

## РАЗВІЦЦЁ МАЎЛЕННЫХ ТЭХНАЛОГІЙ У БЕЛАРУСІ

### Уводзіны

З усяго жывога толькі чалавек валодае дарам маўлення, дзякуючы якому ў яго атрымалася гэтак значна развіць свае інтэлектуальныя здольнасці і, на думку многіх філосафаў, стаць чалавеку чалавекам. Нешта падобнае адбываецца на нашых вачах і з камп'ютарам, які інтэнсіўна авалодвае шырокім спектрам маўленчых тэхналогій: ад кадавання тэлефонных сігналаў да сінтэзу, распазнавання і разумення вуснага маўлення. Цалкам рэальна спадзявацца, што далейшае развіццё чалавека-машынных сістэм маўленчых зносін няўхільна прывядзе да сталага дасканалення моўнай «пісьменнасці» камп'ютараў і, у выніку, да бесперапыннага набліжэння іх інтэлектуальных здольнасцей да чалавечых. Прывядзём дарэчную ў дадзеным кантэксце цытату кнігі акадэміка РАН Вячаслава Усеваладавіча Іванова «Лінгвістыка трэцяга тысячагоддзя»:

***“Мы ўпершыню ў гісторыі віду пачынаем шырока карыстацца тэхнічнымі прыладамі, якія размаўляюць – інакш кажучы, не толькі вырабляць прылады (чым чалавек адрозніваецца ад жывёл), але і навучаць іх нашай мове (чым мы пачынаем адрознівацца ад усіх людзей, якія жылі раней)”.***

Аднак на шляху стварэння практычна запатрабаваных сістэм, якія рэалізуюць маўленчыя тэхналогіі, усё яшчэ неабходна пераадолець нямала цяжкасцей. Характэрны прыклад выказвання Віцэ-прэзідэнта кампаніі Microsoft Г. Сінха: «... Біл Гейтс засмучаны тым, наколькі доўга тэхналогіям распазнавання маўлення даводзіцца "прабіваць" сабе дарогу. Бо карпарацыя ўкладвае грошы ў гэты праект, прынамсі, з 1991 года. ...» (па матэрыялах CNET News).

Сучасныя маўленчыя тэхналогіі ўмоўна можна падзяліць на наступныя віды:

- Распазнаванне маўлення (у т.л. каманд, ключавых слоў, ланцужкоў слоў, спантаннага маўлення).
- Сінтэз маўлення па тэксце (у т.л. шматгаласавога, персаналізаванага, выразнага, мультымадальнага).
- Галасавая біямэтрыя (ідэнтыфікацыя і верыфікацыя асобы па голасе і маўленні).
- Дыягностыка асобных характарыстык і стана моўцы.
- Медыцынская дыягностыка (дыягностыка захворванняў па голасе і маўленні).

У наш час сфера прымянення маўленчых тэхналогій увесь час пашыраецца і закранае ўсё новыя і новыя вобласці чалавечай дзейнасці. Так, напрыклад, у 2012 г. міжнародны аэрапорт Шарамецьева ў Маскве абвясціў пра запуск галасавой сістэмы інфармавання пра статус рэйса. Сэрвіс працуе па нумары +7 (495) 956-46-66. Запуск выклікаў заканамерную цікавасць карыстальнікаў: зручны прыязны інтэрфейс выгодна адрознівае яго ад іншых аўтаматычных сэрвісаў, якія працуюць у айчынных кампаніях.

## **1. Пачатковы этап развіцця маўленчых даследаванняў**

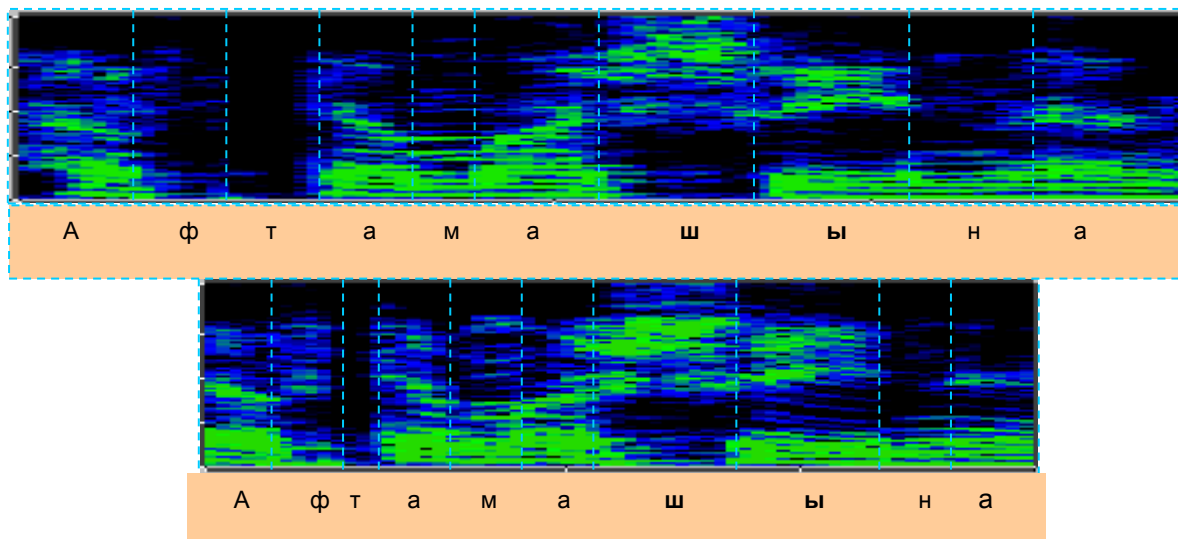
Пачатак сучаснай гісторыі маўленчых даследаванняў у СССР датуецца сярэдзінай 60-х гадоў мінулага стагоддзя, калі ўпершыню пачала працаваць Усесаюзная школа-семінара па аўтаматычным распазнаванні слыхавых выяў (АРСВ), якая збірала ў лепшыя гады да 300 удзельнікаў. Да гэтага ж часу адносіцца і пачатковы этап развіцця маўленчых даследаванняў у Беларусі. У 1965 годзе ў навуковай лабараторыі кафедры радыёпрыёмных прылад Мінскага радыётэхнічнага інстытута пад кіраўніцтвам аўтара гэтага артыкула была арганізавана група даследавання маўленчых сігналаў. У той час у яе ўваходзілі М.П. Дзегцяроў, Б.В. Панчанка, М. Фацееў і інш., якія яшчэ доўгі час, а некаторыя з іх і дагэтуль, працуюць у гэтым кірунку.

Першыя даследаванні групы былі злучаны з распрацоўкай агульных прынцыпаў аналізу маўленчых сігналаў і вылучэнні інфарматыўных

прыкмет, якія дазволілі б прадставіць бесперапынны маўленчы сігнал паслядоўнасцю фанетычных сегментаў. Вынікі гэтых даследаванняў былі абагульнены ў дысертацыі аўтара "Некаторыя пытанні аналізу маўленчых сігналаў", абароненай у 1968 г. у Маскоўскім ДзяржНДІ Радые. Найболей важныя вынікі гэтай працы пазней былі апублікаваны ў аўтарытэтных міжнародных часопісах [1-2]. На базе гэтых даследаванняў упершыню ў СССР было распрацавана адносна простае ўстройства распазнавання маўленчых каманд "СЕЗАМ - 2", якое атрымала ў 1968 г. срэбны медаль ВДНГ СССР. Устройства складалася з двух блокаў: аналізатара прыкмет маўленчага сігнала, такіх як "Галасавы", "Шумны", "Галосны" і інш., і лічыльніка колькасці прыкмет у маўленчай камандзе. Дасягнута досыць высокая надзейнасць распазнавання 20 каманд (уключаючы назвы лічбаў) незалежна ад голасу дыктара, гучнасці і тэмпу вымаўлення. У той жа перыяд часу былі распрацаваны спецыялізаваныя прыборы для эксперыментальна-фанетычных даследаванняў маўлення: аналізатар дынамічных спектраў і інтанограф, з дапамогай якіх у наступныя гады праведзены шматлікія даследаванні ў фанетычных лабараторыях Інстытута мовазнаўства АНБ і Мінскага інстытута замежных моў.

Даследаванні дынамічных спектраў маўлення далі штуршок да развіцця нелінейных метадаў супастаўлення слоў вуснага маўлення, якія распазнаюцца, з іх эталонамі. Спектральныя малюнкi маўлення, у адрозненне ад звычайных візуальных малюнкаў аб'ектаў, могуць падвяргацца некантралюемым нелінейным скажэнням часовай восі. На мал. 1 прыведзены два спектральныя малюнкi слова "аўтамашына", вымаўленых адным і тым жа дыктарам з розным тэмпам маўлення. З малюнка відаць, што пры паскораным тэмпе (ніжняя спектраграма) пры агульным скарачэнні працягласці слова на 30% працягласць гукаў націскага склада "ШЫ" практычна не змянілася, у той час як гукі "Т" і "Н" у ненаціскных складах скараціліся больш чым у 2 разы. З прыведзенага прыкладу відаць, што простага маштабавання спектральных малюнкаў недастаткова для іх

надзейнага распазнавання. Сітуацыя, вобразна кажучы, падобна да той, якая



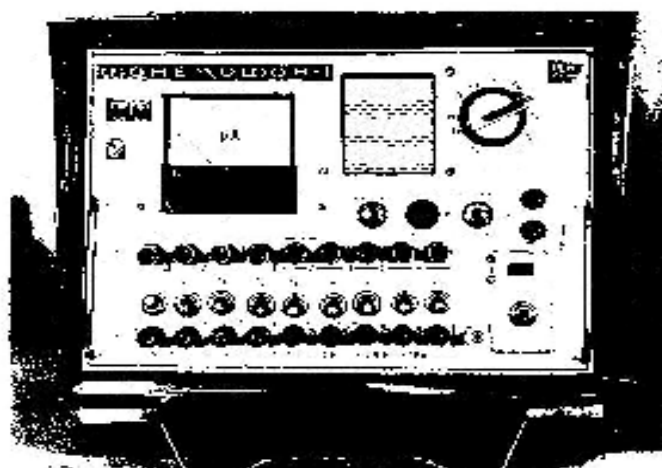
Мал.1. Спектральныя выявы маўлення

магла б паўстаць ва ўмовах "крывога люстэрка" пры распазнаванні зрокавых вобразаў. Рашэнне фундаментальнай праблемы распазнавання маўлення, злучанай з нелінейнымі скажэннямі часавай восі, было прапанавана незалежна і практычна адначасова Г.С. Слуцкерам (Маскоўскі ДзяржНДІ Радые) і Т. Вінцюком (Інстытут кібернетыкі АН УССР) у другой палове 60-х гадоў. Сутнасць прапанаванага рашэння заключалася ў знаходжанні метадам дынамічнага праграмавання (ДП-метада) аптымальнага шляху на матрыцы лакальных адлегласцей паміж часавымі адлікамі вектараў распазнавальнага і эталоннага спектраў. У 1969 годзе аўтарам сумесна з супрацоўнікамі Маскоўскага ДзяржНДІ Радые быў апублікаваны артыкул [3], у якім дадзена далейшае развіццё ДП-метаду для выключна важнага практычнага выпадку, калі межы распазнавальнага слова невядомыя, г.зн. для рашэння задачы выяўлення і распазнавання гукаспалучэнняў у бесперапынным маўленчым сігнале. ДП-метада атрымаў шырокае прызнанне замежных даследчыкаў і, нароўні з метадам схаваных Маркаўскіх мадэляў (СММ), дагэтуль выкарыстоўваецца ў сучасных сістэмах распазнавання маўлення.

Да канца 60-х гадоў адносіцца таксама пачатак прац па стварэнні сінтэзатараў маўлення. Стымулам паслужыла ўсведамленне таго, што распрацоўка і даследаванне мадэляў сінтэзу маўлення - гэта прамы шлях да

атрымання больш дэталёвых ведаў пра прыроду адукацыі і ўласцівасцяў маўленчага сігналу, абапіраючыся на якія ў далейшым можна будзе пабудаваць больш дасканалыя алгарытмы аналізу і распазнавання маўлення. Немалаважную ролю ў засваенні сусветнага тэхналагічнага ўзроўню сінтэзу маўлення таго часу адыграла навуковая стажыроўка аўтара гэтага артыкула ў 1970-м годзе ў Лабараторыі прафесара Лорэнца (Эдынбургскі ўніверсітэт), дзе была распрацавана адна з першых фармантных мадэляў сінтэзу маўленчых сігналаў. З дапамогай сінтэзатара гэтай лабараторыі былі ўпершыню атрыманы высакаякасныя ўзоры сінтэзаванага рускага маўлення.

Першая, пакуль яшчэ не цалкам дасканалая мадэль сінтэзатара рускага маўлення па тэксце "ФАНЕМАФОН-1" (гл. мал 2), "загаварыла" напачатку 70-х гг. і поспех у яе стварэнні звязаны, у першую чаргу, з распрацоўкай новых метадаў апаратнай рэалізацыі фармантнага сінтэзу маўленчых сігналаў. Прынцып фармантнага



Мал. 2. Сінтэзатар «Фанемафон-1»

сінтэзу маўленчых сігналаў заснаваны на мадэляванні ўласцівасцей крыніц узбуджэння (галасавога і шумавога) і рэзанансных (фармантных) характарыстык маўленчага апарата чалавека. У выніку эксперыментальных даследаванняў быў створаны поўны набор фармантных "партрэтаў" фанем, які дазволіў упершыню ажыццявіць сінтэз рускага маўлення па адвольным тэксце. Пазней з'явілася палепшаная версія сінтэзатара - "Фанемафон-2", з дадатковым блокам пераўтварэння "фанема - алафон".

## 2. Гісторыя «сярэдніх» гадоў (70-я – 80-я гады)

Вырашальную ролю ў арыентацыі на рашэнне прыкладных задач выкарыстання маўленчых тэхналогій адыграла стварэнне ў 1974 г. Лабараторыі апрацоўкі маўленчых сігналаў у складзе Мінскага аддзела цэнтральнага НДІ сувязі (МГНДІС). У 1976 г. на Усесаюзным семінары АРСВ-9, праведзеным у Мінску, быў упершыню прадэманстраваны прататып аўтаматычнай тэлефоннай даведачнай службы з сінтэзаваным маўленчым адказам, а ўжо з пачатку 80-х гадоў у Мінску працяглы час працавала сістэма аўтаматычнага абзвону даўжнікоў за міжгароднія перамовы (у іх лік часам траплялі і аўтары гэтай распрацоўкі). Да сярэдзіны 80-х гадоў гэтая сістэма была ўкаранёная ў шматлікіх гарадах СССР - ад Брэста да Петрапаўлаўска-Камчацкага. Паспяховаму ўкараненню сінтэзатараў маўлення папярэднічала працяглая праца па ўдасканаленні як якасных паказчыкаў сінтэзаванага маўлення, так і тэхналогіі іх рэалізацыі ў якасці новага класа вонкавых устройстваў ЭВМ. Асноўным недахопам першых мадэляў сінтэзатараў "Фанемафон-1, 2" былі невысокія разборлівасць і якасць сінтэзаванага маўлення, абумоўленыя, у першую чаргу, выкарыстаннем вельмі



Мал. 3. «Фанемафон-3» на выставе ў Жэневе

спрошчаных мадэляў узаемадзеяння гукаў падчас параджэння маўлення (эфектаў каартыкуляцыі і рэдукцыі) і недастаткова прапрацаванай мадэлі інтанавання маўлення па тэксце. У наступнай мадэлі - "Фанемафон-3" былі ўведзены дадатковыя блокі мадэлявання працэсаў артыкуляцыі і інтанавання маўлення, што істотна павялічыла якасныя паказчыкі сінтэзаванага маўлення.

У 1979 г. "ФАНЕМАФОН-3" дэманстравалася на Сусветнай выставе "ТЭЛЕКОМ-79" у Жэневе (гл. мал.3).

Вядомы фантаст Артур Кларк, наведаўшы павільён СССР і азнаёміўшыся з сінтэзатарам маўлення, запісаў у кнігу водгукаў: *"Вы анярэдзілі мае фантазіі з фільма "Касмічная Адысея - 2001", а швейцарская газета "Аглядальнік" апублікавала артыкул: "Зараз рускія вывучаюць замежныя мовы з дапамогай гаворачага камп'ютара"*.

Важную ролю ў стварэнні прамысловых сінтэзатараў маўлення адыграла распрацоўка цалкам лічбавай мадэлі сінтэзатара маўлення "Фанемафон-4", сістэмы распазнавання маўлення "Сезам" і маўленчага тэрмінала - "Марс" (мал. 4).

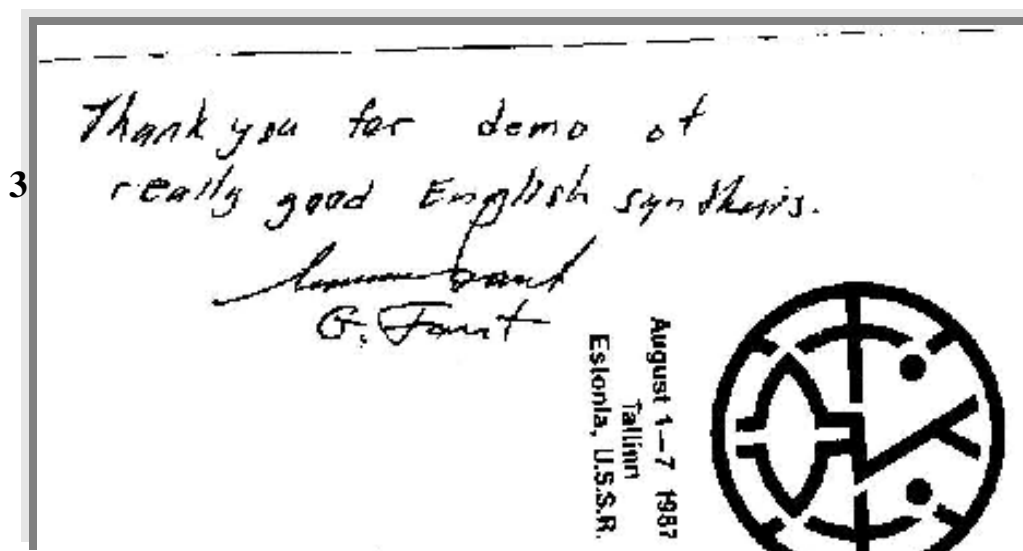


Мал. 4. Маўленчы тэрмінал «МАРС»

Іх серыйны выпуск упершыню ў СССР (1983 г.) быў наладжаны на ПА "Кварц" г. Калінінграда дзякуючы запалу супрацоўнікаў канструктарскага аддзела, які ўзначальваў Валерый Афанасьеў. Ключавую ролю ў іх стварэнні адыгралі распрацоўкі супрацоўнікаў лабараторыі М.Дзегцярова і В.Шатэрніка. У маўленчым тэрмінале "МАРС" упершыню былі інтэграваны функцыі распазнавання і сінтэзу маўлення. У аснову алгарытмаў распазнавання маўлення пакладзены ДП-метады прыняцця славесных рашэнняў на базе набору фармантных прыкмет маўленчага сігналу. Вопытныя ўзоры сістэм "Сезам" і "Марс" былі выкананы на мікрапрацэсарнай аснове і па параметрах прызначэння не саступалі лепшым замежным аналагам таго часу. Арыгінальнасць тэхнічных рашэнняў, выкарыстаных пры стварэнні сістэм "Фанемафон", "Сезам" і "Марс", абаронена шматлікімі аўтарскімі сведчаннямі СССР на вынаходствы.

Да пачатку 1984 г. адносіцца канчатковая фармулёўка, тэарэтычная і эксперыментальная распрацоўка адзінага лінгва-акустычнага падыходу да

рашэння праблемы сінтэзу маўлення па тэксце, яго рэалізацыя ў выглядзе тэхнічных сістэм і практычнае ўкараненне ў складзе аўтаматызаваных сістэм кіравання і сувязі. Вынікі гэтых даследаванняў былі абагульнены ў доктарскай дысертацыі аўтара "Метады аўтаматычнага сінтэзу рускага маўлення па тэксце", абароненай у 1984 г. у Інстытуце электронікі і вылічальнай тэхнікі АН Латвійскай ССР. Пазней атрыманыя вынікі былі адаптаваны для сістэм сінтэзу маўлення на іншых еўрапейскіх мовах. У прыватнасці, да 1987 г. дзякуючы супрацоўніцтву з прафесарам Мінскага інстытута замежных моў Е. Карнеўскай была распрацавана англамоўная версія сінтэзатара [4], якая дэманстравалася на Сусветным кангрэсе фанетычных навук і атрымала высокую адзнаку англамоўных спецыялістаў. Вось факсімільнае водгуку пра гэтую дэманстрацыю аднаго з выдатнейшых у свеце даследчыкаў маўлення Гунара Фанта (гл. мал. 5).



Мал.5. Факсімільнае водгуку Г. Фанта

### 3. Навейшая гісторыя

У 1988 г. на базе лабараторыі МГНДІС у ІТК АНБ была створана лабараторыя распазнавання і сінтэзу маўлення, на пасадку загадчыка якой дырэктарыяй Інстытута быў запрошаны аўтар гэтага артыкула. Як многія яшчэ



добра памятаюць, канец 1980-х гадоў адзначыўся з'яўленнем першых персанальных камп'ютараў, таму натуральна, што ў планах прац лабараторыі з'явілася тэматыка, звязаная з абсталяваннем ПК сістэмай маўленчага ўводу-вываду інфармацыі. Фармантны метадад, які доўгі час меў ключавую ролю ў сістэмах сінтэзу маўлення па тэксце, не падыходзіў для гэтай мэты з-за неабходнасці вялікага аб'ёму вылічэнняў у рэальным часе, што было недасяжна для ПК таго часу. У канцы 80-х гадоў быў прапанаваны новы мікрахвалевы (МХ) метадад сінтэзу маўленчых сігналаў [5], у якім замест вылічэнняў фармантных ваганняў (гукавых хваль) выкарыстоўваўся падрыхтаваны загадзя набор мікрахваль натуральнага маўленчага сігнала. Набор мікрахваль складаўся з адрэзкаў сігнала, роўных працягласці перыяду, а іх колькасць, неабходная для генерацыі любога гуку маўлення, дасягала некалькіх соцень. МХ-метадад увасоблены Аляксандрам Івановым у сінтэзатары "ФАНЕМАФОН-5" у выглядзе спецыялізаванага ПЗ сінтэзатара, арыентаванага на працу з унутранай гукавой платай або з аўтаномным устройствам, якое падключаецца да порта RS-232 (гл. мал. 6). Дзіўная для шматлікіх, кампактнасць яго ПЗ (усяго 64К байт) дазволіла абсталяваць сінтэзам маўлення ўжо першыя IBM PC-XT і нават айчынныя ПК ЕС1840. Сінтэзатар маўлення быў запатрабаваны ў шматлікіх практычных прыкладаннях, але асабліва шырока, і дагэтуль яшчэ, ён выкарыстоўваецца невідучымі карыстальнікамі ПК (больш за сотню камплектаў спецыялізаваных апаратна-праграмных прадуктаў для невідучых былі створаны і распаўсюджаны супрацоўнікам лабараторыі Георгіем Лосікам у Расіі, Украіне і Беларусі ў першай палове 90-х гадоў). Дагэтуль яшчэ яго даволі разборлівае гучанне можна пачуць у Інтэрнэце, ці набыўшы на рынку CD ROM "Гаворачая мыш". У далейшым на аснове МХ-метададу распрацаваны версіі для чэшскай і польскай моў (прыкладна за 3 месяцы знаходжання ў краіне па запрашэнні замоўцаў), а таксама аўтаномны аднаплатны модуль сінтэзу маўлення, украінска-моўная версія якога доўгі час працавала на лініі кіеўскага метро.

Складаная эканамічная сітуацыя, якая склалася ў краіне ў сярэдзіне 90-х гадоў, прымушала шукаць крыніцы фінансавання даследаванняў за мяжой, у першую чаргу, у форме сумесных міжнародных праектаў. Першым з іх стаў міжнародны праект: *"Двухмоўны сінтэз маўлення - нямецкі / рускі"* (1995-1996), які выконваўся сумесна з Дрэздэнскім тэхнічным універсітэтам і які фінансаваўся Германскім навуковым фондам FTU Karlsruhe.

Наступным быў праект *"Аналіз натуральнай мовы і маўлення"* (1996-1997), які выконваўся сумесна з Саарбрукенскім універсітэтам (Германія), Манчэсцэрскім універсітэтам (Вялікабрытанія) і Інстытутам праблем перадачы інфармацыі (Расія) і які фінансаваўся еўрапейскім фондам INTAS. Удзел у гэтым праекце быў звязаны з далейшым развіццём мадэляў сінтэзу маўлення шляхам іх інтэграцыі ў сістэмы апрацоўкі натуральнай мовы метадамі камп'ютарнай лінгвістыкі.

Важную ролю ў інтэграцыі беларускіх даследчыкаў у вобласці лінгвістыкі і маўлення ў еўрапейскую супольнасць стаў удзел у міжнародным праекце *"Развіццё Еўрапейскай камп'ютарнай сеткі па лінгвістыцы і маўленні ва ўсходнім кірунку"* (1997-1998), які фінансаваўся еўрапейскім фондам COPERNICUS. З 1998 г. лабараторыя распазнавання і сінтэзу маўлення АПП з'яўляецца каардынацыйным цэнтрам гэтай сеткі ў Беларусі.

Акрамя еўрапейскіх навуковых арганізацый цікавасць да супрацоўніцтва з лабараторыяй выявілі ў гэтыя гады і некаторыя камерцыйныя арганізацыі. У 1996 г. французская фірма "Секстант Авіёнік" прапанавала рэалізаваць навуковы праект **"Распазнаванне маўленчых каманд ва ўмовах шумоў у кабінце самалёта"**. Праект фінансаваўся фондам Міністэрства абароны Францыі. Нягледзячы на выключную складанасць пастаўленай задачы, праект быў паспяхова выкананы і ў 1997 г. прыняты замоўцам. Асноўныя навуковыя вынікі гэтага праекта выкладзены ў [6]. Іншай камерцыйнай распрацоўкай стаў праект **"Інтэлектуальны тэлефонны аўтасакратар"**, які выконваўся з 1997 г. па дамове з фірмай NovCom NV (ЗША). Сутнасць праекта складалася ў рашэнні задачы распазнавання імёнаў абанентаў, якія

вымаўляюцца па тэлефоне і іншай службовай інфармацыі з тым, каб сістэма змагла выконваць функцыі тэлефоннага аўтасакратара. Праект завершаны ў 1999 г., а яго асноўныя навуковыя вынікі апублікаваны ў [7]. У выкананні гэтых міжнародных праектаў ключавую ролю адыгралі супрацоўнікі лабараторыі Таццяна Леўкоўская, Аляксандр Іваноў і Андрэй Кубашын. Працы па гэтых праектах з'явіліся стымулам да рэанімацыі і далейшаму развіццю алгарытмаў распазнавання маўлення, прапанаваных яшчэ ў 1969 г. [3].

#### **4. Сучасныя даследаванні і распрацоўкі**

Пачыная з 2000 г. навуковыя інтарэсы Лабараторыі ад праблем распазнавання зноў пераходзяць да праблем сінтэзу маўлення. Распрацоўваецца новая мультыхвалевае мадэль сінтэзу маўлення па тэксце, якая забяспечвае высокую якасць і персаналізацыю сінтэзаванага маўлення. Важкі ўклад у стварэнне на аснове гэтай мадэлі сінтэзатара маўлення "Мультыфон" унеслі супрацоўнікі лабараторыі Віталь Кісялёў і Зміцер Жадзінец. Адначасова тэрэтычна абгрунтоўваецца і развіваецца новы навуковы кірунак у маўленчых даследаваннях - камп'ютарнае кланаванне маўлення - для рашэння задач высакаякаснага сінтэзу маўлення па тэксце з максімальна магчымым набліжэннем да голасу і манеры чытання пэўнага чалавека; ствараюцца тэхналогія і адмысловыя праграмныя сродкі кланавання акустычных, фанетычных і інтанацыйных характарыстык маўлення чалавека. Істотны ўклад у стварэнне праграмных сродкаў кланавання маўлення - "Фонакланатар" і "Інтакланатар" унеслі супрацоўнікі лабараторыі Андрэй Давыдаў і Лілія Цырульнік.

Атрыманыя напачатку 2000-х гадоў тэрэтычныя і эксперыментальныя вынікі дазволілі ажыццявіць шэраг сумесных прыкладных праектаў.

У 2005 - 2007 гг. выкананы міжнародны INTAS-праект *"Распрацоўка шматгаласавой і шматмоўнай сістэмы сінтэзу і распазнавання маўлення (мовы: беларуская, польская, руская)"*. Удзельнікі праекта: Беларусь, Германія, Польшча, Расія. У выніку выканання праекта былі закладзены

асновы стварэння шматмоўнай сістэмы аўтаматычнага перакладу вуснага маўлення для 3-х славянскіх моў.

У 2006 - 2007 гг. па заказе ТАА "Інвасэрвіс" (г. Мінск) распрацавана сістэма **"Чытальнік"**, арыентаваная на стварэнне і рэдагаванне "гаворачых" падручнікаў для невідущых вучняў спецшкол Беларусі.

У 2007 - 2009 гг. у межах БРФФД-РФФД выкананы праект **"Распрацоўка мультымедынай сістэмы аўдыявізуальнага сінтэзу маўлення"** - так званай, "персаналізаванай гаворачай галавы" (гл. мал. 7), якая дазволіла істотна палепшыць камфортнасць успрымання маўлення ў шэрагу практычных прыкладанняў (у прыватнасці, у сістэмах дыстанцыйнага навучання).



Мал. 7. «Гаворачая галава»

У 2009 - 2011 гг. у межах БРФФД-РФФД выкананы праект **"Сінтэз выразнага маўлення на базе глыбокага сінтаксічнага аналізу чытаемага тэксту"**. У выніку выканання праекта ўпершыню створана двухмоўная сістэма сінтэзу выразнага маўлення на беларускай і рускай мовах, якая



Мал. 8. Прэм'ер-міністр Беларусі **Міхаіл МЯСНІКОВІЧ** і віцэ-прэзідэнт Расійскай акадэміі навук **Жорэс АЛФЁРАЎ** слухаюць расповед маладога навукоўца **Юрася ГЕЦЭВІЧА** пра распрацоўкі сінтэза беларускага і рускага маўлення. ("Рэспубліка" за 16.11.2011 г.)

дэманстравалася на шэрагу навукова-тэхнічных выстаў (гл. мал. 8).

У гэтыя ж гады праведзены вялікі аб'ём даследаванняў, накіраваных на распрацоўку беларускамоўнай мадэлі сінтэзу маўлення. Асноўную ролю ў

стварэнні сінтэзатара беларускага маўлення адыгралі даследаванні і распрацоўкі Ю.С. Гецэвіча, які зараз выконвае абавязкі загадчыка лабараторыі распазнавання і сінтэзу маўлення.

У 2011 - 2012 гг. па заказе Якуцкага дзяржаўнага ўніверсітэта была распрацавана ўпершыню ў свеце сістэма сінтэзу якуцкага маўлення. Сінтэзатар якуцкага маўлення выкарыстоўваецца ў сістэмах камп'ютарнага навучання якуцкай мове, а таксама ў сістэмах агучвання тэкстаў для сляпых.

Ніжэй прыведзены два прыклады сучасных распрацовак лабараторыі.

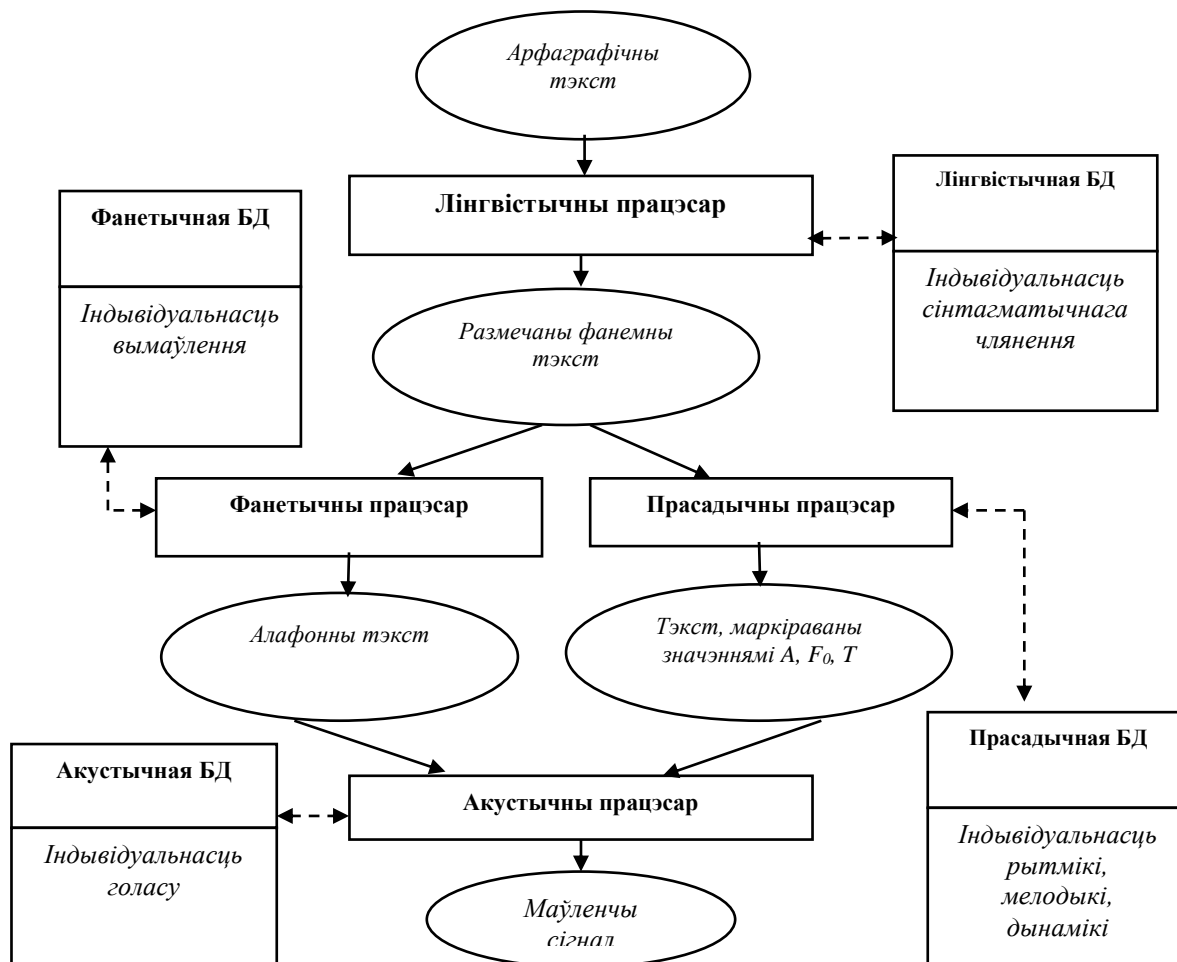
#### ***4. 1. Камп'ютарнае кланаванне персанальнага голасу і дыкцыі***

Шматгадовыя даследаванні, праведзеныя ў ХХ стагоддзі, дазволілі стварыць сінтэзатары, якія забяспечваюць якасць і выразнасць маўлення, даволі прыдатныя для шырокага спектру практычных прыкладанняў. Аднак, нягледзячы на ўсе высілкі, сінтэзаванае маўленне заставалася яшчэ далёкім па якасці ад натуральнага і валодала пазнавальным машынным акцэнтам. Прычынай гэтаму былі не столькі ўзровень нашых ведаў пра працэсы параджэння маўлення і пра фанетыку, колькі недахоп вылічальных рэсурсаў камп'ютараў таго часу. Цяпер мы можам не абмяжоўваць сябе ні аб'ёмам аператыўнай і дыскавай памяці, ні запатрабаваным аб'ёмам вылічэнняў, і гэта з'явілася перадумовай да пастаноўкі задачы сінтэзу маўлення па тэксце з максімальна магчымым набліжэннем па гучанні да голасу і манеры чытання пэўнага чалавека.

Такая пастаноўка задачы, хоць і аддалена, нагадвае шырока вядомую біялагічную праблему кланавання. У нашым выпадку, у адрозненне ад класічнай задачы кланавання, робіцца спроба стварэння блізкай копіі, але не біялагічнай, а камп'ютарнай, і не ўсёй істоты ў цэлым (у дадзеным выпадку чалавека), а толькі адной з яго інтэлектуальных функцый: чытанне адвольнага арфаграфічнага тэксту. Пры гэтым ставіцца задача максімальна поўнага захавання персанальных акустычных асаблівасцей голасу,

фанетычных асаблівасцей вымаўлення і акцэнту, а таксама прасадычнай індывідуальнасці маўлення (мелодыка, рытміка, дынаміка).

Агульная структурная схема сінтэзатара персаналізаванага маўлення, на базе якога ажыццяўляецца кланаванне, паказана на мал.9.



Мал. 9. Схема сінтэзатара персаналізаванага маўлення

Сінтэзатар складаецца з чатырох працэсараў: лінгвістычнага, прасадычнага, фанетычнага і акустычнага. Кожны з працэсараў выкарыстоўвае для пераўтварэнняў, якія ім ажыццяўляюцца, спецыялізаваныя базы дадзеных (БД). У гэтых БД закладзены як агульныя моўныя правілы (лінгвістычныя, прасадычныя, фанетычныя, акустычныя), так і правілы, звязаныя з індывідуальнымі асаблівасцямі голасу і маўлення дыктара.

Кланаванне акустычных характарыстык голасу. Персанальныя акустычныя характарыстыкі голасу чалавека абумоўлены мноствам фактараў, такіх як анатамічныя асаблівасці будовы і функцыянавання элементаў маўленчага апарата (гартань, галасавыя звязкі, глотка, поласць рота і інш.), дынамічныя асаблівасці ўзаемадзеяння ваганняў галасавых звязак і рэзанатараў маўленчага апарата, а таксама шмат іншага. Як вядома, спробы імітацыі персанальных характарыстык голасу ў сістэмах "тэкст - маўленне" на аснове мадэлявання фізіялагічных і акустычных працэсаў параджэння маўлення з-за іх надзвычайнай складанасці дагэтуль не прывялі да адчувальных вынікаў. У сувязі з гэтым найболей разумным уяўляецца выкарыстанне адрэзкаў натуральнай маўленчай хвалі ў якасці мінімальнага "генетычнага матэрыялу" для кланавання голасу. У якасці такога адрэзка мэтазгодна абраць алафон як найболей вывучаную фанетычную субстанцыю, прычым абмежаваны набор алафонаў здольны забяспечыць параджэнне вуснага маўлення адвольнага зместу. Пры гэтым гукавая хваля ўтрымлівае ў сабе ўсе істотныя персанальныя асаблівасці параджэння голасу, якія выяўляюцца ў дадзеным пэўным алафоне.

Кланаванне персанальных фанетычных асаблівасцей вымаўлення. У адрозненні ад персанальных акустычных характарыстык голасу, абумоўленых, у асноўным, статычнымі параметрамі маўленчага апарата, фанетычныя асаблівасці вымаўлення абумоўлены галоўным чынам дынамікай артыкуляцыйных рухаў, якія ажыццяўляюцца падчас маўлення. Уласцівыя дадзенаму індывіду хуткасць артыкуляцыйных рухаў, індывідуальныя асаблівасці артыкуляцыі таго ці іншага гуку (напрыклад /P/), рэгіянальныя ці замежныя акцэнт абумоўліваюць узнікненне своеасаблівых пазіцыйных і камбінаторных адценняў фанем і ствараюць унікальную сістэму алафонаў. Такім чынам, паспяховае кланаванне персанальных фанетычных асаблівасцей вымаўлення можа быць дасягнута шляхам імітацыі асаблівасцей фанемна-алафоннага пераўтварэння, уласцівага дадзенаму чалавеку падчас маўлення.

### Кланаванне персанальных прасадыхных характарыстык маўлення.

Комплекс прасадыхных (інтанацыйных) характарыстык маўлення, які ўключае мелодыку, рытміку і энергетыку, задаецца заканамернымі зменамі ў часе частаты асноўнага тону, працягласці гукаў і амплітуды гукавых сігналаў. Характар гэтых змен вызначаецца не толькі пэўным тэкстам, але і персанальнай манерай яго чытання. Рашэнне задачы кланавання прасадыхных характарыстык маўлення складаецца ў стварэнні досыць поўнага набору персанальных інтанацыйных "партрэтаў" маўлення.

Тэхналогія кланавання і яе прыкладанні. Для паспяховага кланавання персанальных характарыстык голасу і дыкцыі неабходна стварыць досыць поўныя наборы гукавых хваль алафонаў і інтанацыйных "партрэтаў" маўлення. У выпадку, калі дыктар, якога клануюць фізічна даступны, для гэтай мэты выкарыстоўваецца спецыяльна распрацаваны кампактны гукавы масіў слоў і ўрыўкаў тэксту, які начытваецца ім у студыі ці ў звычайных умовах. Калі ж дыктар, якога клануюць недаступны, то выкарыстоўваюцца ўжо наяўныя запісы яго голасу на радыё, тэлебачанні і інш. Першыя вынікі па кланаванні (на прыкладзе персанальнага голасу і дыкцыі аўтара гэтага артыкула) былі атрыманы ў 2000 годзе. У 2001-м годзе атрыманы клон жаночага голасу, а да канца 2005-га набор клонаў складаўся ўжо з 3-х мужчынскіх і 2-х жаночых галасоў. На дадзены момант створана камп'ютарная тэхналогія кланавання фанетыка-акустычных і прасадыхных характарыстык маўлення, якая дазваляе ў значнай ступені аўтаматызаваць і паскорыць працэс стварэння маўленчых клонаў дыктара. Асноўныя навуковыя і практычныя вынікі камп'ютарнага сінтэзу і кланавання маўлення адлюстраваны ў манаграфіі [8]

Адзначым некаторыя магчымыя камерцыйныя і практычныя аспекты камп'ютарнага кланавання. Напэўна знойдзецца вялікая колькасць карыстальнікаў камп'ютара, якія жадаюць, каб іх РС загаварыў яго ўласным голасам ці, напрыклад, голасам блізкага яму чалавека, ці любімага акцёра. Цікавым можа быць таксама праект ажыўлення галасоў вялікіх людзей, якія



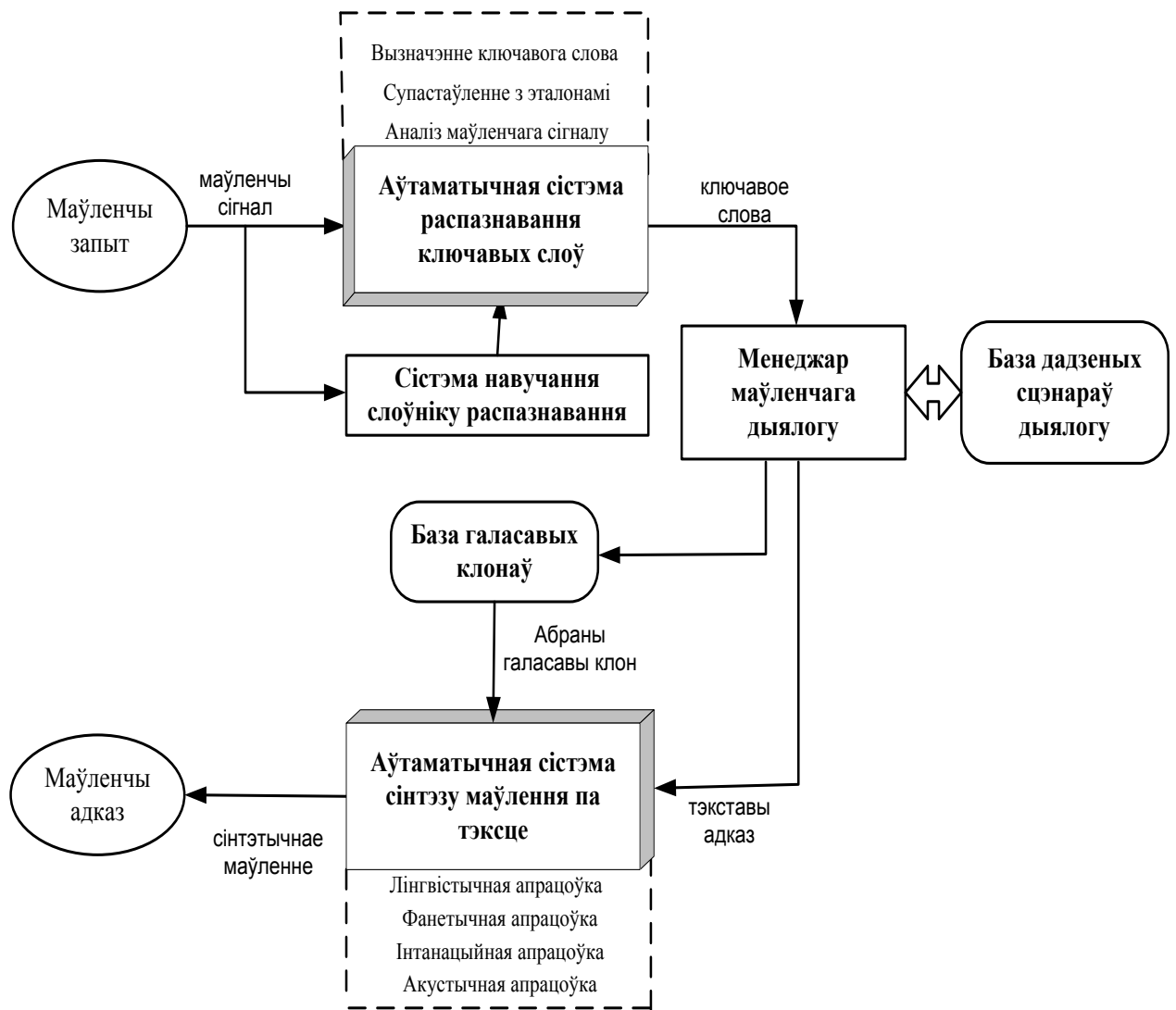
даўно пайшлі ад нас, па пакінутых ад іх грамафонных ці студыйных запісах. Распрацоўка тэхналогіі стварэння галасавых клонаў можа з'явіцца кардынальным сродкам барацьбы з так званым тэлефонным тэрарызмам, забяспечыўшы ідэнтыфікацыю асобы па голасе шляхам аўтаматычнага параўнання аператыўнага запісу голасу з напаўненнем БД галасавых клонаў патэнцыйных правапарушальнікаў.

#### ***4.2. Камп'ютарная мадэль маўленчага віртуальнага суразмоўцы***

Камп'ютарная мадэль вусна-маўленчага віртуальнага суразмоўцы (сістэма "PEBIPC") - новая распрацоўка лабараторыі распазнавання і сінтэзу маўлення, у якой інтэграваны арыгінальныя навукова-тэхнічныя рашэнні, атрыманыя супрацоўнікамі лабараторыі на працягу апошніх гадоў. Сістэма PEBIPC дазваляе ствараць сцэнарыі дыялогаў для разнастайных прыкладанняў і ажыццяўляць іх пры дапамозе вусна-маўленчых чалавека-машынных зносін. Унікальнасць сістэмы PEBIPC заключаецца ў тым, што ў ёй рэалізуецца:

- ✓ надзейнае распазнаванне *ключавых слоў* запыту ў *бесперапынным патоку маўлення*;
- ✓ *шматдыктарнае* распазнаванне *ключавых слоў ва ўмовах акустычных перашкод і скажэнняў*;
- ✓ *шматгаласавы* сінтэз маўлення *па адвольным тэксце*;
- ✓ магчымасць "*кланавання*" *голосу асобы падчас сінтэзу маўлення*;
- ✓ *дуплексны рэжым у рэальным часе* (магчымасць перапынення галасавога адказу).

Структурная схема сістэмы прыведзена на мал. 10



Мал. 10. Структурная схема сістэмы РЕВІРС

Маўленчы сігнал паступае на ўваход сістэмы распазнавання маўлення, якая ажыццяўляе аналіз інфарматыўных прыкмет сігналу, іх супастаўленне з эталонамі ключавых слоў, выяўленне і прыняцце рашэння пра вымаўленае слова. Калі распазнавальны модуль выявіў ключавое слова, яно перанакіроўваецца менеджару маўленчага дыялогу, які фармуе тэкставы адказ. Менеджар маўленчага дыялогу выбірае таксама галасавы клон для сінтэзу маўленчага адказу. Абраны клон і тэкст адказу паступаюць на ўваход сістэмы сінтэзу маўлення, якая ажыццяўляе лінгвістычную, інтанацыйную,

фанетычную і акустычную апрацоўку, у выніку якой тэкст пераўтвораецца ў гучальнае маўленне, якое гучыць зададзенага галасавога клона.

### **Заклучэнне**

Выкладзены ў гэтым раздзеле кароткі нарыс гісторыі стварэння і развіцця маўленчых тэхналогій у Беларусі не прэтэндуе, канешне, на поўнасць асвятлення ўсіх навукова-тэхнічных вынікаў 40-гадовых даследаванняў. Не было магчымасці нават коратка згадаць працы ўсіх аўтараў, якія ўнеслі пэўны ўнёсак у развіццё гэтай галіны ведаў у Беларусі (у агульным імі апублікавана больш за 300 навуковых прац, у т.л. 5 манаграфій).

Нягледзячы на тое, што пройдзены вялікі шлях і атрыманы адчувальныя навуковыя і практычныя вынікі, мяркую, што гісторыя маўленчых даследаванняў у Беларусі на гэтым не скончыцца. Як і 40 гадоў назад, разгадка прыроды чалавечага маўлення застаецца займальнай для маладых даследчыкаў, а магчымыя прыкладанні маўленчых тэхналогій як ніколі становяцца актуальнымі.

### **Спіс літаратуры**

1. Lobanov V.M. More About Speech Signal and the Main Principles of its Analysis // IEEE Transactions on Audio and Electroacoustics.- 1970, N 3. – P. 316-318
2. Lobanov V.M. Classification of Russian Vowels Spoken by Different Speakers // The Journal of the Acoustical Society of America.- 1971, N 2 (2). – P. 521 - 524
3. Лобанов Б.М., Слуцкер Г.С., Тизик А.П. Автоматическое распознавание звукоочетаний в текущем речевом потоке // Труды НИИР, Москва, 1969.- С. 67 - 75

4. Lobanov B.M. The Phonemophon Text-to-Speech System // Proceedings of the XI-th International Congress of Phonetic Sciences, Tallin, 1987.- P. 100 - 104
5. Lobanov B.M, Karnevskaya E.B. MW - Speech Synthesis from Text // Proceedings of the XII International Congress of Phonetic Sciences.- Aix-en-Provence, France 1991.-P. 387 - 391
6. Lobanov B.M., Levkovskaya T.V. Continuous Speech Recognizer for Aircraft Application // Proceedings of the 2<sup>nd</sup> International Workshop “Speech and Computer” – SPECOM’97.-Cluj-Napoca, 1997.-P. 97-102.
7. Lobanov B.M. et al. An Intelligent Answering System Using Speech Recognition // Proceedings of the 5<sup>th</sup> European Conference on Speech Communication and Technology – EUROSPEECH’97. V. 4.- Rhodes-Greece, 1997.- P 1803-1806.
8. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи // Минск: Белорусская Наука, 2008. – 342 с.