ствараюць унікальныя выклікі для распрацоўкі эфектыўных сістэм [1]. Мэтай дадзенай работы з'яўляецца аналіз сучасных падыходаў,

ацэнка вядомых рашэнняў і распрацоўка архітэктуры спецыялізаванай сістэмы для апрацоўкі аўдыя- і відэапатокаў з беларускай мовай.

Спачатку звернемся да аналізу сучасных падыходаў да распазнавання беларускага маўлення. Вядомыя ASR-сістэмы для рэдкіх моў, да якіх адносіцца і беларуская, грунтуюцца на некалькіх асноўных прынцыпах:

– Мадэлі на аснове глыбокіх нейрасетак. Мы лічым, што яны з’яўляюцца найбольш эфектыўнымі сёння. Гэта такія архітэктуры, як, напрыклад, Wav2Vec 2.0, якія дазваляюць выкарыстоўваць папярэдняе навучанне на вялікіх аб’ёмах няразмежаваных аўдыяданых з выкарыстаннем трансдрукераў [1; 2]. Для беларускай мовы такі падыход асабліва прадуктыўны з-за магчымасці кампенсаваць недахоп размечаных даных.

– Канцэпцыя "Transfer Learning", ці перанос ведаў. Гэты падыход прадугледжвае выкарыстанне мадэлей, папярэдне навучаных на буйных мовах (напрыклад, на рускай або англійскай), з далейшай іх адаптацыяй для беларускай мовы. Даследаванні паказваюць, што такі метады можа значна паскорыць распрацоўку і палепшыць якасць распазнавання [3].

– Гібрыдныя мадэлі. Сістэмы, якія камбінуюць нейрасецевыя падыходы з традыцыйнымі метадамі на аснове прыхаваных мадэлей Маркава (HMM), могуць быць эфектыўнымі для апрацоўкі маўлення з шумамі або ва ўмовах нізкай якасці запісу, што актуальна для архіўных відэафайлаў [4].

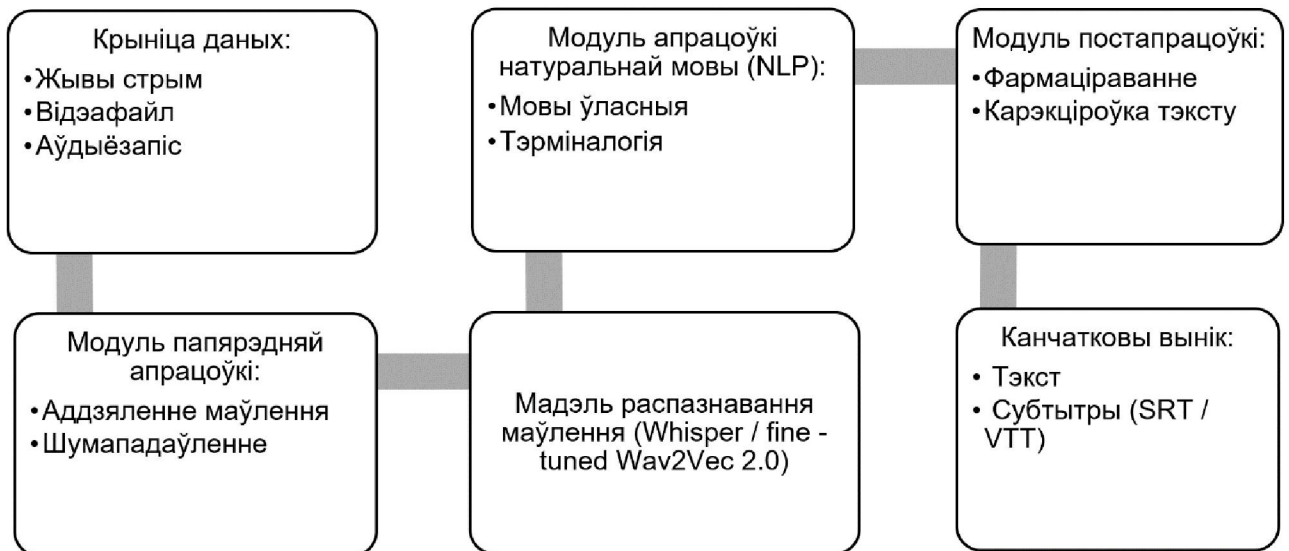
Зараз паспрабуем зрабіць прыблізную ацэнку прыдатнасці вядомых рашэнняў ASR-сістэм для беларускага маўлення. Іх можна падзяліць на тры катэгорыі, прыведзеныя ў Табліцы 1:

Табліца 1 – Параўнальная ацэнка прыдатнасці вядомых рашэнняў ASR-сістэм для беларускай мовы

Катэгорыя рашэння / Прыклад	Узровень падтрымкі беларускай мовы	Тэхналагічная аснова	Прыдатнасць для прамысловага ўжывання	Галоўныя абмежаванні
Міжнародныя воблачныя платформы (Google STT, Yandex SpeechKit)	Адсутная або эксперыментальная	Сучасныя нейрасеткі	Нізкая	Адсутнасць беларускай мовы ў прадукцыйных API
Спецыялізаваныя нацыянальныя праекты (SSR Lab [1; 5])	Поўная (даследчы ўзровень)	Уласныя алгарытмы, магчымасць інтэграцыі адкрытых мадэлей (Wav2Vec)	Сярэдняя	Абмежаваныя вылічальныя рэсурсы, неабходнасць дадатковага наладжвання
Сервісы агучнага тэксту (Natural Reader [2], UniSender)	Поўная (толькі сінтэз маўлення, TTS)	Нейрасеткі	Высокая (для стварэння кантэнту)	Не прызначаны для распазнавання (ASR)

Як відаць з аналізу, для стварэння практычна прыдатнай сістэмы неабходна абавязкова спецыялізаваныя нацыянальныя распрацоўкі або будаваць уласныя рашэнні на аснове мадэлей з адкрытым зыходным кодам, такіх як OpenAI Whisper або адаптаваных версій Wav2Vec 2.0.

На аснове праведзенага аналізу прапануецца модульная архітэктура сістэмы для рэальначасовай і адкладзенай апрацоўкі аўдыя- і відэапатокаў (гл. *Малюнак 1*):



Малюнак 1 – Модульная архітэктара сістэмы для апрацоўкі аўдыя- і відэапатокаў

Звернемся да апісання ключавых модуляў. Модуль папярэдняй апрацоўкі адказвае за ачыстку аўдыясігналу ад шума, нармалізацыю гучнасці, выдзяленне маўленчага патоку з відэафайла і яго сегментацыю. Для гэтага выкарыстоўваюцца алгарытмы на аснове спектральнага аналізу, напрыклад, Spectral Gating. Мадэль распазнавання маўлення – гэта ядро сістэмы. У якасці базавай прапануецца мадэль Whisper ад OpenAI або Wav2Vec 2.0 Large. Гэтыя мадэлі маюць добрую агульную здольнасць да абагульнення і могуць быць эфектыўна адаптаваныя да беларускай мовы шляхам Fine-tuning на спецыяльным карпусе [1; 5; 3]. Fine-tuning павінен уключаць не менш за 1000 гадзін размечанага беларускага маўлення. Модуль апрацоўкі натуральнай мовы (NLP) ажыццяўляе постапрацоўку тэксту: карэкцыю памылак на аснове моўных мадэлей, прывязку да слоўніка беларускай тэрміналогіі для павышэння дакладнасці ў спецыялізаваных даменах (напрыклад, у адукацыйным кантэнце [6, с. 240]), распазнаванне ўласных назваў. Модуль постапрацоўкі фарміруе канчатковы вынік у патрэбным фармаце (тэкставы файл, субтытры SRT/VTT для відэа).

Для павышэння дакладнасці прапануецца камбінаваць некалькі метадаў:

– Data Augmentation: штучнае папаўненне навучальнай выбаркі шляхам дабаўлення розных тыпаў шумоў, змянення хуткасці і тону маўлення;

– Language Model Rescoring: выкарыстанне статыстычнай моўнай мадэлі беларускай мовы для паслядоўнай карэкцыі асноўнай ASR-мадэлі. Гэта асабліва эфектыўна для паляпшэння распазнавання складаных тэрмінаў [7; 8];

– Active Learning: сістэма пазначае няўпэўненыя ўчасткі тэксту для далейшай праўкі чалавекам, што дазваляе таксама паляпшаць мадэль.

Практычныя кейсы ўжывання ASR-мадэлі:

– стварэнне субтытраў для архіўных відэазапісаў Белтэлерадыёкампаніі. Апрацоўка вялікіх аб'ёмаў кантэнту з рознай якасцю запісу;

– інтэлектуальны асістэнт для інклюзіўнай адукацыі, які транскрыбуе лекцыі ў рэальным часе для студэнтаў з парушэннямі слыху [6, с. 239];

– выкарыстанне ў практыцы выкладання замежных моў [9; 10];

– аналітыка вусных медыяперадач для сацыялінгвістычных даследаванняў [11].

Такім чынам, у працы праведзены аналіз сучасных тэхнічных падыходаў да распазнавання беларускага маўлення. Устаноўлена, што, нягледзячы на адсутнасць гатовых камерцыйных рашэнняў, існуюць эфектыўныя метады і мадэлі з адкрытым кодам, прыдатныя для стварэння практычных сістэм. Прапануецца модульная архітэктара, якая абапіраецца на буйныя вядомыя мадэлі (Whisper, Wav2Vec 2.0) з далейшым іх дакладным наладжваннем і спецыялізаванымі модулямі апрацоўкі, што дазволіць стварыць рашэнне для рэальнай і адкладзенай апрацоўкі аўдыя- і відэапатокаў. На нашу думку, найбольш перспектыўнымі кірункамі далейшых даследаванняў з'яўляюцца пытанні пра папаўненне маўленчых карпусоў, распрацоўка спецыялізаваных моўных мадэлей для постапрацоўкі і аптымізацыя гатовых сістэм для разгортвання на абмежаваных вылічальных рэсурсах.

*Спіс выкарыстаных крыніц:*

1. Распрацоўка алгарытмаў дыктара-незалежнага распазнавання беларускага маўлення для робата-самаход [Электронны рэсурс]. – Рэжым доступу: [https://ssrlab.by/wp-content/uploads/2017/11/2017\\_Kazlouskaja-Sybko-Pratasienia-Hiecevic\\_Raspracouka-alharytmau-dyktaraniezalieznaha.pdf](https://ssrlab.by/wp-content/uploads/2017/11/2017_Kazlouskaja-Sybko-Pratasienia-Hiecevic_Raspracouka-alharytmau-dyktaraniezalieznaha.pdf). – Дата доступу: 20.09.2025.
2. Як чытальнікі тэкстаў штучнага інтэлекту могуць дапамагчы тым, хто не ведае англійскай мовы [Электронны рэсурс] // Cudek AI. – Рэжым доступу: <https://www.cudekai.com/be/blogs/>. – Дата доступу: 20.09.2025.
3. Стварэнне базы для распазнавання беларускага маўлення [Электронны рэсурс]. – Рэжым доступу: <https://www.polskieradio.pl/396/7816/artukul/2760862>. – Дата доступу: 20.09.2025.
4. Агрыева, Я. У. Захаванне і распазнаванне вобраза / Я. У. Агрыева [Электронны рэсурс]. – Рэжым доступу: [https://libeldoc.bsuir.by/bitstream/123456789/61259/1/Ahryeva\\_Zahavanne.pdf](https://libeldoc.bsuir.by/bitstream/123456789/61259/1/Ahryeva_Zahavanne.pdf). – Дата доступу: 20.09.2025.
5. Лабараторыя распазнавання маўлення [Электронны рэсурс]. – Рэжым доступу: <https://ssrlab.by/4240>. – Дата доступу: 20.09.2025.
6. Пятрова, Н. Я. Штучны інтэлект у інклюзыйнай адукацыі (на прыкладзе выкладання дысцыпліны «Беларуская мова» у ВНУ) / Н. Я. Пятрова // Непрерывное профессиональное образование лиц с особыми потребностями : сб. ст. V Междунар. науч.-практ. конф., Минск, 14 дек. 2023 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: А. А. Охрименко [и др.]. – Минск, 2023. – С. 239–243.
7. Гецэвіч, М. Метады распазнавання / М. Гецэвіч [Электронны рэсурс]. – Рэжым доступу: [https://libeldoc.bsuir.by/bitstream/123456789/26090/1/Getsev%D1%96ch\\_Metad.PDF](https://libeldoc.bsuir.by/bitstream/123456789/26090/1/Getsev%D1%96ch_Metad.PDF). – Дата доступу: 20.09.2025.
8. Топ-5 сервісаў для агучкі тэксту: агляд і аналіз [Электронны рэсурс]. – Рэжым доступу: [https://timeweb.com/ru/community/articles/top-5-servisov-dlya-ozvuchki-teksta-obzor-i-analiz\\_](https://timeweb.com/ru/community/articles/top-5-servisov-dlya-ozvuchki-teksta-obzor-i-analiz_). – Дата доступу: 20.09.2025.
9. Петрова, Н. Е. Использование интеллектуальных технологий в обучении русскому языку как иностранному в вузе / Н. Е. Петрова // Цифровая гуманитаристика и технологии в образовании (DHTE 2023): сб. статей IV Международной научно-практической конференции, Москва, 16 – 17 ноября 2023 г. / Под ред. В.В. Рубцова, М.Г. Сороковой, Н.П. Радчиковой. – Москва : ФГБОУ ВО МГППУ, 2023. – С. 381–389.
10. Петрова, Н. Е. Применение искусственного интеллекта в преподавании русского языка как иностранного / Н. Е. Петрова // Непрерывная система образования «Школа - университет». Инновации и перспективы : сборник статей VII Международной научно-практической конференции, Минск, 19-20 октября 2023 г. / Белорусский национальный технический университет ; редкол.: О. К. Гусев [и др.]. – Минск : БНТУ, 2023. – С. 213-216.
11. Пытанні распазнавання маўлення [Электронны рэсурс]. – Рэжым доступу: <https://elib.bsuir.by/bitstream/123456789/278223/1/252-257.pdf>. – Дата доступу: 20.09.2025.