

Лихошерстов Д.О.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Лебедев Д.Ю.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ СУЧАСНИХ СИСТЕМ РОЗПІЗНАВАННЯ ДАКТИЛЬНОЇ-ЖЕСТОВОЇ МОВИ ДЛЯ СИСТЕМ СУРДОПЕРЕКЛАДУ

У роботі проаналізовані сучасні рішення для визначення жестів при сурдоперекладі жестових мов, котрі базуються на основі розпізнавання жестів за допомогою різноманітних методів і підходів. Проаналізовані рішення, котрі використовують електронні системи на верхніх кінцівках людини. На противагу електронним системам, які носяться на кінцівках, досліджувалися рішення з окремими відеомодулями, котрі передавали відеоряд у програмні додатки обробки даних. Досліджувалася система сурдоперекладу із залученням людей в якості перекладача в реальному часі. Також розглядалися системи сурдоперекладу, що працюють на основі комп'ютерного зору. Сформовані основні, концептуальні недоліки кожного із рішень, які унеможливають інтеграцію досліджуваних систем у побут пересічних користувачів.

Доведено, що зазначені проблеми диктують необхідність створення єдиного програмно-апаратного підходу для систем розпізнавання дактильно-жестової мови під час сурдоперекладу та визначенно шляхи їх удосконалення, зокрема у напрямку спрощення використання. Розроблено удосконалену структурно-функціональну організацію системи розпізнавання дактильно-жестової мови.

На основі проведеного аналізу, запропоновані нові технічні рішення та розроблена блок-схема структури системи сурдоперекладу, що дозволить повноцінно вирішити процеси комунікації людей із вадами слуху та мовлення у суспільстві, а також у майбутньому, дозволить гарантувати людині з вадами слуху, що не потрібно буде проводити хірургічне втручання для встановлення кохлеарного імплантату. Обґрунтовано розширення функціональних можливостей системи розпізнавання жестів та сформовані напрями для їх подальшого вдосконалення.

Ключові слова: сурдопереклад, дактиль, абетка жестів, передача даних, розпізнавання жестів.

Постановка проблеми. За даними Всесвітньої організації охорони здоров'я, понад 430 млн людей у світі страждають на глухоту чи порушення слуху. 165 млн із них – особи у віці старше 65 років, 32 млн – діти молодші за 15 років. За оцінками, до 2050 р. понад 700 млн людей або кожен десятий буде мати проблеми зі слухом [1].

Відмінність людей із вродженими вадами слуху від людей, що втратили слух через різноманітні фактори, полягає у комунікації за допомогою мови жестів. Впорядкований рух руками та головою легко відтворює думки глухого і дана методика є значно зручнішою та швидшою, ніж методика комунікації набором тексту на клавіатурі цифрового пристрою. Складність виникає під час спілкування глухої людини з людиною без вад слуху, коли здорова людина не володіє мовою жестів.

В Україні згідно із різними даними налічується понад 120 тис. людей із порушенням слуху – глухих, котрі у повсякденному житті спілкуються за допомогою мови жестів.

Сучасні реалії України вказують на глибині проблеми комунікації людей із вадами слуху в суспільстві. За даними Всеукраїнської організації інвалідів «Українське товариство глухих», є лише 4 атестовані перекладачі [2] у Волинській області, і це більше ніж на 1 млн жителів [3]. Кількісно неможливо описати доступність об'єктів інфраструктури чи адаптованість середньої спеціальної або вищої освіти для осіб із вадами слуху. Тобто особливі потреби людей із вадами слуху майже повністю не адаптовано у суспільне життя в Україні.

Слід зазначити, що люди із вадами слуху використовують знакову мову спілкування. У сучасному світі існує декілька знакових систем спілкування для глухих:

- жестова мова – повноцінна знакова система й один із трьох засобів комунікації людей із вадами слуху;
- калькуюча жестова мова – будується на основі граматики словесної мови;
- дактильна азбука – кожен жест відповідає буквам алфавіту (рис. 1).



Рис. 1. Українська дактильна абетка

Для української дактильної абетки існують 33 дактильні жести (кожен жест відповідає кожній букві української абетки). Позначення дактилем в українській дактилології передається трьома способами:

- конфігурацією пальців;
- рухом пальців;
- рухом зап'ястя руки.

23 дактилями зображаються лише конфігурацією пальців: А, Б, В, Г, Е, Ж, И, І, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ш, Ч, Ю, Я.

9 дактилем – поєднанням конфігурації пальців і руху зап'ястя руки: Д, Є, З, Ї, Й, К, Ц, Щ, Ь.

Одна дактилема Г – поєднанням статичної конфігурації пальців і руху великого пальця. У роботі Ю.В. Крака, Ю.В. Барчукова, Б.А. Троценко «Побудова моделей дактилем для синтезу дактильної інформації» [4, с. 147] наведено приклад дактилем «Г – Г – Ь», що зображено на рис. 2.

Дактилеми	Зображення
Г – Г – Ь	

Рис. 2. Приклад дактилем «Г – Г – Ь»

Жестами можуть вважатися будь-які рухи та/або зміна положення кінцівок тіла. Вважається, що жестова мова є одним із найбільш популярних підходів у комунікації із людьми, що мають вади слуху.

Лише останніми роками держава, суспільство та наука почали працювати для покращення комунікативних умов життя глухонімих людей. Так, на сайті українського товариства глухих є набір відеороликів, котрі демонстру-

ють, як правильно відтворювати жестами слова української абетки [5].

Зазначені вище проблеми диктують необхідність проведення порівняльного аналізу наявних сьогодні систем розпізнавання дактильно-жестової мови для сурдоперекладу та визначення шляхів їх удосконалення, зокрема у напрямку популяризації та спрощення використання.

Аналіз останніх досліджень і публікацій. У сучасному світі існують різні технології автоматичного розпізнавання мовних жестів. Будемо вважати жестами будь-які рухи або зміни положення частин тіла.

По-перше, необхідно розглянути системи, що використовують додаткові засоби для зчитування мови глухих. Допоміжним засобом може виступати носима електроніка на кінцівках людини для визначення жестів особи із вадами слуху. Так, у праці «Інерційна система розпізнавання жестів» М.О. Ярошенко, А.Ю. Варфоломеев, П.О. Яганов [6] розглядають інерційну систему розпізнавання жестів, котра реалізує процес ідентифікації жестів на основі алгоритму функціонування нейронної мережі BLSTM і використовує рукавицю, що одягається на руку людини.

Подібну концепцію було реалізовано групою студентів Вашингтонського університету у вигляді технологічної рукавиці SignAloud [7]. Набір датчиків у рукавицях фіксують положення і рух рук і надсилають дані через Bluetooth на центральний комп'ютер, який аналізує дані за допомогою різних послідовних статистичних регресій. Коли знайдено збіг із жестом, відповідне слово чи фраза відтворюється через динамік. В обох випадках основна ціль рукавиці – фіксувати положення долоні та пальців і таким чином передавати на обчислювальний комплекс набір жестів, котрі будуть перекладені. Такий підхід дозволяє у відносно простий спосіб визначати набір жестів і миттєво їх перекладати.

Основним недоліками таких рішень є рукавиця – у повсякденному житті людям не завжди зручно носити рукавицю із набором датчиків. Недоліком можна вважати й експлуатацію рукавиці у побуті.

По-друге, потрібно розглянути системи, що виконують переклад тексту чи голосу жестовою мовою за допомогою спеціалізованих додатків. У різних країнах на державному рівні реалізовані додатки, здатні виконувати переклад зі звукової мови знаковою із певними обмеження. Такі програми будемо називати однонаправленими, оскільки переклад виконується лише у напрямку жестової мови.

Як приклад, комп'ютерна програма ProDeaf (WebLibras), розроблена у Португалії у 2010 р., зчитує звуки або текст і перекладає португальською мовою жестів. Основним недоліком цього рішення є однонаправленість – можливість перекладати лише зі слів у жести. Зворотній процес не реалізовано.

В Україні Міністерство юстиції України інтегрує мобільний додаток ConnectPRO, що дозволить людям із порушенням слуху у більш зручний для себе спосіб отримати безоплатну правову допомогу на базі Четвертого київського місцевого центру з надання правової допомоги. Сама система працює у пілотному режимі за таким алгоритмом:

1. Адміністратор викликає перекладача жестової мови через додаток ConnectPRO.
2. Після встановлення відеозв'язку із перекладачем жестової мови адміністратор повертає пристрій до людини з вадами слуху.
3. Людина пояснює проблему жестами – оператор озвучує адміністратору жести та перекладає жестовою мовою слова адміністратора. Алгоритм зображено на рис. 3.



Рис. 3. Алгоритм роботи

Основними недоліками виступає людський фактор. Ніхто не може гарантувати наявність достатньої кількості сурдоперекладачів.

Нині актуальними є системи, які автоматично дозволяють виконувати сурдопереклад. Слід зазначити, що під автоматичним процесом перекладу розуміється фіксація у реальному часі жестів особи з вадами слуху в потоковому відеоряді, певним чином жести ідентифікувати і за допомогою обчислювальних систем ідентифіковані жести інтерпретувати у вигляді словесного/текстового вигляду. Також система повинна бути двонаправленою.

По-третє, розглянемо рішення, котрі використовують методи комп'ютерного зору для зчитування жестів у реальному часі.

Дослідники Китайської Академії наук разом зі спеціалістами Microsoft Research [8] використовують периферійний пристрій, сенсор рухів Microsoft

Kinect Xbox 360, для зчитування мови жестів осіб із вадами слуху, що стоять перед сенсором із подальшим перекладом його у розмовний текст.

Система, реалізована у праці В.Е. Нагапетяна та В.М. Хачумова [9, с. 59–66], працює із використанням сенсора-камери Asus Xtion Pro Live, котра зчитує положення руки та передає набір стопкадрів на обчислювальний комплекс для подальшого перекладу.

Загальним недоліком досліджуваних систем перекладу є відсутність універсальності, потреба у спеціалізованому обладнанні, що унеможливає використання їх у повсякденному житті.

Постановка завдання. На основі проведеного аналізу визначено параметри, що будуть покращуватися:

1. Слід розробити систему, яка не використовує специфічних підходів у реєстрації жестів. Замість камер або датчиків на кінцівках необхідно використовувати камери смартфонів. У повсякденному житті майже кожен має смартфон із камерою.

2. Розробити функціональний додаток, у якому реалізувати повний функціонал системи. Тобто додаток звертається до камери та зчитує потокове відео. Обробляє його, класифікує жести та виконує переклад. Відобразити для користувача переклад жестів.

3. Система повинна працювати й у зворотному напрямку – додаток використовує мікрофон для запису слів із подальшим перекладом жестами, котрі будуть показуватися на екрані смартфона для людини із вадами слуху.

Виклад основного матеріалу дослідження. Проаналізувавши частину наявних рішень, пропонуємо підхід, що дозволить повноцінно вирішити процеси комунікації людей із вадами слуху та мовлення у суспільстві, а також у майбутньому дозволить гарантувати людині з вадами слуху, що не потрібно буде проводити хірургічне втручання для встановленню кохлеарного імплантата.

Для сурдоперекладу потрібно виділити на потоковому відеоряді об'єкт, який ідентифікують як «жест». Потім цей образ потрібно класифікувати. Крім того, у системах реального часу потрібно знати початок і кінець жесту у потоці відеокadrів. Для вирішення всіх складностей виникає потреба використання нейронних мереж.

Рішення базується на використанні смартфонів, на яких буде встановлений додаток сурдоперекладу, розробка якого буде виконана. Додаток залучає камеру смартфона для зчитування жестів. Далі відбувається обробка зчитаних даних на серверній частині та подальше відображення на дисплеї смартфона результатів перекладу.

Розроблювана повноцінна система, що буде функціонувати у реальному часі і здійснювати переклад між:

- об'єкт А – особа, що має вади слуху/мовлення;
- об'єкт Б – здорова особа.

В обох об'єктах чи в одного з об'єктів є смартфон, на якому запущений додаток, що виконує процеси цифрового сурдоперекладу. Кількість смартфонів не є ключовою, і запропонований алгоритм буде працювати на будь-якому пристрої із камерою та дисплеєм. У додатку будуть реалізовані режими роботи, які ґрунтуються на залученні одного або декілька смартфонів.

Слід зазначити, що ця система є двонаправленою, тобто здійснюється переклад того, що сказав об'єкт А до об'єкта Б та у зворотному напрямку. Отже, сам алгоритм можна описати такими процесами (будемо описувати алгоритм перекладу від об'єкта «А» до об'єкта «Б»):

1. Камера направлена на об'єкт А та здійснює фільмування об'єкта А. У реальному часі відеоряд передається на серверну частину, де на вхід нейромережі подається відеоряд.

2. Нейромережа з відеоряду класифікує жести, що відповідають за букви класичного алфавіту. Зображення із кожного кадра передається в таблицю відповідності, де кожному жесту ставиться у відповідність певна літера/слово. Сформований набір перекладеної інформації відправляється на смартфон іншого учасника.

3. Об'єкт Б бачить у додатку на смартфоні у текстовому форматі те, що сказав об'єкт А за допомогою жестів.

Удосконалена структурно-функціональна організація системи розпізнавання дактильно-жестової мови представлена на рис. 4, де літерою «П» позначають скорочення слова «процес», а цифра вказує на номер процесу в алгоритмі перекладу. У цій схемі представлено двох співрозмовників, у кожного є смартфон зі встановленим додатком.

Процес № 1: Відбувається зчитування інформації, котру потрібно буде перекласти. Якщо переклад від об'єкта «А» до об'єкта «Б», то використовуємо камеру смартфона, а якщо навпаки, то мікрофон або текстовий набір на дисплеї смартфона.

Процес № 2: Формування файлу вхідної інформації. Процес необхідний для визначення коректності введених даних.

Процес № 3: Сурдопереклад

– Процес № 3.1 Передача інформації із мобільного додатку на сервер для подальшої обробки може відбуватися за допомогою кодування даних [10, с. 81]. На сервері виконується обробка поточкового відео (із відеоряду визначається набір жестів, що відповідає українській дактильній абетці) та передача набору стоп-кадрів на вхід системи сурдоперекладу.

– Процес № 3.2 Оброблена інформація відправляється на смартфон іншого учасника.

Процес № 4: Виведення перекладеної інформації у додатку.

Процес № 5: Зчитування об'єктом інформації.

Висновки. Проведено аналіз наявних рішень для сурдоперекладу та встановлено, що нині не існує єдиного програмно-апаратного рішення, яке би вирішувало всі проблеми комунікації між особами з вадами слуху.

Наведені приклади систем демонструють, що вони не є універсальними та потребують значних покращень.

На основі порівняльного аналізу визначені основні критерії покращення систем сурдоперекладу, а саме удосконалена структурно-функціональна організація системи для спрощення використання.

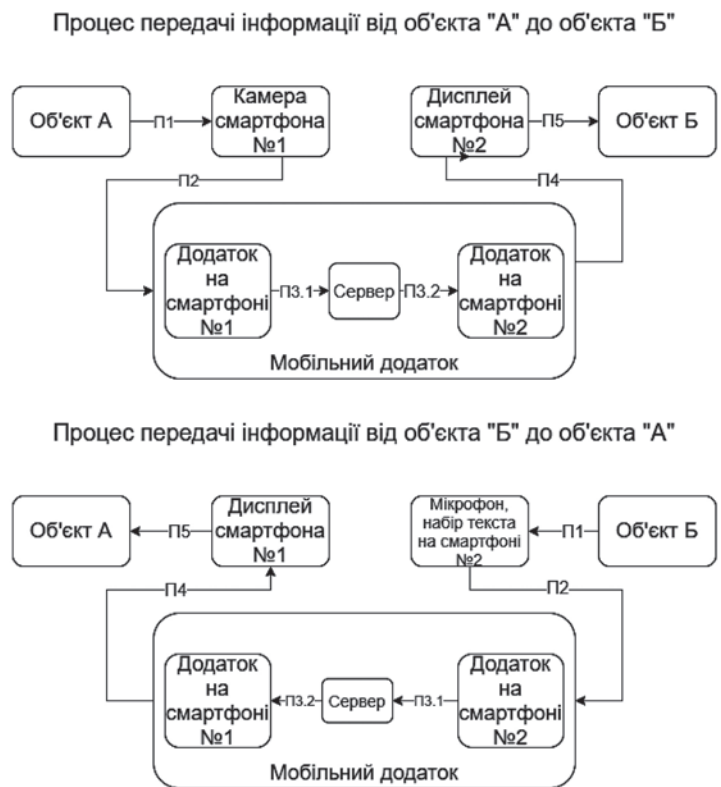


Рис. 4. Удосконалена структурно-функціональна організація системи розпізнавання дактильно-жестової мови

Напрямом подальших досліджень є реалізація запропонованих у роботі удосконалень, направлених на розширення функціональних можливостей систем розпізнавання дактильної-жестової мови для сурдоперекладу з елементами звукового контролю.

Список літератури:

1. Всесвітня організація охорони здоров'я. Deafness and hearing loss. 2021. URL: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
2. Українське товариство глухих. Реєстр перекладачів жестової мови. 2021. URL: <https://utog.org/perekladachi/reestr-perekladachiv>
3. Українське товариство глухих. Чисельність населення на 1 грудня 2020 року. 2021. URL: http://www.lutsk.ukrstat.gov.ua/18_1_11.htm
4. Крак Ю.В., Барчукова Ю.В., Троценко Б.А. Побудова моделей дактилем для синтезу дактильної інформації. *Штучний інтелект*. 2011. № 3. С. 147-155.
5. Про набір жестів в асоціації глухих. URL: <https://utog.org/zhestova-mova/vivchaemo-zhestovu-movu/>
6. Yaroshenko M.O., Varfolomeiev A.Y., Yaganov P.O. Інерційна система розпізнавання жестів. *Мікросистеми, електроніка та акустика*. 2019. № 24 (5). С. 42–47.
7. Рукавиці SignAloud. URL: <https://www.washington.edu/news/2016/04/12/uw-undergraduate-team-wins-10000-lemelson-mit-student-prize-for-gloves-that-translate-sign-language/>
8. Microsoft uses Kinect to interpret sign language from deaf people. URL: <https://www.pcworld.com/article/448415/microsoft-uses-kinect-to-interpret-sign-language-from-deaf-people.html>
9. Нагапетян В.Е., Хачумов В.М. Автоматичне перетворення жестів російської ручної азбуки в текстовий вигляд. *Штучний інтелект і прийняття рішень*. 2013. № 3. С. 59–67.
10. Шаповал І.В., Лебедев Д.Ю. Алгоритм роботи пристрою AES шифратора. *Проблеми інформатизації та управління*. 2016. № 1 (53). С. 87–91.

Lyksherstov D.O., Lebedev D.Yu. ANALYSIS OF MODERN DACTYL-GESTURE SPEECH RECOGNITION SYSTEMS FOR SURDOTRANSLATION WITH ELEMENTS OF SOUND CONTROL

The work analyzes modern solutions for the definition of gestures in sign language, which work on the basis of reading gestures using different approaches. In this paper, a review of solutions that use wearable electronics on the upper extremities. In contrast to wearable electronics, solutions with separate video modules that transmitted video for analyze to data processing software applications. Was studied the system of sign language translation with the involvement of people as a translator in real time. Computer-based sign language translation systems were also considered. Was formed the main, conceptual shortcomings of each of the solutions that make it impossible to integrate research systems into the lives of ordinary users.

It is proved that these problems dictate the need to create a single software and hardware approach for recognition of dactyl-sign language during sign language translation and identify ways to improve them, in particular, in the direction of popularization and simplification of use.

Based on the analysis, proposed ways to improve the technical solutions and developed a block diagram of the structure of the sign language translation system, which will fully address the communication processes of people with hearing and speech impairments in society.

This will guarantee a person with hearing impairments in the future that it will not be necessary to perform operations to install a cochlear implant. The expansion of the functional capabilities of the gesture recognition system is substantiated and vectors are formed for their further improvement.

Key words: sign language translation, dactyl, sign alphabet, data transfer, gesture recognition.