

ІДЕНТИФІКАЦІЯ ПЕРЕВАЖНОЇ МОВИ СПІЛКУВАННЯ ЛЮДИНИ

Яна Лазоренко, Ігор Сініцин, Віктор Шевченко

Зі збільшенням обсягів переміщення населення, як тимчасового, так і на тривалий час через військові дії, роботу тощо, виникає потреба в удосконаленні процесів контролю переміщення та ідентифікації груп людей. Першочерговою потребою є виявлення приналежності людини до певної національності чи території переважного перебування. Крім того, це може бути корисно й для наукових соціо-політичних досліджень, а також сфери туризму, адже така ідентифікація людей за мовним середовищем може говорити й про культурне середовище, що дозволить краще передбачати вподобання цілих груп клієнтів.

Обробку аудіозапису мовлення диктора було розбито на 3 етапи, кожен із яких містить визначену покроково інструкцію процесу підготовки та обробки даних. Спочатку було проведено аналіз самого звуку, адже голос і вимова є одними із найшвидших і при цьому найефективніших інструментів ідентифікації особи чи групи осіб. Потім аудіозапис було конвертовано в текст і проведено аналіз лексичного складу досліджуваного фрагменту розмови. У кінці результати роботи програми, отримані в попередніх двох етапах, були комплексно оцінені за визначеними в ході теоретичного дослідження критеріями з присвоєнням їм вагових коефіцієнтів. Як підсумок, було отримано припущення про мовне середовище диктора.

У роботі були розглянуті чинники, що впливають на формування вимови та побудови різних інтонацій, особливості взаємозв'язку тексту та звуку, а також були математично формалізовані деякі залежності між звуковими параметрами та вимовою звуків у російській і українській мовах.

Тема роботи широкою, оскільки розглядає не тільки лексичні, а й акустичні особливості мовлення. Тож враховуються не лише мінливі чи ситуативні якості, наприклад, словниковий запас розмови, а й властиві людині незалежно від контексту розмови, її емоційного стану чи того, якою мовою вона говорить у даний момент. Розглянуті у роботі критерії дозволяють робити припущення про «природну», звичну диктору мову, тобто ту, якою він спілкується давно й більшість часу.

Ключові слова: ідентифікація мовця, аналіз тексту, звукові характеристики, акустичний спектр звуків.

With the great increase of population movement, caused either by temporarily needs due to travelling or by long-term ones due to work, etc., there appears a need to improve the processes of movement control and identification of groups of people. The primary need is to identify if a person belongs to a certain nationality or territory of primary residence. In addition, it can be useful for scientific social and political researches, as well as for the field of tourism and entertainment. Such identification of people by language environment can also help to find out more about the cultural environment, which will make possible to predict the preferences of entire groups of customers better.

The analysis of the audio recording of the speaker's speech was divided into 3 stages, each of which contains a defined step-by-step instruction for the processes of data preparation and handling. Firstly, the sound itself was analyzed, because voice and pronunciation are one of the fastest and most effective tools for identifying a person or a group of people. Then the audio recording was converted into text and an analysis of the lexical composition of the studied fragment of the conversation was done. At the end, the results of the program, got in the previous two stages, were compared and the complex evaluations were done according to the criterias determined in the theoretical researches. The mentioned criterias were used with weighting factors assigned to them. As a result, an assumption about the speech environment of the speaker was given by the program.

The work also describes the factors which affects the formation of pronunciation and the change of various intonations, the relationship between text and sound. Some dependencies between sound parameters and the pronunciation of sounds in the Russian and Ukrainian languages were mathematically formalized.

The topic of the work is wide, as it considers not only lexical, but also acoustic features of speech. In this way, not only changeable or situational qualities are taken into account, for example, the vocabulary of the conversation, but also those ones that are related to the person regardless of the context of the conversation, his or her emotional state or what language he or she is speaking at the very moment. The criterias considered in the work can help to make assumptions about the speaker's «natural», most familiar language, in other words the language he or she uses most of the time in life.

Keywords: speaker identification, text analysis, sound characteristics, acoustic spectrum of sounds.

Вступ

Щороку транспортно-логістична та й туристична галузі активно розвиваються, адже люди релокуються у зв'язку із військовими діями, зі зміною роботи, переїжджають через клімат і проблеми зі здоров'ям тощо.

Тому автоматизація процесів контролю переміщення та ідентифікації груп людей стає все більш актуальною й вимагає удосконалення наявних інструментів і застосування нових підходів. Одним із найшвидших засобів є визначення приналежності людини до певної національності чи території переважного перебування за вимовою. Для цього лише необхідно проаналізувати короткий уривок розмови, що займає мінімум часу. Разом із тим такий метод є досить ефективним, бо вимова формується протягом тривалого часу й так само потрібно багато часу, аби її скоригувати. Деякі ж відмінності у вимові й акценті закладаються ще в дитинстві й зберігаються назавжди. Це дозволяє визначити, яка мова була для людини першою, або якою мовою вона переважно говорить. Крім того, вимова є маркером і територіального походження людини, бо навіть різним регіонам у межах однієї країни притаманні різні говірки.

Так, обробка мовлення людини задля ідентифікації її особистості є актуальним і досить широким завданням, тому що охоплює аналіз як тексту, так і звуку, а також як вроджених особливостей (характеристики власне голосу людини), так і набутих або ситуативних (засилля діалектів чи кальки, характерної для певної місцевості, або інтонація і т. д.). Однак наразі відсутні автоматизовані засоби, які використовують комплексний аналіз таких

факторів, що робить поки неможливим масове використання такого способу ідентифікації. Проте для вирішення цієї проблеми важливо здійснювати дослідження й розробки, адже мова й мовлення – це першочерговий і дуже наочний маркер багатьох спільних рис соціальних груп суспільства.

Аналіз стану в прикладній галузі

Більшість прикладних моделей або експертних систем для обробки живого мовлення людини сфокусовані саме на визначенні мови диктора та розпізнаванні його мовлення. Тобто фактично основною задачею цих рішень є транскрибування аудіозапису в текст. Надалі ж у тексті аналізується переважно логічний зміст, тобто завданням є визначити, про що йдеться в тексті й вибрати необхідні дані з нього.

Натомість програм чи застосунків, які б зосереджувалися саме на особливостях мовних конструкцій, використаних у тексті, або на специфічних ознаках вимови дуже мало. І знову ж таки, вони вирішують загалом одну прикладну задачу – розпізнавання тексту. Однак це не дозволяє співвідносити отримані дані з особистістю мовця й не надає ніяких його характеристик, крім того, власне, якою мовою він зараз говорить. У реальності ж зовнішність і мовленнєві навички дають чи не найбільше інформації як про людину окремо, так і про групи людей, залежно від обраної ознаки для класифікації.

Отже, найперше, що ми маємо – це голос. Він містить риси, притаманні виключно конкретній людині. Тому зараз проводиться досить багато досліджень і проєктується систем для ідентифікації особистості за голосом. У своїй спільній роботі М.В. Ткаченко, Р.М. Федоренко, Ю.В. Кондратенко й І.Г. Зотова(1) детально розглянули характеристики голосу, принципи їх формування та можливість використання цих даних для визначення диктора. Є й уже готові системи, якими користуються в банківських, медичних системах і частині державних органів США, наприклад, розробки від компаній Nuance Communications Inc. [2] і Biocatch [3]. Дані матеріали акцентують увагу на взаємозв'язку унікальних голосових параметрів та ідентифікації мовця. Водночас результати цих робіт є корисними й для класифікації груп людей на основі їхнього мовлення, де потрібно узагальнити їхні спільні мовленнєві риси й знехтувати зазначеними унікальними рисами. Крім того, голос людини частково змінюється протягом життя залежно від віку, хвороб, емоційного стану тощо. Отже, це теж потрібно врахувати й «згладжувати» такі похибки при аналізі мовлення цілої групи людей. Це дозволить запобігти великій кількості помилок і уникнути ефекту «перенавчання» моделі, заснованої на нейронних мережах.

Над більш загальними характеристиками голосу й закономірностями їхніх змін працювали Ю.С. Ямненко й О.В. Смирнов. У спільній роботі авторами наведено принципи зміни голосу залежно від емоцій та психологічного стану мовця [4]. Це один із найважливіших аспектів, який необхідно врахувати при спробі класифікації людей на групи за мовленням, адже нетипові за звучанням фрагменти аудіозапису можуть свідчити не про відмінність людини від певної групи, а всього-на-всього про особливості її психічного стану в конкретний момент. Якщо ж деяка схожа емоційність насправді присутня на аудіозаписах більшості людей групи, але відомо, що вони перебувають у різних психологічних станах, то можна говорити саме про інтонацію мовлення. Це один із важливих інструментів для ідентифікації груп людей, оскільки кожна мова відрізняється своєю інтонаційністю й рівнем вираженості цього явища. При цьому зазвичай люди лишають у мовленні інтонації своєї рідної мови, навіть коли переходять на іноземні мови. І виправити таку «звичку» доволі складно. Таким чином можна класифікувати людей за їх рідною мовою чи мовою, якою вони спілкуються більшість часу, вирізняти специфічність їхньої говірки й відповідно характеристики середовища й території переважного перебування.

Доктор філологічних наук Багмут Алла Йосипівна багато років досліджувала інтонаційні особливості мови саме теоретично. У своїх працях вона надає не лише ґрунтовні теоретичні матеріали щодо інтонації як явища зміни голосу, а й дуже детально розбирає правила побудови інтонації й емоції в мовленні за синтаксичними й семантичними особливостями речень. Так, декотрі монографії філологині присвячені розбору лише одного виду речення, приміром, розповідного речення з рівною інтонацією [5], [6]. Разом із цим Міннігалімову Руслану Тимуровичу вдалося формалізувати деякі з цих закономірностей у роботі «Аналіз та синтез українського мовлення» [7], де він математично описав закономірності зміни частот основного тону для розповідних стверджувальних і заперечних, а також питальних речень. Однак цього недостатньо для комплексного аналізу мовлення людини.

Отже, цілком роботи є виявлення закономірностей зміни характеристик голосу притаманних для певних мов; порівняння відмінностей між мовленням дикторів різними мовами та їх формалізація; аналіз як звукових складових мовлення, так і лексичних задля комплексного аналізу мовлення людини; визначення послідовності кроків аналізу мовлення за голосом і текстом при класифікації дикторів за мовою, яку вони переважно використовують. Для досліджень було обрано порівняння мовців української та російської мов, проте загальні принципи підходу є універсальними й можуть застосовуватися для опрацювання інших мов.

Наявні теоретичні підходи

Переважна більшість методів ідентифікації мовця за голосом і аналізу текстів базується на алгоритмах машинного навчання. Такий підхід є досить результативним, й до того є досить гнучким і залежним від вхідних навчальних даних. Отже ефективність роботи програми може суттєво коливатися залежно від обраної вибірки для навчання. Не правильно обчислені статистичні дані теж можуть впливати на роботу такої системи. Натомість поєднання машинного навчання з чітко визначеними правилами та математично формалізованими закономірностями завжди підвищує якість рішення й робить його стабільнішим.

Багато програм розпізнавання мовлення використовують метод динамічного програмування, що працює за допомогою порівняння зразків аудіо чи тексту з еталоном, і ряд додаткових методів відокремлення окремих елементів (звуків чи фонем), які становлять склади та морфеми. Широкого вжитку набули нейронні мережі, Марківські моделі та методи дискримінантного аналізу.

Також для розбиття аудіозапису на синтаксично цілісні фрагменти за змістом використовують розбиття за паузами, що зазвичай свідчать про знаки пунктуації в тексті. Приміром, після коми йде мінімальна пауза, а після тире або крапки – більша [8]. За цими паузами, їх частотою та тривалістю роблять припущення про емоційний стан мовця, його мовленнєві проблеми тощо, але при цьому не враховують те, наскільки паузи взагалі характерні для тієї чи іншої мови. Отож, залишається неможливим охарактеризувати мовця в контексті соціуму – тільки його особисті вади чи переживання в конкретний момент.

Методи ідентифікації людини за текстом побудовані на виявленні найчастіше вживаних автором слів і конструкцій. Таким чином це так само, як і з аудіозаписами, дає змогу охарактеризувати конкретну людину особистісно, але не її в соціумі. Тому що окремі слова, яким надає перевагу конкретна людина, чи її особистий стиль мовлення не означають, що риси цієї людини притаманні всій групі людей, із котрими вона взаємодіє. Для виявлення ознак групи необхідно зосереджуватися не на унікальності певних мовців, а навпаки на схожості їхнього мовлення в межах групи, сформованої, скажімо, за національною ознакою. Узагальнивши отримані знання, можна вивести риси унікальності для всієї групи, що вже робить уможливлене ідентифікацію цієї групи серед інших груп соціуму та подальше співвіднесення окремих мовців до тієї чи іншої групи.

Отже, головна суперечність між описаними підходами й практичними потребами полягає в тому, що принципи ідентифікації й характеристики мовців базуються винятково на особистих особливостях мовлення людини й розглядають саму людину лише як окрему сутність, незалежну від інших. Проте на мовлення людини великою мірою впливає її оточення та місце проживання. Це вже говорить про те, що кілька людей можуть мати ідентичні ознаки мовлення. Крім того, людина підсвідомо здатна розпізнавати різні говірки, діалекти, акценти тощо буквально за кількома фразами, однак робочих програмних засобів для цього фактично немає. Із цього можна зробити висновок, що мовлення групи людей можливо узагальнити й формалізувати за деякими звуковими й текстовими характеристиками. А знайдені залежності між цими мовними характеристиками та мовним середовищем людини, територією проживання, родом занять тощо використовувати для практичних задач, де можна знехтувати особистими якостями людей. Але водночас важливою є саме їх класифікація як груп відповідно до соціальних, політичних або географічних критеріїв. Така автоматизація може бути інтегрована у різні галузі людського побуту, у процеси статистичних соціологічних досліджень тощо.

Мовні критерії для класифікації дикторів за мовленням

Щоб розпізнавати приналежність мовця до певної групи за його мовленням у різних галузях і для різних прикладних задач, важливо аналізувати саме побутову розмову, таку, якою людина говорить у житті за невимушених обставин [9]. Варто зазначити, що ефективніше розпізнавати мову побутової розмови за звуком, аніж за текстом, особливо в контексті обраних мов для даної роботи. Це зумовлено тим, що, якби там не було історично, але наразі українська, російська і частково білоруська досить сильно змішані на території України. Тобто фактично більшість людей говорить суржилом, десь ближчим до української, десь – до російської, а в північних регіонах до білоруської. До того ж прикордонні регіони Росії мають говірки, де лексика дещо «українізована», тобто теж говорять суржилом. Тому простий аналіз тексту може давати суттєві похибки. Його радше доцільно застосовувати, як один із критеріїв визначення мови із певним ваговим коефіцієнтом.

Яскравим маркером мови може бути наголос – у російській і українській наголоси значно відрізняються і в дуже багатьох словах, які використовуються саме в повсякденних побутових розмовах [10]. Тож наголос – це такий проміжний елемент між аналізом тексту і звуку, бо треба розпізнати на якому складі підсилюється звук у мовленні, а тоді порівняти із тим, на якому складі він звично підсилюється в одній та іншій мові. Звісно, це теж не абсолютний показник, бо через тривале зросійщення ми звикли деякі слова наголошувати на російський лад. Але все-таки наголос варто включити до сукупності критеріїв, і до того ж із досить високим коефіцієнтом. Він фактично визначає, в якому мовному середовищі зростала людина, а отже, із часом ця ознака як маркер впевнено закріплюється в мовленні людини.

Розбираючи мову на менші фрагменти врешті-решт доходимо до звуків. Тут відмінностей дуже багато і вони суттєві. Тому, власне, багато питома українських слів дійсно неможливо вимовити російською вимовою, і ті, хто не мав україномовного оточення, просто не знають, як вимовити, скажімо, начебто той самий звук «и», але українською. Тому аналіз звукових відмінностей мови є складнішим, але натомість більш однозначним і ефективним, тому що навіть коли людина говорить суржилом, тобто лексично вживає слова іншої мови, вимова все одно залишається однаковою для всіх слів.

Власне складність запропонованого методу полягає перш за все в тому, що розпізнавати й відловлювати окремі звуки в мовленні важче, аніж цілі слова. Крім того, насправді «технічної» літератури з філології, готових бібліотек для ідентифікації мови за окремими звуками чи програмних алгоритмів для такої задачі досить мало. Усі правила вимови звуків і відмінності між звуками пояснюються філологами переважно інтуїтивно. Наприклад, якщо ми чуємо багато м'якого «л», то це говірка Полтавщини, якщо «в» зм'якшується до «у», то це Захід України. Однак фізики допомагають філологам і дають перелік визначених критеріїв, як саме ми це чуємо, які характеристики відповідають за зміну цієї «м'якості» звуку. Слід підкреслити, що більшість алгоритмів працюють або зі звуком, або з текстом. У даній задачі між звуком і текстом тісний зв'язок, тож обробляти, аналізувати треба і

текст, і звук у парі відносно одне одного. Тому основою рішення є формалізації фонетики й фонології відносно аудіохарактеристик і подальше упорядкування алгоритмів.

1. Із усього розглянутого вище можна сформуванати базову послідовність кроків обробки мовлення:
2. підготовка аудіозапису, у тому числі фільтрація сторонніх шумів;
3. виявлення й відкидання емоції, оскільки вона говорить про психічний стан мовця, а не про особливості мови чи говірки;
4. транскрибування аудіозапису в текст і його аналіз:
 - a. визначення мови тексту;
 - b. виявлення слів, які мають однакове чи схоже написання в обох мовах;
 - c. виявлення нехарактерних для визначеної мови слів, що говорить про кальку, суржик або діалекти;
5. порівняння наголосу диктора в словах, які схожі за написанням в обох мовах, із правильним наголосом цих слів у одній та іншій мові;
6. виокремлення з аудіозапису кількох фрагментів зразків окремих звуків (особливо відмінні в українській і російській голосні), згладжування їхніх звукових характеристик і порівняння отриманих даних із «еталонними» характеристиками цих звуків у обох мовах;
7. припущення про мовне середовище чи територію переважного перебування на основі алгоритму аналізу отриманих даних за критеріями з урахуванням певних вагових коефіцієнтів для них;
8. можливо, порівняння результату даної класифікації з класифікацією мовця методами машинного навчання, оцінка похибок і коригування припущень.

Загальна обробка аудіозапису

Широкий спектр емоцій людини можна звести до трьох основних груп: позитивні, негативні та нейтральні. Перші використовуються для позначення щастя, задоволення, гарного настрою. Досягається це за допомогою підвищення тону голосу, іноді його гучності й ритмічності. Водночас інтонація фраз плавно йде вгору, без стрибків.

Негативні емоції навпаки виявляють невдоволеність, неприємність і погане самопочуття. Емоції цієї групи більше відрізняються одне від одної за зміною звукових параметрів, оскільки людина в стані пригнічення і людина в стані запалу може відчувати негативні емоції, але передаватися вони будуть інакше. Проте в цілому основною їхньою відмінністю від позитивних є відсутність плавності у звучанні, ступінчастість переходів між фрагментами на аудіозаписі. Окрім цього, для багатьох негативних емоцій відчутне зниження висоти звуку [4], тоді як гучність може коливатися залежно від яскравості емоції й контексту її використання.

Ознаки емоцій перших двох груп необхідно виявляти й відкидати, дещо згладжуючи запис, аби вони не спотворювали результати подальшого аналізу мовлення диктора. Іншими словами, аудіозапис необхідно за виразністю наблизити до нейтральних емоцій, які передають спокійний, розслаблений і впевнений стан мовця. У такому аудіозаписі будуть відсутні стрибки звукових характеристик, зберігатиметься загальна плавність.

Для того, аби відфільтрувати емоції, у роботі аналізувалися цілі речення, адже частіше за все вони містять у собі лише одну значущу синтагму, тобто фрагмент, який є семантично і емоційно неподільним. Тож масиви значень звукових характеристик представлялися як узагальнені функції ліній поліноміальної апроксимації третього порядку, адже кубічна функція показала оптимальні результати в ході роботи: забезпечувалося необхідне згладжування графіків, але разом із тим зберігалися найбільші перепади значень. Потім отримані залежності порівнювалися з закономірностями зміни звукових характеристик для позитивних і негативних емоцій [11]. Якщо емоція визначалася як позитивна чи негативна, то запис згладжувався ковзним середнім, інакше – лишався без змін і приймався за емоційно нейтральний.

Продовжуючи роботу безпосередньо зі звуком, перейдемо до розбиття звукового ряду на малі фонетичні одиниці. Раніше зазначалося, що найбільш відмінні у вимові російською та українською голосні звуки. Тому відокремлюємо з уривків сигналу саме їх. Для цього ділимо аудіозапис на склади, адже склад завжди містить один і тільки один голосний звук. Просто ділити запис на однакові за тривалістю проміжки не можна, тому що склади різняться за тривалістю звучання. Тому використаємо спосіб відкидання приголосних, тобто спочатку знайдемо склади за приголосними. Потім виберемо з уривку аудіозапису, де є тільки один склад, ділянку, де звукові характеристики мають більші значення, що свідчить про використання саме голосу для вимови голосних звуків.

На голосному звуці всі параметри зростають, зокрема, й гучність, тоді як на приголосному знижуються й стають майже нульовими на шиплячих і свистячих. Це говорить про те, що можна розбити звукову доріжку на уривки за складами, використавши значення певного порогу, коли вважатимемо, що всі значення нижче нього відповідатимуть приголосним звукам, тоді як вище – голосним. Найкраще спрацювало порогове значення у мінус 30-40дБ (dBFS) тривалістю 40-50мс, однак це, звісно, залежить від початкової гучності аудіозапису. Продемонструємо це графічно на масиві гучностей якогось запису в порівнянні з текстовим варіантом фрази. Розбиваємо його за проміжками та виберемо для прикладу один із складів, як показано на рис.1.

Далі опрацюємо склади окремо, аналогічно фразі, але встановлюємо дещо вищий поріг гучності й скорочуємо тривалість проміжків. На рис.2 показано, що таким чином ми відбираємо центральну частину фрагменту-складу, а затінені на зображення сірим шматочки відкидаємо, бо це приголосні звуки. У підсумку після розбиття запису на склади та відкидання «умовно безголосих» частинок отримуємо масив, що складається виключно з голосних звуків.

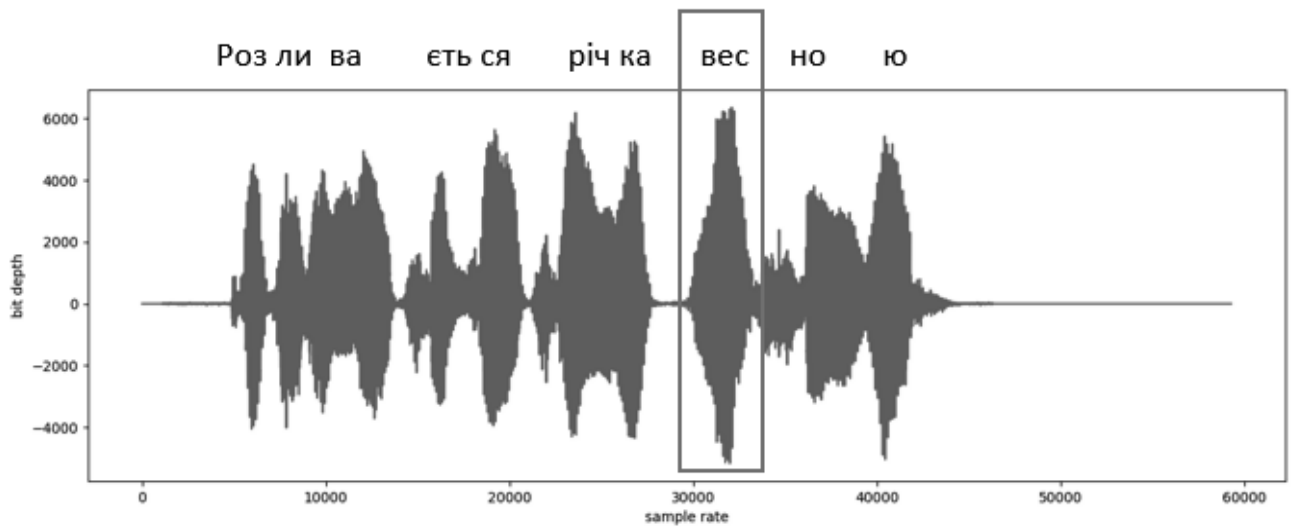


Рис.1. Графічна відповідність тексту й голосу



Рис.2. Графічна відповідність літер (звуків) складу й фрагменту голосу

Порівняння голосних звуків на аудіозаписі

Для того, аби відрізнити звуки, зокрема голосні, між собою, а також їх звучання в українській та російській мові, треба їх апроксимувати поліномом 6-го порядку (адже для розбору звуків необхідна більш точна деталізація, тому 3-го порядку вже не достатньо) і узагальнити результат. А крім того, визначити, що брати за еталон. Для цього розглянемо систему звуків української та російської мов.

Отож українська містить 6 голосних фонем: а, о, у, е, і, и. Російська - теж 6: а, о, у, э, и, ы. Тому співставляти звуки можна попарно. Розглянемо, як змінюються звукові характеристики для кожного з них із точки зору акустики. Олександр Іщенко у дослідженні «Акустична класифікація українських голосних звуків» наводить графіки зміни гучності звуків за формантами (призвуками, що надають звучанню музичного інструмента або голосу характерного забарвлення – тембру). Аналогічні дослідження проводили С.В.Князев і С.К.Пожарицька. Апроксимувавши графічні залежності, знайдені авторами, поліномом 6-го порядку, отримуємо наступні функції зміни інтенсивності звукових характеристик (рис.3-8).

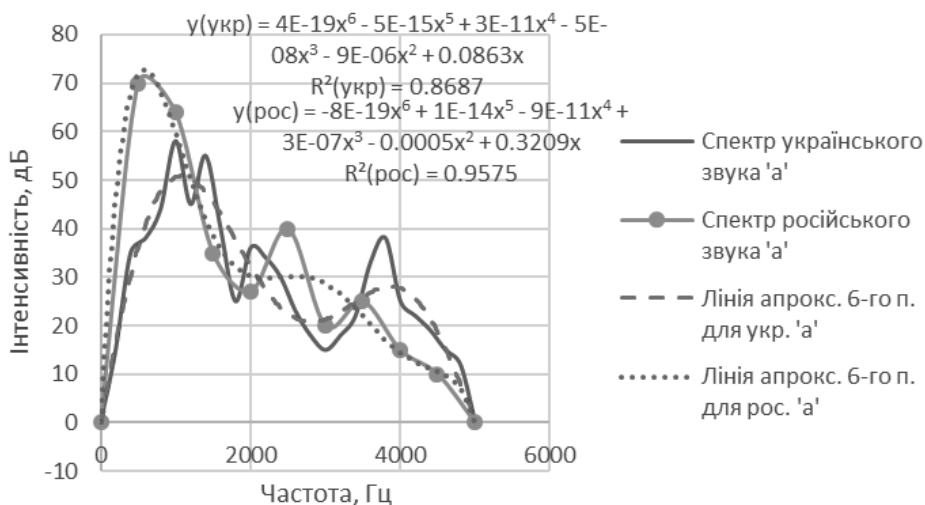


Рис.3. Порівняння українського звука 'а' та російського звука 'а'

Із даних графіків бачимо, що російська вимова звуку 'а' має значно вищу інтенсивність на початковому проміжку.

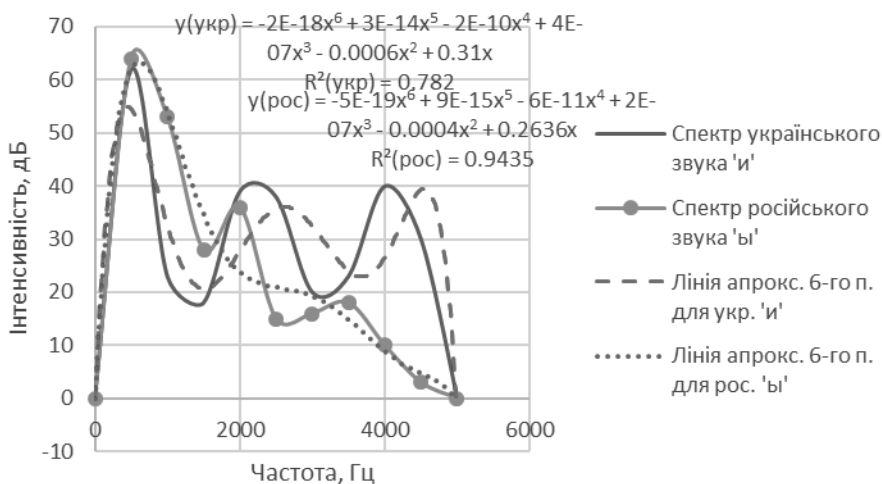


Рис.4. Порівняння українського звука 'и' та російського звука 'ы'

Тут знову бачимо суттєву відмінність – український звук більш рівномірний на всьому проміжку, тоді, як російський відповідник має виражену «голосну» частину тільки на початку.

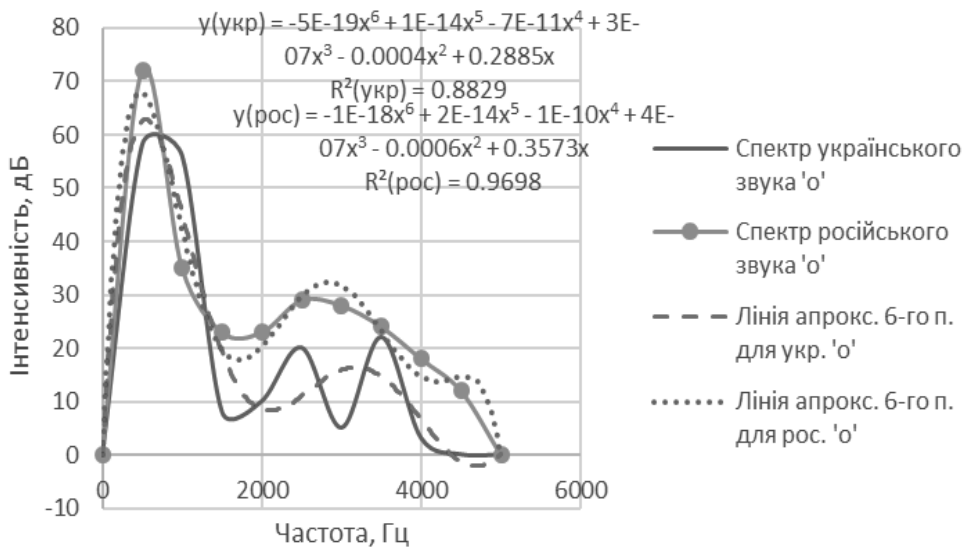


Рис.5. Порівняння українського звука 'о' та російського звука 'о'

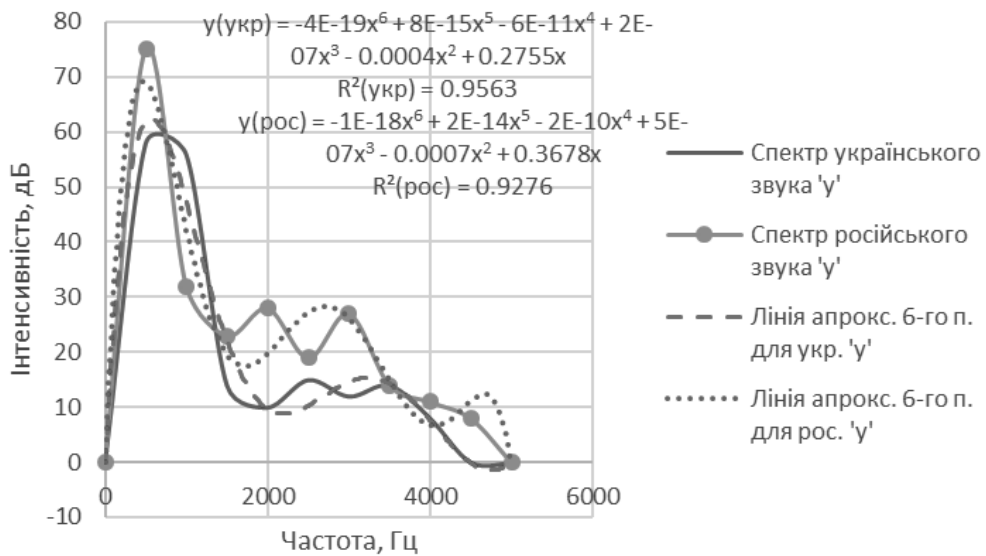


Рис.6. Порівняння українського звука 'у' та російського звука 'у'

Українська вимова звуків 'о' та 'у' більш усічена наприкінці, гучність спадає більш різко.

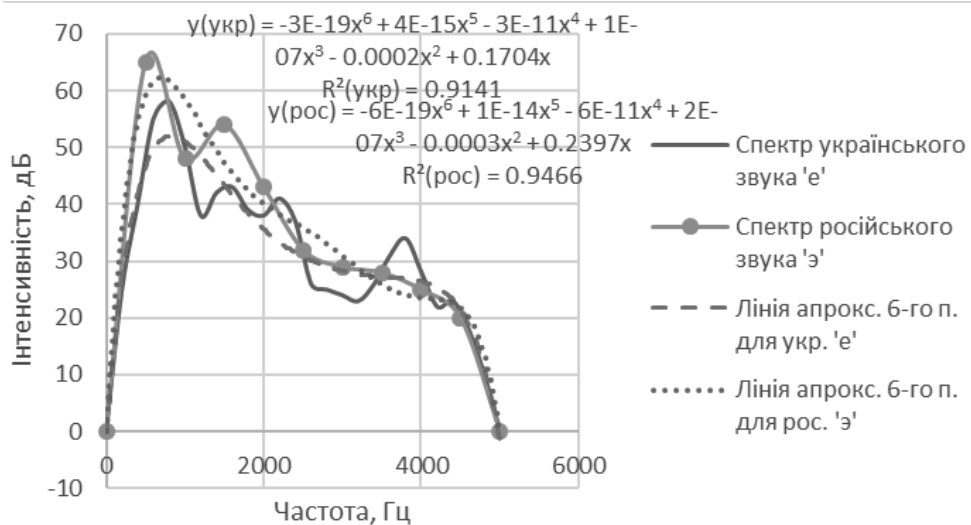


Рис.7. Порівняння українського звука 'е' та російського звука 'э'

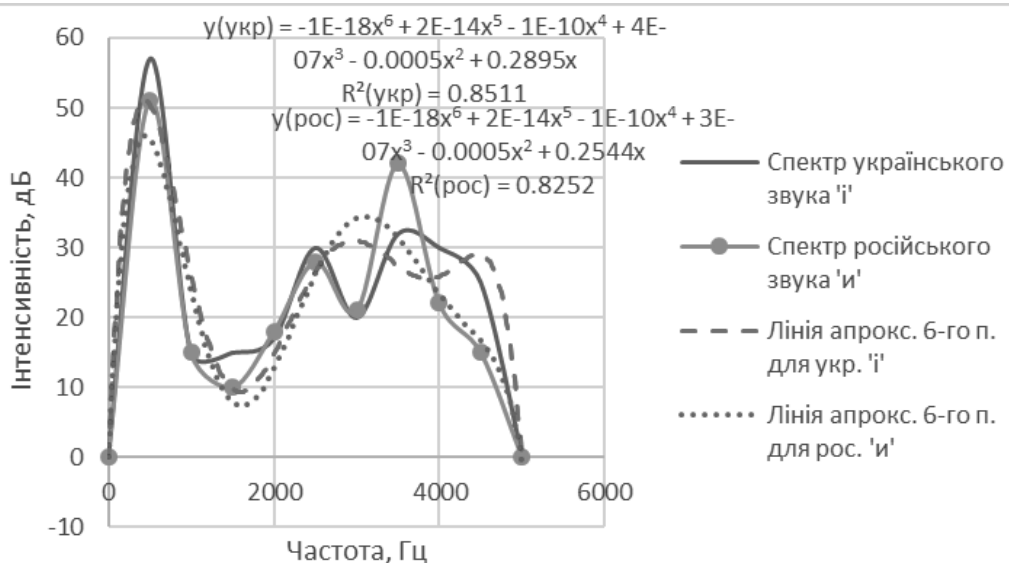


Рис.8. Порівняння українського звука 'і' та російського звука 'и'

На відміну від попередніх звуків, звуки 'е' та 'э' або 'і' та 'и' практично не відрізняються, тому при подальшому аналізі цей звук не повинен мати великого вагового коефіцієнту.

Отже, на даному етапі уже можна робити припущення, яка вимова для диктора ближча. Ще одна значима відмінність російської від української – це те, що в ній значно більше звуків 'а' та 'і', тому навіть кількісні характеристики отриманих масивів потрібно враховувати як один із критеріїв комплексного аналізу мовлення. Надалі для ідентифікації особливостей говірки мовця перейдемо до обробки й аналізу тексту.

Лексична обробка тексту

Для того, аби проаналізувати мовлення людини з точки зору лексики, необхідно конвертувати аудіозапис у текст. Відтак розбиваємо його на масив слів і укладаємо словник з пар «слово – кількість уживань», враховуючи не лише повнозначні слова, а й службові також, оскільки сполучники, прийменники й частки також відрізняються в російській і українській мовах. Тим часом люди саме в них дуже часто роблять помилки, використовуючи скальковані службові слова, уподібнюючи їх до звучання їхньою рідною мовою.

Тепер проаналізуємо отриманий словник слів і розділимо його на слова, типові як для української, так і для російської, питомо українські та питомо російські, а також нетипові для жодної з мов за класичним словником, тобто це скальковані слова, діалекти тощо. І тут важливо звернути увагу не лише на кількісні характеристики груп слів, а й, так би мовити, на якісні. Бо кількох неправильних сполучників чи діалектів достатньо, аби в купі з отриманими результатами аналізу аудіозапису зробити припущення про переважну мову спілкування мовця, навіть якщо зараз він говорить іншою.

Крім того, на початку зазначалося, що наголоси можуть слугувати відмінними маркерами для ідентифікації мовця. Тут було використано попередньо підготований словник повнозначних слів, які мають однакове написання в обох мовах, але наголос у них падає на різні склади. Таким чином отримали колекцію пар «слово, масив наголосів», де масив наголосів – це масив із двох елементів, де першим є порядковий номер наголошеного складу в українській мові, а другим – у російській. Ці слова потім зворотно знаходилися на аудіозаписі й розбивалися на склади за описаним у попередніх розділах алгоритмом. За наголошений склад обирався той, на якому звукові характеристики мали вищу інтенсивність.

Підсумкові дані, отримані програмою під час усіх етапів обробки звуку й тексту треба зібрати докупи й зробити певні припущення щодо мовця. Для цього було створено окремий модуль, який надає кожному з розглянутих критеріїв певне числове значення ваги, робить необхідні перетворення з результуючими значеннями, перемножує їх і таким чином наближено визначає, яка мова природніша для диктора – за замовчуванням. Наприклад, беремо, що, чим більше кінцеве значення, тим вища вірогідність, що людина переважно використовує українську в щоденному спілкуванні.

Однак, звісно, результати припущення коливаються залежно від вагових коефіцієнтів і алгоритму співставлення критеріїв в цілому. Тому програма все ще на стадії удосконалення, зокрема, дослідження міри впливу взятих до уваги критеріїв на точність ідентифікації мовця.

Використані програмні засоби

Програмний алгоритм на основі описаних вище виявлених рішень і формалізованих закономірностей написаний мовою C# на платформі .NET. Основними класами застосунку є

Звук із набором характеристик таких, як гучність, тривалість, частота тощо;

Склад, який має як загальні характеристики, у тому числі позначку наголошеності-ненаголошеності, так і масив об'єктів типу Звук;

Емоція, що містить масиви коефіцієнтів зміни основних характеристик голосу;

Аудіофрагмент, що складається з масиву Склад-об'єктів, загальних характеристик, а також об'єкту Емоція; та інші допоміжні класи-оброблювачі для всіх цих сутностей.

Для часткової передобробки аудіозаписів було використано бібліотеки C# у вигляді Nuget-пакетів, такі як NAudio, SoundFingerprinting, SoundTouch.Net, а для виконання власне алгоритмів із масивами використовувалися користувацькі алгоритми, де для базових операцій замість циклів було застосовано методи бібліотеки Math.NET Numerics.

Для передобробки тексту, створення словників вживаних мовцем слів і тому подібне було використано деякі засоби з бібліотек Lucene.Net і SimMetrics.

Усі інші алгоритмічні операції, покроково описані в попередніх розділах, виконувалися за допомогою користувацьких методів.

Висновки

У представленій роботі було досліджено лексику та вимову як складові мовлення людини, розглянуто основні фактори, за якими мовця чи групу мовців можна ідентифікувати в контексті мовного середовища та території переважного проживання.

На основі отриманих закономірностей і формалізацій було створено покрокові алгоритми обробки мовлення як за звучанням, так і за лексичним різноманіттям. Результатом роботи є програмне рішення, яке може використовуватися як окремо, так і в сукупності з іншими компонентами.

У подальших дослідженнях планується визначити міру впливу описаних критеріїв на правильність ідентифікації припущень про мовне середовище диктора, підвищити точність оцінки мовця програмою, а також дослідити можливість розширення й універсальності підходу щодо інших мов.

Література

1. М.В.Ткаченко, Р.М.Федоренко, Ю.В.Кондратенко, І.Г.Зотова. Методи автоматичної ідентифікації диктора за голосом. Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України ім. Івана Черняховського. 2018. № 3 (64). С. 131-135.
2. Офіційний сайт Nuance Communications Inc.: <https://www.nuance.com/index.html>
3. Офіційний сайт Biocatch: <https://www.biocatch.com>
4. Ю.С.Ямненко, О.В.Смирнов. Визначення емоційного стану людини на базі аналізу голосових характеристик. Магістерська дис.: Кафедра промислової електроніки Факультету електроніки НТУ «КПІ ім. Ігоря Сікорського». 2018. 124 с.
5. А.Й.Багмут Інтонційна будова простого розповідного речення в слов'янських мовах. Київ: Наукова думка. 1970. 173 с.
6. А.Й.Багмут Структура і функціонально-семантичний аспект інтонації простого розповідного речення у слов'янських мовах. Академія наук СРСР Інститут слов'язнавства й балканістики. 1980. 48 с.
7. Р.Т.Міннігалімов Аналіз та синтез українського мовлення / Р.Т.Міннігалімов, Київ: магістерська дис.: Кафедра радіоконструювання та виробництва радіоапаратури Радіотехнічного факультету НТУ «КПІ ім. Ігоря Сікорського». 2015. 107 с.
8. Л.А.Близниченко Інтонція мовлення. Київ: Наукова думка. 1968. 203 с.
9. І.Р.Глюк Дослідження акустичних особливостей інтегральних характеристик української мови. Магістерська дис.: Кафедра акустики та акустоелектроніки Факультету електроніки НТУ «КПІ ім. Ігоря Сікорського». 2019. 91 с.
10. С.Я.Єрмоленко Українська літературна вимова і наголос. Словник-довідник. Академія наук Української РСР Інститут мовознавства ім. О. О. Потебні. Київ: Наукова думка. 1973. 724 с.
11. В.Л.Шевченко, Я.С.Лазоренко, О.М. Боровська. Інтонційна виразність тексту при програмному озвучуванні. Науковий журнал «Проблеми програмування». 2021. №2. С. 76-84.
12. С.В.Князев, С.К.Пожарицька Сучасна російська літературна мова. Фонетика, графіка, орфографія, орфоепія. Москва: Академічний проект. 2005. 426 с.

References

1. M.V. TKACHENKO, R.M. FEDORENKO, Y.V. KONDRATENKO, I.H. ZOTOVA (2019) Methods of automatic identification of the announcer by voice. Collection of scientific works of the Center for Military and Strategic Studies of the National Defense University of Ukraine named after Ivan Chernyakhovskyi. (3) P. 131-135. (in Ukrainian)
2. Official website of NUANCE COMMUNICATIONS INC.: <https://www.nuance.com/index.html>
3. Official website of BIOATCH: <https://www.biocatch.com>
4. Y.S.YAMNENKO, O.V.SMIRNOV (2018) Determination of a person's emotional state based on the analysis of voice characteristics. Master's thesis: Department of Industrial Electronics, Faculty of Electronics, NTU "Igor Sikorsky KPI". 124 p. (in Ukrainian)
5. A.Y. BAGMUT (1970) The intonation structure of a simple narrative sentence in Slavic languages. Kyiv: Naukova dumka. 173 p. (in Ukrainian)
6. A.Y. BAGMUT (1980) The structure and functional-semantic aspect of the intonation of a simple narrative sentence in Slavic languages. Academy of Sciences of the USSR Institute of Slavic and Balkan Studies. 48 p. (in Ukrainian)
7. R.T.MINNIGALIMOV (2015) Analysis and synthesis of Ukrainian speech. Master's thesis: Department of Radio Design and Radio Equipment of the Radio Engineering Faculty, NTU "Igor Sikorsky KPI". 107 p. (in Ukrainian)
8. L.A. BLYZNYCHENKO (1968) Speech intonation. Kyiv: Naukova dumka. 203 p. (in Ukrainian)
9. I.R. GLYUK (2019) Study of acoustic features of the integral characteristics of the Ukrainian language. Master's thesis: Department of Acoustics and Acoustic Electronics of the Faculty of Electronics, NTU "Igor Sikorsky KPI". 91 p. (in Ukrainian)
10. S. Y. YERMOLENKO (1973) Ukrainian literary pronunciation and accent. Dictionary-reference. Academy of Sciences of the Ukrainian SSR Institute of Linguistics named after O. O. Potebny. Kyiv: Naukova dumka. 724 p. (in Ukrainian)
11. V.L. SHEVCHENKO, Y.S. LAZORENKO, O.M. BOROVSKA (2021) Intonational expressiveness of the text at program sounding. Scientific magazine "Problems of programming". (2) P. 76-84. (in Ukrainian)
12. S.V. KNYAZEV, S.K. POZHARYTSKA (2005) Modern Russian literary language. Phonetics, graphics, spelling. Moscow: Academic Project. 426 p. (in Russian)

Одержано 02.09.2022

Про авторів

Лазоренко Яна Сергіївна,

студентка бакалаврату.

Кількість публікацій в українських виданнях – 1.

<http://orcid.org/0000-0002-3987-2338>

Сініцин Ігор Петрович,

доктор технічних наук,

старший науковий співробітник.

Кількість публікацій в українських виданнях – понад 80.

Індекс Хірша – 1.

<https://orcid.org/0000-0002-4120-0784>

Шевченко Віктор Леонідович,
доктор технічних наук, професор.
Кількість публікацій в українських виданнях – понад 300.
Кількість зарубіжних публікацій – 17.
Індекс Хірша – 4.
<https://orcid.org/0000-0002-9457-7454>

Місце роботи авторів:

Інститут програмних систем НАН України,
03187, м. Київ-187,
проспект Академіка Глушкова, 40.
Тел.: (38)(044) 526-21-48
E-mail: ukrprog@isofts.kiev.ua

Київський національний університет імені Тараса Шевченка
Вул. Богдана Гаврилишина, 24, Київ, Україна, 02000.
Тел.: (38)(066) 174-11-34
E-mail: yana_lazorenko@knu.ua

Прізвища та ім'я авторів і назва доповіді англійською мовою:

Lazorenko Ya. S. , Sinitsyn I.P., Shevchenko V.L.
Identification of the native language of a person

Прізвища та ім'я авторів і назва доповіді українською мовою:

Лазоренко Я. С., Сініцин І. П., Шевченко В. Л.
Ідентифікація переважної мови спілкування людини