Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)
https://doi.org/10.36930/40360216
Article received: 05.02.2026, accepted: 01.04.2026
Article published: 30.04.2026
UDC 004.[02+4+67+9]
Correspondence author: Yu. Yu. Bilas yurii.y.bilas@lpnu.ua

Ю. Ю. Білас, В. В. Різник
Національний університет "Львівська політехніка", м. Львів, Україна

МЕТОД ГЕНЕРУВАННЯ ШТУЧНОЇ МОВИ НА ОСНОВІ КОМБІНАЦІЇ НАЯВНИХ

Проаналізовано підходи до створення штучних мов (ручні, ШІ-орієнтовані та алгоритмічні) і встановлено, що наявні рішення або потребують значних витрат часу та експертної участі, або демонструють недостатню відтворюваність результатів, або зберігають впізнаваність мови-оригіналу. Розроблено метод генерування тарабарщини на основі комбінації наявних мов із використанням фонетичних даних у форматі IPA. Для побудови методу застосовано багатомовні словники "ipa-dict" (близько 25 мов), очищення транскрипцій від невалідних і просодичних символів, поділ фонемних послідовностей на склади за набором фонотактичних правил та формування частотної бази складів за їх довжиною і позицією у слові (початок, середина, кінець). Наведено механізм детермінованого відображення вхідного складу на згенерований склад із використанням хешування MD5 для ініціалізації псевдовипадкового вибору, що забезпечує відтворюваність перетворення за однакових вхідних даних. Для зниження шуму в частотній моделі відкинуто рідкісні складові шаблони з низькою частотою появи, а генерування цільових складів реалізовано шляхом зваженого вибору з відповідних частотних груп. Проведено експериментальне оцінювання властивостей висловлювань, згенерованих запропонованим методом, за участі 32 респондентів у чотирьох етапах: розпізнавання мови, розпізнавання емоцій у тарабарщині, відмінність штучної мови від справжньої та ідентифікація базової мови. Встановлено, що точність розпізнавання шести базових емоцій у згенерованому мовленні перевищує 80 %, а частка правильного визначення штучної мови в парі з малознайомою природною мовою становить 59,4 %, що свідчить про складність надійного розрізнення штучного й природного мовлення в запропонованій постановці. З'ясовано, що за використання однієї базової мови її фонетичний вплив часто виявляється під час сприйняття, тоді як комбінація багатьох мов істотно ускладнює ідентифікацію мовного джерела. Досліджено швидкодію підсистеми: тривалість генерування тарабарщини приблизно лінійно залежить від тривалості вхідного аудіо та становить близько 0,1 с на 1 с запису (тобто опрацьовує дані швидше, ніж у реальному часі). Запропоновано структуру наскрізної системи перетворення мовлення на штучне зі збереженням емоцій і голосу мовця; у межах цієї роботи реалізовано блоки розпізнавання фонем у мовленні та їх перетворення на тарабарщину, а також експериментально оцінено результати генерування тарабарщини, а інтеграцію TTS і перенесення просодичних характеристик визначено як напрям подальших досліджень.

Ключові слова: штучна мова; сконструйована мова; тарабарщина; синтез мовлення; емоції в мовленні.

Вступ / Introduction

У сучасних системах синтезу мовлення та мультимедійних застосунках важливу роль відіграє можливість генерування мовлення штучними мовами. До таких мов належать конструйовані мови та різні форми тарабарщини (англ. *Conlangs*, *Gibberish*, *Jabberwocky*), які створюються навмисно для використання реальними мовцями або вигаданими персонажами [20].

Технології синтезу мовлення TTS (англ. *Text-To-Speech*) досягли значних результатів для окремих природних мов. Проте в низці застосувань важливішими є не семантичний зміст висловлювання, а емоційне забарвлення голосу або універсальність звучання, незалежного від рідної мови користувача. Це характерно, зокрема, для комунікації з дітьми, звукового дизайну, робототехніки та відеоігор [1, 5, 19]. У таких випадках штучні мови або тарабарщина можуть використовуватися для передачі емоцій без прив'язки до конкретної природної мови.

Використання штучних або частково беззмістовних мов поширене в культурі, медіа та технологіях. У дитячих телепрограмах прикладом є вигадана мова персонажів серіалу "Teletubbies" (1997). У художній літературі та кінематографі створення штучних мов застосовують для формування унікальних світів, зокрема у всесвітах "The Lord of the Rings" (1954), "Star Wars" (1977) та "Avatar" (2009), а також у фільмі "Arrival" (2016). Подібні підходи використовують і в соціальній робототехніці: роботи Probo [19] та Kismet [5] застосовують синтетичне емоційне мовлення або псевдомовні вокалізації для комунікації з людьми. Окрім цього, псевдомови та вигадані тексти використовують в експериментальному мистецтві, наприклад у книзі "Codex Seraphinianus" (1981) або в музичних композиціях, таких як "To Ashes and Blood" (2024). Ці приклади демонструють широкий інтерес до мовлення без чіткого семантичного змісту та вказують на актуальність досліджень у цій галузі.

Створення штучних мов традиційно здійснюється вручну та передбачає розроблення фонетики, граматики

Білас Юрій Юрійович, аспірант, кафедра автоматизованих систем управління.
Email: yurii.y.bilas@lpnu.ua; https://orcid.org/0009-0006-0051-5081
Різник Володимир Васильович, д-р техн. наук, професор, кафедра автоматизованих систем управління.
Email: volodymyr.v.riznyk@lpnu.ua; https://orcid.org/0000-0002-3880-4595
Цитування за ДСТУ: Білас Ю. Ю., Різник В. В. Метод генерування штучної мови на основі комбінації наявних. Науковий вісник НЛТУ України. 2026, т. 36, № 2. С. 144–151.
Citation APA: Bilas, Yu. Yu., & Riznyk, V. V. (2026). A method of generating artificial language based on a combination of existing. *Scientific Bulletin of UNFU*, 36(2), 144–151. https://doi.org/10.36930/40360216

і словника. Такий процес є складним, потребує залучення фахівців-лінгвістів і може тривати місяці [4, 21]. Альтернативою є алгоритмічні методи генерування тарабарщини [24, 25], однак наявні підходи мають низку обмежень. Частина з них орієнтована на збереження впізнаваності базової мови [24], інші обмежуються генеруванням коротких або односкладових слів, потребують ручного редагування або не забезпечують відтворюваності результатів роботи [25].

Питання автоматичного генерування тарабарщини на основі комбінації кількох мов, із маскуванням базової мови та збереженням можливості передачі емоцій у мовленні, залишаються недостатньо дослідженими.

З огляду на це, актуальним є розроблення методів генерування штучної мови на основі комбінації наявних мов, які б забезпечували маскування базової мови, відтворюваність результатів та придатність отриманої тарабарщини для передачі емоцій.

Актуальність дослідження зумовлена потребою у методах генерування штучної мови, придатних для озвучення мультимедійного вмісту, відеоігор, робототехніки та інших застосувань, де важливе передавання емоцій без семантичного змісту, що уможливить зниження витрат на переклад аудіовмісту.

Об'єкт дослідження – генерування тексту штучною мовою (тарабарщиною) та його емоційне сприйняття людиною.

Предмет дослідження – методи та засоби алгоритмічного генерування тарабарщини на основі комбінації наявних мов, що дають змогу маскувати базову мову та зберігати придатність для передачі емоцій у мовленні.

Мета роботи – розробити метод генерування штучної мови на основі комбінації наявних мов формування емоційно виразного безсемантичного мовлення, придатного до використання у мультимедійних системах і зниження витрат на багатомовне озвучення.

Для досягнення зазначеної мети визначено такі основні *завдання дослідження*:

- проаналізувати підходи до створення штучної мови та алгоритмічного генерування тарабарщини, що забезпечить виявлення їхніх переваг, обмежень і невирішених проблем;
- розробити метод генерування тарабарщини на рівні складів IPA та комбінації різних природних мов, що дасть змогу формувати відтворювану штучну мову з маскуванням базової мови;
- провести натурний експеримент із сприйняттям тарабарщини за участі респондентів (у форматі онлайн-опитування: розпізнавання мови, емоцій, відмінність від справжньої мови, ідентифікація базової мови) та описати результати, що уможливить оцінювання практичної придатності запропонованого методу;
- запропонувати структуру наскрізної системи перетворення мовлення природною людською мовою на штучну мову зі збереженням емоцій та голосу мовця, що дасть можливість в майбутньому розробити програмне рішення для здешевлення виробництва аудіовмісту, яке охоплюватиме більшість кроків такого процесу.

Аналіз останніх досліджень та публікацій. *Створення штучної мови вручну.* Зазвичай штучні мови конструюють лінгвісти. В огляді [3] зазначено, що створення художньої мови вимагає опрацювання кожного аспекту – від звукової системи через морфологію та синтаксис до семантики та прагматики. У роботі [15] досліджено вплив звукової структури на сприйняття (наприклад, "оркська" мова в "The Lord of the Rings" сприймалась агресивною). Дж. Р. Р. Толкін створив мови "кенья" та "синдарин" з повною фонетикою та граматикою. Він також створив множину кореневих слів, з яких пізніше утворив похідні слова, також вигадав для носіїв мов історії їх народів, міфологію тощо. Це допомагало автору сформувати бачення того, як його вигадані мови еволюціонували. Толкін також враховував культурні та соціальні аспекти, які впливали на розвиток мов, адже мова – це не тільки засіб спілкування, але й важлива частина ідентичності народу [3]. Вартість створення мови професіоналами є немалою: на популярному вебсайті для творців мов conlang.org середня вартість створення мови станом на 2026 р. становить 5000 USD. Створення штучних мов також часозатратно: мова "Na'vi" для фільму "Avatar" (2009) створювалась місяцями [21], і цікаво, що вона досі розвивається, отримуючи нові слова; мова "Simlish" з серії ігор "The Sims" – близько шести місяців [4]. Отже, ручне створення штучних мов є трудомістким і дорогим задоволенням.

Генерування штучної мови за допомогою ШІ. ChatGPT-4 можна використовувати для створення мов [6], і такі сконструйовані мови навіть підкорятимуться закону Зіпфа (Ципфа), а також частотність слів нагадуватиме частотність слів у природніх і штучних людських мовах; проте такий підхід не забезпечує стабільну відповідність слів і перекладів для текстів великих розмірів.

Алгоритмічне генерування тарабарщини. Семантично вільні висловлювання (засоби слухової взаємодії, які дають змогу виражати емоції та наміри, що складаються з вокалізацій і звуків без семантичного змісту або мовної залежності) поділяють на тарабарщину (англ. *Gibberish*), нелінгвістичні висловлювання (англ. *Nonlinguistic utterances*), музичні висловлювання (англ. *Musical utterances*) та паралінгвістичні висловлювання (англ. *Paralinguistic utterances*). Серед підходів створення тарабарщини виділяють [25]:

- генерування "бурмоковтних" слів (англ. *Jabberwocky*), тобто таких, які не існують, проте відповідають законам мови і є інтуїтивно зрозумілими;
- створення ненаявних односкладових слів (логатом), з допомогою заданих правил англійської мови, які визначають які слова є "легальними" в конкретній людській мові з точки зору орфографії чи фонотактики;
- зміна слів в аудіозаписі з допомогою випадкового склеювання (англ. *Random splicing*), тобто рандомізованого відтворення коротких аудіофрагментів задля того, щоб заплутати зміст мовлення.

З опису авторів роботи [25], перший метод є працезатратним, бо потребує ручної перевірки реалістичності звучання. Недоліком другого методу є його застосовуваність тільки до логатом, а також необхідність складної пост-фільтрації. Третій метод одразу генерує звуки, проте в них легко розпізнати мову-оригінал, чутно акцент, набір доступних звуків є обмеженим тим, що чутно в оригінальному записі, якість аудіозапису сильно впливає на результат, а також, за словами авторів, інколи все ще можна розпізнати лексичний зміст.

У роботі [16] запропоновано алгоритм для генерування рядків висловлювань, які складаються зі слів різної довжини, і кожне з них містить комбінацію відкритих складів типу ПГ або ППГ (П = приголосний, Г = голосний). У роботі [5] запропоновано схожий алгоритм, який створює висловлювання у вигляді рядків різної довжини, випадково підбираючи приголосні та голосні для набору простих форматів складів (Г, ПГ, ППГ, ГГ).

У роботі [9] описано алгоритм, який створює склади довжиною 1-3, і для обирання звуків складу використовує наперед визначені ймовірності того, який тип звуку буде наступним у складі, зважаючи на вже обрані попередні. Важливо зазначити, що, на відміну від попередніх згаданих алгоритмів, цей використовує міжнародний фонетичний алфавіт IPA (англ. *International Phonetic Alphabet*), що матиме дуже різноманітний набір звуків. Недоліками цих двох алгоритмів є обмежений набір допустимих видів складів, а також великі ймовірності, що результат може звучати нереалістично.

Передача емоцій у синтезованому мовленні. У роботі [26] наведено метод вибору одиниць (англ. *Unit Selection*), який передбачає синтез мовлення одночасно з передачею емоцій, обираючи звуки з бази даних, яка відповідає певній емоції. Недоліком цього підходу є потреба у величезній базі даних. У роботі [13] пропонують використовувати приховані моделі Маркова *HMM* (англ. *Hidden Markov Models*) для передачі емоцій під час синтезу мовлення. Даний метод працює з наявними людськими мовами, проте не підходить для штучно сконструйованої мови.

Передача емоцій у штучній мові. У роботі [8] автори використовували випадково обрані просодичні параметри: швидкість мовлення, висота тону та гучність. У роботі [16] використано ці ж самі параметри, а також контур тону, ймовірність наголошення. Недоліками цих підходів є низька реалістичність результатів.

Критичний аналіз і виявлення невирішених проблем. Отже, аналіз відомих досліджень свідчить про те, що наявні підходи до створення штучної мови трудомісткі та дорогі (ті, що ґрунтуються на ручній розробці), або мають обмежену придатність для великих обсягів тексту та слабку відтворюваність результатів (генерування штучної мови за допомогою ШІ), або з обмеженим набором допустимих форм складів чи легкістю розпізнавання мови-оригіналу (алгоритмічне генерування тарабарщини); а методи передачі емоцій зазвичай орієнтовані на справжні мови, потребують великих баз даних, чи мають низьку реалістичність. Водночас, залишається недостатньо дослідженим питання автоматичного генерування тарабарщини на основі комбінації кількох мов із маскуванням базової мови, відтворюваністю результатів та експериментальним оцінюванням придатності для передачі емоцій. Це зумовлює доцільність розроблення методів та засобів алгоритмічного генерування тарабарщини на основі комбінації наявних мов, що і є предметом дослідження даної роботи.

Матеріали та методи дослідження. У роботі застосовано методи аналізу та узагальнення для дослідження підходів до генерування штучних мов і синтезу мовлення. Теоретичною основою є методи генерування штучних мов та стохастичні моделі, що використовуються для моделювання послідовностей складів у форматі IPA. Для побудови моделей використано дані проєкту ipa-dict, що охоплюють близько 25 мов. Оцінювання ефективності запропонованого підходу здійснено експериментально за участі 32 респондентів шляхом онлайн-опитування з подальшим аналізом результатів сприйняття згенерованого мовлення.

Результати дослідження та їх обговорення / Research results and their discussion

Проаналізуємо особливості розроблення наскрізної інформаційної системи перетворення звичайного мовлення на мовлення штучною мовою зі збереженням емоційного забарвлення та голосу мовця. На наведеному нижче рисунку зображено схему концепції такої системи. Система є наскрізною (англ. *End-to-End*), тобто здатна виконати всі завдання процедури перетворення однієї мови (справжньої) на іншу (штучно сконструйовану), не потребуючи при цьому додаткових сторонніх інструментів для, наприклад, синтезу чи перенесення голосу або емоцій, а враховуючи усе, що потрібно для цього. У сучасному світі велика частина технологій, пов'язаних з обробленням мовлення та його синтезом, використовує мову програмування Python через велику наявність прикладних бібліотек. Тож пропонована система не є винятком, і вибір технологій компонентів орієнтований на цю мову.

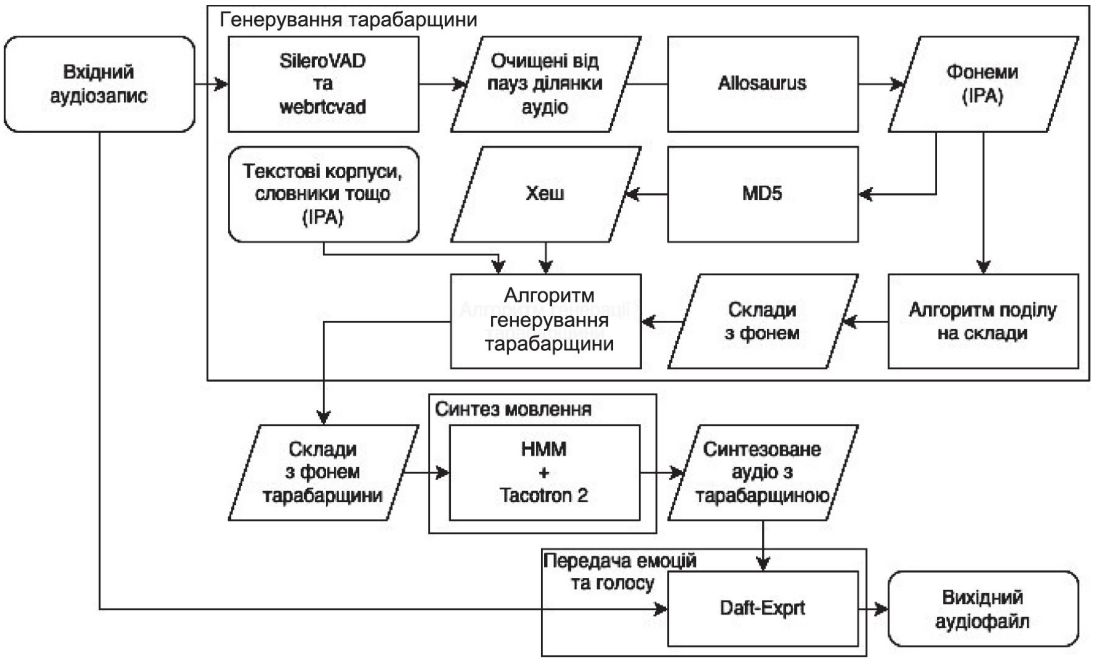

Рисунок. Концептуальна схема інформаційної наскрізної системи перетворення звичайного мовлення на мовлення штучною мовою зі збереженням емоцій та голосу мовця / Conceptual diagram of end-to-end intormational system for converting ordinary speech to artificial speech while preserving the speaker's emotions, voice and rhythm

На рисунку прямокутниками зі заокругленими краями позначено зовнішні вхідні та вихідні дані, паралелограмами – внутрішні дані, які утворюються в межах роботи системи, прямокутниками – програмні компоненти системи, стрілками – напрямок руху даних.

У систему на вхід потрапляє не текст, як зазвичай буває в дотичних дослідженнях, а аудіозапис. Використавши звук як вхідні дані, спрощено систему, уникнувши оброблення текстових даних. Також у вхідному звуці міститься й інша інформація: емоції, голос та ритм, – які потрібні на останніх етапах роботи системи.

У звуковому записі можуть бути паузи з тишею між фразами. Наступний етап розпізнавання фонем для кращих результатів потребує запису без тиші. Тому застосовується тандем технологій "Silero VAD" та "webrtcvad". Вибір припав на них через легке налаштування, гарні результати та підтримку мови програмування Python.

Отримавши звуки, очищені від пауз, виконано розпізнавання фонем. Не розпізнаємо слова, бо це не потрібно, адже тоді система може бути агностичною відносно мови спілкування мовця. Для розпізнавання фонем обрано бібліотеку "Allosaurus" для мови Python. Вона працює з алфавітом IPA. Це є важливою перевагою, адже набір фонем IPA дуже широкий, і це дає змогу генерувати дуже різноманітне мовлення. "Allosaurus" працює дуже швидко, і розпізнає фонеми задовільно (достовірність розпізнавання 80 %, і цього достатньо, адже потім ці фонеми перетворюються на інші, і невеличкі похибки допустимі).

Зважаючи на попередньо виконаний аналіз наявних підходів до генерування штучної мови, обрано алгоритмічний метод через його швидкість, дешевизну та відносну простоту. Схоже до попередньо згаданих авторів робіт [5, 9, 16], прийнято рішення використовувати ймовірнісний підхід з обмеженням структури складів. Проте, на відміну від проаналізованих вище алгоритмів, запропонований в даній роботі підхід не генерує окремі фонеми для складів, а підбирає цілі склади, про що буде згодом нижче.

Отримавши розпізнане мовлення у форматі IPA, відбувається поділ рядка символів на склади. Задача поділу на склади для будь-якої мови є доволі складною, бо в кожній мові ці правила відрізняються. Для простоти, вирішено створити набір простих правил, що базується на правилах, які можна помітити в англійській мові. У майбутньому можна обрати інший підхід, проте в межах цієї роботи зроблено рішення зупинитись на такому варіанті, і результати поділу були достатньо хорошими. Алгоритм працює так:

1. Проходимо по всіх фонемах послідовно.
2. Якщо натрапили на голосну фонему на позиції *i*, то…
2.1. Якщо голосних є дві підряд, то об'єднати їх.
2.2. Перебирати послідовно (індекс *j*) назад фонеми, починаючи з *i*, і щоразу перевіряти, чи виходить приступ (початок складу, англ. onset) з дозволеною структурою. Приступ утворюється з фонем від *j* до *i*-1 включно. Щойно приступ стає недозволеної структури, перебір кандидатів на приступ зупиняється.
2.3. Отриманий приступ видаляємо з останнього складу.
2.4. Новий склад починається з приступу та знайденої голосної фонеми.
3. В іншому випадку – додаємо знайдену фонему до останнього складу.

В алгоритмі згадується "недозволена" та "дозволена" структура приступу складу. Дозволеність визначається простими правилами:

1. Будь-який приступ дозволений на початку складу.
2. Приступ не може бути довшим, ніж 3 фонеми.
3. Приступ не може містити заборонені комбінації:
3.1. Фрикативна приголосна після проривної.
3.2. Проривна приголосна після проривної.
3.3. Проривна приголосна після фрикативної (крім фрикативної "s" на початку приступу).
3.4. Фрикативна приголосна після назальної.
3.5. Назальна приголосна після назальної.
3.6. Проривна приголосна після назальної на початку приступу.
3.7. Проривна приголосна після плавної приголосної.
3.8. Фрикативна приголосна після плавної приголосної.
3.9. "f" після "w".

Для генерування тарабарщини використано багатомовні словники "ipa-dict" (містять транскрипції слів у форматі IPA 25-ти мов). Величезний масив даних очищено від слів, які в поганому форматі, які містять фонеми, що не підтримуються бібліотекою "Allosaurus", а також видаляються просодичні символи. Слова з очищеного набору даних поділяються на склади тим ж алгоритмом, що описано вище. Для отриманих транскрипцій складів трьох груп: на початку, всередині, в кінці слів – збираються частоти. Для уникнення шуму, частоти нижче 10 відкидаються. Після опрацювання, отримано базу даних, яка дає змогу для вказаної довжини складу та його позиції у слові дізнатись його ймовірність появи. Під час генерування тарабарщини для розпізнаних у мовленні складів обираємо відображення у цій БД шляхом зваженого випадкового вибору з відповідної частотної групи. Обрано підхід створення тарабарщини, склеюючи певні готові склади, бо так утворювались слова, що краще читаються (порівняно з тими, що запропоновані у роботах [5, 9, 16]), а також такий метод є набагато простішим у налаштуванні, бо не потребує людської ручної перевірки читабельності складів або складних випадків їх генерування.

Для забезпечення стабільності та відтворюваності результатів використано алгоритм хешування MD5. Хеш отримується з розпізнаного фонетичного складу, і використовується як зерно (англ. Seed) генератора випадкових чисел в алгоритмі генерування тарабарщини. Додатковою важливою перевагою використання хешування є те, що утворена тарабарщина також підкоряється закону Зіпфа, як і справжні людські мови.

Далі, згідно із пропонованою концепцією, виконано процес синтезу мовлення та перенесення емоцій з оригінального запису голосу. Проаналізувавши ряд досліджень із синтезу мовлення [14, 18, 22, 26] та перенесення емоцій [7, 13, 23, 27], обрано тандем HMM з Tacotron 2 [14] та Daft-Exprt [27] відповідно. Вибір припав через їх простоту використання (без додаткового донавчання), якість, а також неймовірну швидкодію. Daft-Exprt, зокрема, є унікальною технологією через те, що може переносити одразу три властивості: емоції, голос та ритм, – навіть якщо при цьому текст мовлення відрізняється.

Проте в межах цієї роботи не проводилось налаштування етапів синтезу мовлення та пересення емоцій (це заплановано у наступних роботах). Натомість, отримавши тарабарщину у форматі IPA, озвучено її з допомогою голосової акторки, якій поставили завдання відтворити просодичні параметри з оригінального запису.

Структура експерименту. Експеримент оформлено у вигляді онлайн-форми з опитуванням, що містить чотири етапи.

Перший етап покликаний зрозуміти, чи учасникам важко помітити штучну мову серед справжніх, а також чи буде штучна мова сприйматись схожою до тієї, на основі якої її сконструйовано. Їм запропоновано 11 аудіозаписів різними мовами, серед яких було чотири штучних мови (на основі англійської, іспанської, французької та комбінації всіх 25-ти мов з "ipa-dict") та поставлено запитання "Яка мова тут використана?". Серед варіантів відповіді запропоновано: ряд справжніх мов, штучна мова та інше. На кожному зі записів промовляється фраза "нехарактерний запах" відповідною мовою. Респондентів повідомлено, що фрази промовлені акторкою без акценту, мови можуть повторюватись, серед них є штучні мови, а також деякі з варіантів відповідей можуть не бути присутніми серед аудіозаписів.

Другий етап дає змогу перевірити, чи підходить тарабарщина для передачі емоцій від аудіозаписів деякою невказаною мовою (насправді це була тарабарщина на основі всіх 25-ти мов з "ipa-dict"). Акторці поставлено завдання передати в цих аудіозаписах шість базових емоцій: здивування, радість, злість, страх, сум, огиду. Учасникам запропоновано як варіанти відповідей ті ж самі шість емоцій. Також їм повідомлено, що емоції трапляються рівно один раз.

Наступний, третій етап, мав ту ж саму мету, що й перший. Учасникам запропоновано прослухати два аудіозаписи. Один з них містив мовлення киргизькою мовою (бо жоден з опитуваних не знає її) та тарабарщиною на основі всіх 25-ти мов з "ipa-dict". Текст, який промовлявся – вірш Тараса Шевченка "І день іде, і ніч іде". Цей вірш є коротким і не містить слів, які могли б підказати слухачам про тему повідомлення (наприклад, слів "кобзар", "Вкраїна" тощо). Користувачам повідомлено, які саме мови тут є, проте записи були розміщені у випадковому порядку, згенерованому генератором випадкових чисел. Завдання респондентів – розпізнати, який із записів – реальна мова, а який – штучна, створена алгоритмічно.

Останній, четвертний етап, дав змогу зрозуміти, чи легко виявити базову справжню мову, на основі якої сконструйовано штучну, а також чи можливо замаскувати її, використавши за основу комбінацію з кількох мов. Учасникам запропоновано прослухати чотири аудіозаписи штучними мовами на основі: англійської, іспанської, французької та комбінації з 25-ти мов з "ipa-dict". Серед варіантів для вибору були тільки справжні мови. Учасників повідомлено, що всі мови в аудіозаписах є штучними, проте побудовані на основі справжніх мов, і поставлено завдання спробувати розпізнати, на основі якої мови сконструйований кожен з варіантів.

Аналіз отриманих результатів експерименту. За тиждень зібрано 32 відповіді учасників. Усі вони родом з України. Серед них 11 жінок та 21 чоловік. Вік градіюється від 15 до 49 років. Більшість з них є теперішніми або колишніми студентами, колегами та товаришами авторів роботи. Майже всі (81 %) опитувані знають англійську мову хоча б на базовому рівні. Значна частина знає основи польської мови (34 %) та іспанської (19 %).

Результати першого етапу експерименту можна побачити в табл. 1. З них видно, що українську мову можуть розпізнати усі учасники, адже це їх рідна мова. Значна частина опитуваних змогла розпізнати англійську через її популярність. Цікаво також, що багато хто розпізнав арабську, незважаючи на те, що жоден з респондентів не знає її. Це свідчить про особливість її фонологічної системи. Схожа ситуація з португальською та турецькою. А з данською мовою стався конфуз: учасники часто сплутували її з німецькою. Ймовірно, це пов'язане з тим, що в аудіозаписі використано слово "lugt", яке є також в німецькій мові, яка більш знайома слухачам. Стосовно ж штучних мов, то "штучно-французька" сприймалась як її базова – французька, "штучно-іспанська" – як португальська (дуже схожа на іспанську), "штучно-англійська" – як англійська. А от штучна мова на основі комбінації багатьох мов часто здавалась схожою слухачам на данську та польську (хоча польської та жодної слов'янської мови не було використано для побудови цієї штучної мови). Фраза штучною мовою звучала як */jatanatanska stɔkna/* і закінчується на "-ска", що часто трапляється у слов'янських мовах (хоча жодна з мов цього регіону не присутня в "ipa-dict"), що, ймовірно, й спричинило такий результат. Також звертаємо увагу на те, що частка людей, які сприйняли штучну мову, як штучну, є маленькою, а також вона не сильно відрізняється від частки людей, які сприйняли штучну мову, як її базову мову (наприклад, для "штучно-англійської" це 28,1 % проти 25 %, що є дуже маленьким розривом, зважаючи на те, що англійську мову респонденти чують дуже часто і вважають, що можуть розпізнати).

Табл. 1. Відповіді учасників на запитання "Яка мова тут використана?" /
Answers of survey participants on question "What language is being used here?"

	uk	en	ar	da	ko	pt	tr	fr	it	de	pl	gi
uk	100	0	0	0	0	0	0	0	0	0	0	0
en	3,1	31,3	6,3	6,1	0	6,3	6,3	3,1	15,6	15,6	0	9,4
ar	0	0	43,8	3,1	0	3,1	31,3	0	0	3,1	3,1	12,5
da	0	0	0	15,6	3,1	6,3	6,3	6,3	3,1	25	6,3	28,1
ko	0	0	21,9	3,1	28,1	6,3	25	0	0	0	0	15,6
pt	3,1	0	6,3	12,5	9,4	28,1	6,3	0	9,4	0	12,5	12,5
tr	3,1	3,1	6,3	6,3	3,1	0	21,9	6,3	6,3	0	3,1	40,6
gi-fr	0	0	9,4	18,8	3,1	6,3	12,5	28,1	0	0	0	21,9
gi-es	0	0	0	6,3	3,1	25	12,5	3,1	15,6	3,1	0	31,3
gi-en	0	25	0	6,3	9,4	9,4	3,1	3,1	6,3	3,1	6,3	28,1
gi-all	3,1	0	3,1	18,8	6,3	12,5	3,1	0	0	6,3	21,9	25

Ще хочемо вказати, що один з учасників експерименту, який щодня працює з різними мовами та прослуховує мовлення ними, поділився, що йому було доволі складно визначити мову, адже в аудіозаписах не було акценту. Цей цінний відгук допомагає зрозуміти, що акцент бере велику участь у розпізнаванні мови. З цього етапу робимо висновок, що мови мають певні характеристики фонологічної системи, які їх вирізняють з-по-

між інших, і, якщо мова побудована на основі справжньої мови, то вона переймає ці характеристики, і теж часто сприймається слухачами, як її базова. Якщо ж штучна мова побудована на комбінації з багатьох мов, то її важко сприйняти, як якусь конкретну, а вона більше сприймається як якась незнайома ("gi") або її статистика сприйняття наближена до рівномірного розподілу серед мов-кандидатів.

У табл. 1 заголовки колонок – це вибір учасників, водночас як заголовки рядків – це мова, яку дійсно було використано в аудіозаписі, а на перехрещенні – частка учасників, які розпізнали мову, вказану у колонці, в аудіозаписі з мовою, вказаною у рядку; у заголовках та рядках вказано коди мов: uk – українська, en – англійська, ar – арабська, da – данська, ko – корейська, pt – португальська, tr – турецька, fr – французька, it – італійська, de – німецька, pl – польська, gi – штучна, gi-fr – штучна на основі французької, gi-es – штучна на основі іспанської, gi-en – штучна на основі англійської, gi-all – штучна на основі 25-ти мов з "ipa-dict"; жирним позначено клітинки, які відповідають тій самій мові в дійсності та в сприйняті учасників.

У табл. 2 наведено результати другого етапу опитування. Результати чітко кажуть, що слухачі дуже добре сприймають емоції, незважаючи на те, що мова – штучна. Це доводить, що штучні мови чудово справляються у завданнях, де потрібно передати емоції, а не семантичний сенс повідомлення.

Табл. 2. Відповіді учасників на запитання "Яка емоція тут використана?" / Answers of survey participants on question "What emotion is being used here?"

	Злість	Здивування	Огида	Радість	Страх	Сум
Злість	78,1	6,3	9,4	3,1	3,1	0
Здивування	3,1	96,9	0	0	0	0
Огида	0	0	81,3	3,1	0	15,6
Радість	0	3,1	0	75	18,8	3,1
Страх	6,3	0	0	9,4	81,3	3,1
Сум	3,1	0	12,5	3,1	0	81,3

У табл. 2 заголовки колонок – це вибір учасників, водночас як заголовки рядків – це емоція, яку дійсно було використано в аудіозаписі, а на перехрещенні – частка учасників, які розпізнали емоцію, вказану у колонці, в аудіозаписі з емоцією, вказаною у рядку.

За результатами третього етапу, 59,4 % слухачів правильно вказали, де є штучна мова, а де справжня (киргизька). 40,6 % слухачів обрали протилежну відповідь. Розподіл дуже близький до 50:50, що свідчить про те, що слухачам доволі важко визначити, яка мова є штучною, а яка просто незнайомою. З іншого боку, є невеличкий нахил у бік успішного розпізнавання мови, що є показником того, що є куди покращуватись, і потрібно попрацювати над правилами побудови слів, аби вони звучали краще. Для цього потрібно провести серії додаткових експериментів зі справжніми слухачами, щоб своїми відповідями вони допомогли нам налаштувати підсистему до ще кращого рівня.

Результати останнього етапу наведено в табл. 3. З них робимо висновки, що якщо штучна мова побудована на певній мові, то вона з великою ймовірністю буде відчуватись саме, як ця мова, або споріднені їй. Особливим також є результат 40,6 % сприйняття комбінації різних мов, як польську. Але, як вже було зазначено раніше, це може бути пов'язано з тим, що згенероване слово містило суфікс, притаманний слов'янським мовам. Проте, незважаючи на це, слухачі обрали багато інших мов-кандидатів.

Табл. 3. Відповіді учасників на запитання "Яка мова є базовою для тарабарщини з цього запису?" / Answers of survey participants on question "What language has been used as basis for gibberish from this audio recording?"

	Іспанська	Французька	Англійська	Комбінація
Українська	3,1	0	0	3,1
Англійська	0	0	40,6	0
Арабська	15,6	3,1	6,3	6,3
Данська	6,3	12,5	12,5	21,9
Італійська	25	3,1	9,4	6,3
Іспанська	28,1	6,3	0	0
Корейська	3,1	3,1	12,5	0
Німецька	0	0	0	6,3
Польська	3,1	0	6,3	40,6
Турецька	6,3	28,1	9,4	9,4
Французька	0	37,5	0	3,1
Інше	9,4	6,3	3,1	3,1

У табл. 3 заголовки колонок – це базова мова, на основі якої згенерована тарабарщина, водночас як заголовки рядків – це мова, яку, на думку слухачів, було використано для генерування тарабарщини, а на перехрещенні – частка учасників, які впізнали мову, вказану у рядку, в аудіозаписі з базовою мовою, вказаною у колонці.

Результати швидкодії. Тривалість генерування тарабарщини лінійно залежить від тривалості вхідного аудіо; генерування приблизно вдесятеро швидше за реальний час (приблизно 0,1 с на 1 с аудіо на системі Intel Core i5 11-го покоління, 8 GB RAM).

Обмеження методу. Метод працює тільки на фонетичному/складовому рівні: морфологія, синтаксис і семантика не моделюються. Поділ на склади виконано за базовими правилами, орієнтованими на англійську; для інших мов можливі помилки.

Обговорення результатів дослідження. У роботі [9] досліджено сприйняття людиною штучного мовлення, згенерованого на основі фонем IPA у вигляді тарабарщини, зокрема вплив просодії (яка визначалась трьома параметрами синтезатора, значення яких обиралось випадково) та фонетичного складу на емоційне сприйняття. Результати експерименту показують відсутність універсальних міжкультурних закономірностей у сприйнятті таких параметрів, що ускладнює побудову універсальних моделей синтезу мовлення та вказує на необхідність гнучких підходів до генерування.

У роботі [11] наведено бібліометричний аналіз досліджень у галузі штучних мов, який демонструє зростання інтересу до цієї тематики, особливо у контексті мистецтва та експериментальної лінгвістики. Встановлено, що більшість досліджень орієнтовані на навчання мов та лінгвістичні експерименти, тоді як прикладні методи автоматизованого генерування мов залишаються менш дослідженими.

У роботі [12] запропоновано генератор штучних мов на основі методів оброблення природної мови, зокрема n-грам, рекурентних нейронних мереж та attention-моделей. Автори вказують на важливість збереження семантичної структури та правдоподібності мови, але зазначають складність налаштування таких систем і необхідність значних обчислювальних ресурсів. Також пропонований метод орієнтований на попереднє навчання для однієї конкретної мови.

У дослідженні [2] наведено підхід до генерування штучних мов із використанням LLM у вигляді багатокрокового конвеєра (фонологія, морфологія, синтаксис, лексика). Такий підхід дає змогу отримувати узгоджені та різноманітні мови, проте залежить від складних пропрієтарних моделей (які мають платню за використання), не гарантує повної відтворюваності результатів, не працює в режимі реального часу, а також не працює без наявного інтернет-з'єднання.

У навчальному посібнику [17] та книзі [10] детально описано класичний процес створення штучної мови, враховуючи поетапний вибір фонетики, граматики та лексики. Ці підходи є ефективними для створення повноцінних мов, проте автори вказують на складність і трудомісткість ручного створення мов, що потребує значних експертних знань і часу. Це підтверджує доцільність автоматизованих підходів, таких як запропонований у даній роботі метод генерування тарабарщини.

Отже, обговорення результатів дослідження підтверджує доцільність проведення цієї роботи, оскільки в проаналізованих джерелах недостатньо наведено підходи до генерування тарабарщини, що поєднують багатомовні фонетичні дані з детермінованими механізмами побудови мовлення. Наявні рішення або орієнтовані на ручне конструювання мов, або використовують складні моделі штучного інтелекту, що не забезпечують відтворюваності результатів та контролю над фонетичною структурою, а також не працюють в режимі реального часу. Отримані результати частково заповнюють цю прогалину, демонструючи можливість побудови штучної мови на основі комбінування наявних мов зі збереженням емоційної виразності мовлення та маскуванням базової мови.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – розроблено метод генерування штучної мови на основі комбінування фонетичних одиниць кількох природних мов у форматі IPA з використанням детермінованого відображення складів, який забезпечує одночасно відтворюваність отриманих результатів, маскування базової мови та збереження фонетичної правдоподібності мовлення, що дає змогу формувати варіативну тарабарщину без прив'язки до конкретної мови.

Практична значущість результатів дослідження – можна застосувати запропонований метод у системах синтезу мовлення, відеоіграх, освіті, терапії, робототехніці та мультимедійних застосунках для створення штучного мовлення без семантичного змісту, але з передаванням емоцій, що дасть змогу зменшити витрати на переклад аудіоконтенту та забезпечити універсальність взаємодії з користувачем незалежно від його рідної мови.

Висновки / Conclusions

Розроблено метод генерування штучної мови на основі комбінації наявних мов формування емоційно виразного безсемантичного мовлення, придатного до використання у мультимедійних системах і потенційного зниження витрат на багатомовне озвучення. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Проаналізовано підходи до створення штучних мов (ручні, ШІ-орієнтовані та алгоритмічні) і встановлено їхні ключові обмеження: значну трудомісткість/вартість, недостатню відтворюваність або обмежену керованість фонетичною структурою. Це підтвердило доцільність розроблення детермінованого алгоритмічного підходу на фонетичному рівні.
2. Розроблено метод генерування тарабарщини на рівні складів IPA на основі комбінації даних проєкту "ipadict" (близько 25 мов), поділу на склади та частотного відображення складів за їх довжиною і позицією у слові (початок/середина/кінець). Використання MD5 для ініціалізації вибору забезпечило відтворюваність результатів і стабільність перетворення вхідних складових послідовностей.
3. Проведено експеримент із 32 учасниками у чотирьох етапах (розпізнавання мови, розпізнавання емоцій, відмінність штучної мови від справжньої, ідентифікація базової мови) та встановлено: точність розпізнавання базових емоцій у тарабарщині перевищує 80 %, а розпізнавання штучної мови у парі з маловідомою слухачам мовою становить 59,4 %. Для комбінації багатьох мов спостерігається розмиття ідентифікації базової мови, тоді як для одноосновних варіантів вона частіше вгадується.
4. Запропоновано структуру наскрізної системи перетворення звичайного мовлення на штучне зі збереженням емоцій та голосу мовця (розпізнавання фонем, генерування тарабарщини, синтез мовлення, перенесення просодичних ознак). У межах цієї роботи реалізовано блоки розпізнавання фонем та їх перетворення тарабарщини, і експериментально оцінено другий з них – блок генерування тарабарщини; інтеграцію TTS і перенесення емоцій/голосу заплановано на наступні етапи.
5. Встановлено, що підсистема генерування тексту штучною мовою працює швидше за реальний час (1 с аудіо опрацьовується за приблизно 0,1 с), що підтверджує практичну придатність підходу для мультимедійних застосувань і створює підстави для подальшого розширення можливостей системи.

References

1. Akyazı, K. G., & Baştemur, Ş. (2024). Interactive Robots: Therapy Robots. *Psikiyatride Güncel Yaklaşımlar – Current Approaches in Psychiatry*, *16*(1), 16–30. https://doi.org/10.18863/pgy.1242958
2. Alper, M., Yanuka, M., Giryes, R., & Begus, G. (2025). ConlangCrafter: Constructing Languages with a Multi-Hop LLM Pipeline.arXiv *preprint*. https://doi.org/10.48550/arXiv.2508.06094
3. Beinhoff, B. (2023). Design intentions and actual perception of fictional languages: Quenya, Sindarin and Navi. In: I. Noletto, J. Norledge, & P. Stockwell (Eds.), *Reading Fictional Languages* (pp. 76–92). Edinburgh University Press. https://doi.org/10.1515/9781399529167
4. Bracchi, L. (2023). Invented languages and their reflections on the world of video games. *Thesis for: Languages, Literatures and Intercultural Studies*, pp. 1–34 https://doi.org/10.13140/RG.2.2.14042.31687
5. Breazeal, C. L. (2000). Sociable machines: Expressive social exchange between humans and robots. *PhD thesis, Massachusetts Institute of Technology, Cambridge, United States*. order number AAI0801833. URL: https://dl.acm.org/doi/10.5555/932790
6. Diamond, J. (2023). "Genlangs" and Zipfs Law: Do languages generated by ChatGPT statistically look human? *arXiv preprint*. https://doi.org/10.48550/arXiv.2304.12191
7. Dutta, S., & Ganapathy, S. (2024). Zero Shot Audio To Audio Emotion Transfer With Speaker Disentanglement. *ICASSP 2024 – 2024 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 10371–10375. https://doi.org/10.1109/ICASSP48485.2024.10445962
8. Gonzalez, A. G. C., Lo, W., & Mizuuchi, I. (2022). Talk to Kotaro: a web crowdsourcing study on the impact of phone and prosody choice for synthesized speech on human impression. *2022*

31st IEEE International Conference on Robot and Human Interactive Communication (*RO-MAN*), pp. 244–251. https://doi.org/10.1109/RO-MAN53752.2022.9900685
9. Gonzalez, A. G. C., Lo, W.-S., & Mizuuchi, I. (2023). The Impression of Phones and Prosody Choice in the Gibberish Speech of the Virtual Embodied Conversational Agent Kotaro. *Applied Sciences*, *13*(18), article ID 10143. https://doi.org/10.3390/app131810143
10. González, C. (2025). Inventing Languages: An Introduction to Constructed Languages. *Cambridge: Cambridge University Press*, 402 p. https://doi.org/10.1017/9781108864015
11. Gonzalez, S. (2024). A Network Analysis Approach to Conlang Research Literature.arXiv *preprint*. URL: https://arxiv.org/abs/2407.15370
12. Heyer, F. (2021). Generating Immersive Conlangs. *Thesis for B. Sc. Computer Science, Kiel University, Kiel, Germany*, pp. 1–26. https://doi.org/10.13140/RG.2.2.22160.28168
13. Liu, R., Şişman, B., & Li, H. (2021). Reinforcement Learning for Emotional Text-to-Speech Synthesis with Improved Emotion Discriminability. *Interspeech 2021*, 4648–4652. https://doi.org/10.21437/Interspeech.2021-1236
14. Mehta, S., Székely, É., Beskow, J., & Henter, G. E. (2022). Neural HMMs Are All You Need (For High-Quality Attention-Free TTS). *ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 7457–7461. https://doi.org/10.1109/ICASSP43922.2022.9746686
15. Mooshammer, C., Bobeck, D., Hornecker, H., Meinhardt, K., Olina, O., Walch, M. C., & Xia, Q. (2024). Does Orkish Sound Evil? Perception of Fantasy Languages and Their Phonetic and Phonological Characteristics. *Language and Speech*, *67*(4), 961–1000. https://doi.org/10.1177/00238309231202944
16. Oudeyer, P.-Y. (2003). The production and recognition of emotions in speech: Features and algorithms. *International Journal of Human-Computer Studies*, *59*(1–2), 157–183. https://doi.org/10.1016/S1071-5819(02)00141-6
17. Peterson, J. (2025). How to Create a Language: The Conlang Guide. *Cambridge: Cambridge University Press*, 420 p. https://doi.org/10.1017/9781108991827
18. Ping, W., Peng, K., Gibiansky, A., Arik, S. O., Kannan, A., Narang, S., Raiman, J., & Miller, J. J. (2017). Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning. *6th International Conference on Learning Representations* (*ICLR 2018*), pp. 1–16. https://doi.org/10.48550/arXiv.1710.07654
19. Saldien, J., Goris, K., Yilmazyildiz, S., Verhelst, W., & Lefeber, D. (2008). On the Design of the Huggable Robot Probo. *Journal of Physical Agents*, *2*(2), 3–11. https://doi.org/10.14198/JoPha.2008.2.2.02
20. Schreyer, C. (2021). Constructed Languages. *Annual Review of Anthropology*, 50, 327–344. https://doi.org/10.1146/annurev-anthro-101819-110152
21. Suprycheva, I. (2022). Linguistic Creativity in Cinema: A Case Study of the Navi Language. *Three year degree thesis, University of Padua*, pp. 1–64. URL: https://thesis.unipd.it/handle/20.500.12608/60158
22. Vainer, J., & Dušek, O. (2020). SpeedySpeech: Efficient Neural Speech Synthesis. *Interspeech 2020*, 3575–3579. https://doi.org/10.21437/Interspeech.2020-2867
23. van Niekerk, B., Carbonneau, M.-A., & Kamper, H. (2023). Rhythm Modeling for Voice Conversion.arXiv *preprint*. https://doi.org/10.48550/arXiv.2307.06040
24. Yilmazyildiz, S., Latacz, L., Mattheyses, W., & Verhelst, W. (2010). Expressive Gibberish Speech Synthesis for Affective Human-Computer Interaction. *Lecture Notes in Computer Science*, 6231, 584–590. https://doi.org/10.1007/978-3-642-15760-8_74
25. Yilmazyildiz, S., Read, R., Belpeame, T., & Verhelst, W. (2016). Review of Semantic Free Utterances in Social Human-Robot Interaction. *International Journal of Human-Computer Interaction*, *32*(1), 63–85. https://doi.org/10.1080/10447318.2015.1093856
26. Yin, Z. (2018). An Overview of Speech Synthesis Technology. *2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control* (*IMCCC*), pp. 522–526. https://doi.org/10.1109/IMCCC.2018.00116
27. Zaïdi, J., Seuté, H., van Niekerk, B., & Carbonneau, M.-A. (2022). Daft-Exprt: Cross-Speaker Prosody Transfer on Any Text for Expressive Speech Synthesis. *Interspeech 2022*, 4591–4595. https://doi.org/10.21437/Interspeech.2022-10761

Yu. Yu. Bilas, V. V. Riznyk

Lviv Polytechnic National University, Lviv, Ukraine

A METHOD OF GENERATING ARTIFICIAL LANGUAGE BASED ON A COMBINATION OF EXISTING

Approaches to artificial language creation (manual, AI-driven, and algorithmic) were analyzed, and their key limitations were identified: high labor and expert involvement, limited reproducibility, or insufficient masking of the source language in generated output. A method for gibberish generation based on a combination of existing languages was developed using phonetic data in IPA (International Phonetic Alphabet). The method relies on multilingual "ipa-dict" resources (about 25 languages), cleaning of transcription entries, syllabification of phoneme sequences with rule-based phonotactic constraints, and construction of syllable-frequency groups by syllable length and word position (initial, medial, final). A deterministic mapping mechanism from an input syllable to an output syllable was introduced using MD5 hashing to initialize pseudo-random selection, which ensures reproducible transformation for identical inputs. To reduce noise, low-frequency syllable patterns were filtered out, and target syllables were generated through weighted sampling from the corresponding positional frequency groups. An experimental evaluation of properties of utterances generated by proposed method with 32 participants was conducted in four stages: language recognition, emotion recognition in gibberish, discrimination between artificial and natural speech, and base-language identification. The results showed that recognition accuracy for six basic emotions in gibberish exceeded 80 %, confirming practical suitability for affective speech scenarios where semantic content is not required. The share of correct artificial-language detection in a pair with a less familiar natural language was 59.4 %, indicating that robust discrimination between generated and natural speech remains difficult in this setup. It was also found that when gibberish is generated from a single base language, listeners more often infer that linguistic origin, whereas combining many languages substantially blurs base-language identification. Runtime performance was investigated and showed approximately linear dependence on input duration, with generation speed (of gibberish) around 0.1 s per 1 s of audio (faster than real time). In addition, a structure of an end-to-end speech conversion system to artificial speech while preserving speaker emotion and voice was proposed. Within this study, only the phonemes detection and their transforming into gibberish blocks were implemented and gibberish generation block was evaluated; integration with TTS and explicit prosody-transfer components is defined as future work.

Keywords: artificial language; conlang; gibberish; speech synthesis; emotions in speech.