

УДК 004.451**DOI: 10.31866/2617-796x.3.1.2020.206107****Ткаченко Олександр,**

кандидат фізико-математичних наук,
доцент кафедри інженерії програмного забезпечення,
Національний авіаційний університет,
Київ, Україна
aatokg@gmail.com
<https://orcid.org/0000-0001-6911-2770>

Ткаченко Костянтин,

кандидат економічних наук,
старший викладач кафедри інформаційних технологій,
Державний університет інфраструктури та технологій,
Київ, Україна
tkachenko.kostyantyn@gmail.com
<https://orcid.org/0000-0003-0549-3396>

Щербатюк Владислав,

магістрант кафедри інформаційних технологій,
Державний університет інфраструктури та технологій,
Київ, Україна
vladscherbatiuk@gmail.com
<https://orcid.org/0000-0002-4592-6400>

ДЕЯКІ ПРОБЛЕМИ ТА ПЕРСПЕКТИВИ РОЗРОБКИ КРОСПЛАТФОРМНИХ ГОЛОСОВИХ СИНТЕЗАТОРІВ

Метою статті є дослідження і розгляд загальних проблем і перспектив розробки голосових синтезаторів на основі їхньої кросплатформності, опис та аналіз процесів розробки авторської програми синтезу мови, яка може підтримувати роботу на максимальній кількості систем, спроможних до звуковідтворення та незалежних від типу операційної системи.

Методами дослідження є методи семантичного аналізу основних понять розглянутої предметної сфери (голосові синтезатори). У статті розглянуто підходи до розробки кросплатформних голосових синтезаторів.

Новизною проведеного дослідження є аналіз математичного апарату, що може застосовуватися для синтезу голосу (голосних, різних типів приголосних звуків), та опис деяких аспектів розробки власної авторської системи, що синтезує голос людини.

Висновки. Переважна більшість програм голосового синтезу розроблена спеціально для визначених типів комп'ютерів або прив'язана до конкретної операційної системи. У роботі досліджено голосовий синтезатор, визначено його математичне обґрунтування та перспективи кросплатформності.

Ключові слова: синтез голосу; модулятор; демодулятор; цифрова модуляція; інформаційно-технологічна підтримка синтезу голосу; голосовий синтезатор; кросплатформний голосовий синтезатор.

Вступ. Голос людини складається зі звуків, що утворюються під час відкриття та закриття глотки голосовими зв'язками, що фізично створює періодичну форму хвилі з багатьма гармоніками (https://www.porticus.org/bell/belllabs_kitsss.html; Изотов, 2015). Потім цей основний звук фільтрується носом і горлом (складна резонансна трубопровідна система), щоб керувати різницею вмісту гармонік (формантів), створюючи широкий спектр звуків, які необхідні для вимови тих чи тих слів, лексем тощо. Існує ще один набір звуків, відомий як беззвучні звуки, які створюються або модифікуються по-різному.

На сьогодні є великий вибір програм, які пропонують синтез голосу, але переважна кількість з них реалізована для однієї чи декількох операційних систем. Більшість програм, що підтримують голосовий синтез, покладається на online-сервіси обчислення даних через центральні сервери та є сумісними тільки з певними наборами інструментів розробки. Особливістю таких програм є й те, що вони мають переважно закритий вихідний код.

Результати дослідження. Серед поширених методів синтезу голосу є підхід, при якому синтезатор мови аналізує, вимірюючи зміни в її спектральних характеристиках, які відбуваються в часі (https://uk.wikipedia.org/wiki/Синтез_мовлення). Усе це призводить до серії сигналів, що представляють такі змінені частоти в будь-який конкретний момент часу, коли користувач говорить (вимовляє слова). Інакше кажучи, сигнал розбивається на низку частотних смуг (чим більша ця кількість, тим точнішим є аналіз), а рівень сигналу, що присутній у кожному діапазоні частот, забезпечує миттєве подання вмісту спектральної енергії.

Щоб відтворити мовлення, синтезатор мови (голосовий синтезатор) створює зворотний процес, обробляючи широкочастотні джерела шуму, пропускаючи їх через етап, на якому здійснюється фільтрація вмісту частоти на основі унікальних спеціально сформованих числових рядів (https://uk.wikipedia.org/wiki/Синтез_мовлення).

У звуковому декодері вхідні дані передаються через фільтр з багатьма діапазонами, після цього кожна смуга частотного сигналу пропускається через модулятор так званих «огинаючих» частот, а сигнали управління від модулятора частот передаються в декодер. Декодер застосовує ці амплітудні керівні сигнали до відповідних підсилювачів каналів фільтра для повторного синтезу голосу (звуку, інколи шуму).

Інформація про миттєву частоту вихідного голосового сигналу (на відміну від його спектральної характеристики) відкидається. Ця характеристика має невисокий рівень важливості для забезпечення суто оригінального призначення використання синтезатора як допоміжного засобу передачі інформації. Саме цей аспект процесу синтезу мови, що відбувається у голосовому синтезаторі, визначив високий рівень його корисності під час створення спеціальних голосових ефектів

у різних сферах культури, зокрема в попкультурі, музиці, аудіорозвагах та аудіо-навчанні.

Процеси синтезатора формують та надають лише параметри голосової моделі по каналу зв'язку, а не відтворюють окремі елементи (точки) звукового сигналу. Через те, що параметри, які змінюються, порівнюються з вихідною формою хвилі мовлення не миттєво, а повільно, пропускна здатність, необхідна для передачі мови, може бути зменшена (https://www.porticus.org/bell/belllabs_kitsss.html; Изотов, 2015). Це дає змогу більшій кількості каналів, по яким передається звук (голос користувача, звуки природи, звуки музики тощо), використовувати прості канали зв'язку, до яких, зокрема, можна зарахувати радіосигнал, одноконтактний кабель.

Звукові синтезатори умовно можна поділити на аналогові та цифрові (https://uk.wikipedia.org/wiki/Синтез_мовлення). *Аналогові синтезатори* аналізують вхідний сигнал, розбиваючи сигнал на декілька частотних каналів чи діапазонів частот. Модуляційний сигнал та сигнал-носіє передаються через низку цих налаштованих смугових фільтрів. На прикладі типового голосу робота модуляційний сигнал – це мікрофон, а сигнал-носіє – шум або пилоподібний хвильовий сигнал (https://www.porticus.org/bell/belllabs_kitsss.html). На практиці зазвичай використовують 8 або 12 каналів (Изотов, 2015).

Амплітуда модулятора для кожної окремої смуги аналізу звуку враховується при генерації сигналу, який використовується під час управління підсилювачами для кожної з відповідних смуг основного (головного) сигналу. Результат модуляції полягає в тому, що частотні компоненти сигналу, що є модульованими, відображаються на основний сигнал у вигляді дискретних змін амплітуди в кожній частотній смузі.

Дублювальна смуга часто не використовується, хоча і є підпорядкованою. Такі дублювальні смуги використовуються для звукових частот, які знаходяться поза діапазонами аналізу для типового (найбільш поширеного) мовлення, але все ще важливі в мовленні. Прикладами є слова, які починаються з букв «с», «ф», «ч» або будь-якого іншого так званого «шиплячого» звуку. Вони можуть бути змішані з результатом частоти-носія для підвищення чіткості відображення звуку. Результатом є мовлення, яке може зрозуміти людина, хоч воно і має дещо механічне звучання.

Голосові синтезатори часто містять систему, призначену для генерування не дуже чутних або взагалі нечутних звуків, використовуючи генератор шуму замість основної частоти звукового сигналу.

Синтезатор мовлення, що використовує канальний алгоритм, серед двох компонентів аналітичного сигналу враховує лише амплітудні компоненти та ігнорує схильності фазових компонентів до створення шумових сигналів, що призводить до нечіткого звукового відображення голосу, який буде звучати нечітко, мовби з радіоперешкодами (https://uk.wikipedia.org/wiki/Синтез_мовлення).

Модуляція – це процес зміни властивостей звукової хвилі, так званого сигналу-носія, модульованим сигналом, який зазвичай містить інформацію, що передається. Більшість радіосистем використовують частотну або амплітудну модуляцію для радіомовлення. *Модулятор* виконує модуляцію, а *демодулятор*

виконує демодуляцію, зворотну модуляції (<http://ittrap.com/modemmodulation-and-demodulation/>).

Аналогова модуляція слугує передачі аналогового базового (низькочастотного) сигналу, наприклад аудіо- або телесигналу, через аналоговий смуговий канал з різною частотою, наприклад, через обмежений радіочастотний діапазон або мережевий канал кабельного телебачення.

Цифрова модуляція слугує передачі цифрового бітового потоку через аналоговий канал зв'язку, наприклад, через телефонну мережу комутації загального користування. Аналогова та цифрова модуляція полегшують мультиплексування частотного поділу частот, де кілька сигналів інформації з низькою прохідністю передаються одночасно через один і той же загальний фізичний носій, використовуючи окремі канали пропускання (кілька різних частот-носіїв) (<http://ittrap.com/modemmodulation-and-demodulation/>).

Методи цифрової модуляції базової смуги (наприклад, лінійне кодування) полягають у передаванні цифрового бітового потоку по каналу базової смуги (нефільтрованому мідному проводу, послідовній шині чи провідній локальній мережі). У синтезаторах музики модуляція може застосовуватися для синтезу форм хвиль із широким спектром обертонів. У цифровій модуляції аналоговий сигнал-носіє модулюється дискретним сигналом (https://www.porticus.org/bell/belllabs_kitsss.html; Изотов, 2015). Методи цифрової модуляції передбачають цифро-аналогове перетворення, а відповідна демодуляція – аналого-цифрове перетворення.

Основні методи цифрової модуляції поділяються на (<http://ittrap.com/modemmodulation-and-demodulation/>):

- PSK – фазова модуляція – зі скінченною кількістю фаз.
- FSK – частотна модуляція – зі скінченною кількістю частот.
- ASK – амплітудна модуляція – зі скінченною кількістю амплітуд.
- QAM (квадратурна амплітудна модуляція) – використовує скінченне число принаймні двох фаз і принаймні двох амплітуд.

У QAM синфазний сигнал (приклад – косинусоїда) і квадратурний фазовий сигнал (приклад – синусоїда), амплітуда модулюється з кінцевою кількістю амплітуд і потім підсумовується. Тобто це можна розглядати як двоканальну систему, в якій кожен канал використовує ASK. Отриманий сигнал еквівалентний до комбінації PSK і ASK. У всіх перерахованих вище методах кожній з цих фаз, частот або амплітуд присвоюється унікальна схема двійкових біт.

Якщо алфавіт складається з $M = 2^N$ альтернативних символів, кожен символ є повідомленням, що складається з N біт. Якщо швидкість передачі символів становить f_s символів за секунду, швидкість передачі даних становить $N * f_s$ біт за секунду. Зокрема, для алфавіту, що складається з 16 символів, кожен символ представляється 4 бітами.

У PSK, ASK або QAM, де частота модульованого сигналу-носія є постійною, алфавіт модуляції часто представлений на зірковій діаграмі, що показує амплітуду I-сигналу на осі x та амплітуду сигналу Q на осі y для кожного символу (рис. 1 та рис. 2).

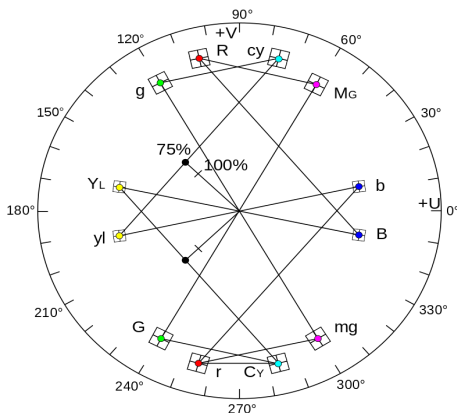


Рис. 1. Приклад зіркової діаграми алфавіту модуляції

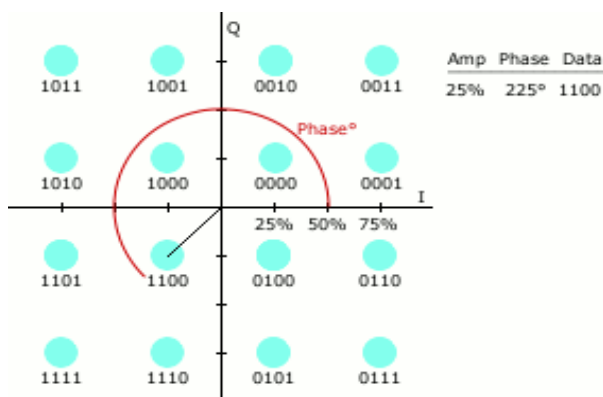


Рис. 2. Приклад діаграми відображення символів алфавіту модуляції

У сигналі QAM один носій відстає від іншого на 90°, а його амплітудну модуляцію заведено називати *фазовим компонентом*, позначеним $I(t)$. Інша модулювальна функція – квадратурна складова, $Q(t)$. Отже, складова форма хвилі математично може бути відображена таким чином (https://www.porticus.org/bell/belllabs_kitsss.html; Изотов, 2015):

$$s_s(t) \triangleq \sin(2\pi f_c t) I(t) + \underbrace{\sin\left(2\pi f_c t + \frac{\pi}{2}\right)}_{\cos(2\pi f_c t + \frac{\pi}{2})} Q(t)$$

або

$$s_s(t) \triangleq \sin(2\pi f_c t) I(t) + \underbrace{\sin\left(2\pi f_c t + \frac{\pi}{2}\right)}_{-\sin(2\pi f_c t + \frac{\pi}{2})} Q(t),$$

де f_c – частота-носії. Когерентний демодулятор множить прийнятий сигнал окремо як косинус, так і синусоїдальний сигнал для отримання одержаних оцінок $I(t)$ і $Q(t)$. Наприклад (https://www.porticus.org/bell/belllabs_kitsss.html; Изотов, 2015):

$$\begin{aligned} r(t) &\triangleq s_c(t) \cos(2\pi f_c t) \\ &= I(t) \cos(2\pi f_c t) \cos(2\pi f_c t) \\ &\quad - Q(t) \sin(2\pi f_c t) \cos(2\pi f_c t) \end{aligned}$$

Використовуючи стандартні тригонометричні тотожності, можна записати це у такому вигляді:

$$\begin{aligned} r(t) &= \frac{1}{2} I(t) [1 + \cos(4\pi f_c t)] - \frac{1}{2} Q(t) \sin(4\pi f_c t) \\ &= \frac{1}{2} I(t) + \frac{1}{2} [1 + \cos(4\pi f_c t)] \\ &\quad - Q(t) \sin(4\pi f_c t) \end{aligned}$$

У фонетиці *формант* – це спектральне формування, яке є результатом акустичного резонансу голосового тракту людини. Однак в акустиці визначення форманта іноді відрізняється, оскільки його можна визначити як пік, або локальний максимум, у спектрі. Для гармонічних звуків формантна частота є частотою гармоніки, що посилюється резонансом. Різниця між цими визначеннями полягає в тому, що «форманти характеризують механізми вироблення звуку» або «форманти характеризують сам вироблений звук».

Форманти – відмінні частотні компоненти акустичного сигналу, що виробляються мовою чи співом. Інформація, необхідна людині для розрізнення мовних звуків, може бути представлена в числовому вигляді.

Форманти поділяються на F1, F2, F3 (йдуть від найнижчої частоти F1). Так звані «проривні» звуки модифікують розміщення формантів у навколишніх голосних. Білабіальні звуки (наприклад, «b» та «p» у «ball» чи «sap») викликають опусканням тону формантів; веларові звуки (англ. «k» та «g») показують F2 і F3; альвеолярні звуки (англ. «t» та «d») викликають менші зміни в сусідніх формантах голосних (залежно від того, який голосний присутній).

Спектрограми можуть бути використані для візуалізації формантів. У спектрограмах важко відрізнити форманти від природних гармонік, коли хтось співає.

Спроби Д. Джонса систематизувати артикуляцію голосних призвели до того, що фактичний форматний простір може бути трикутним чи подібним до трапеції. Показаний сюжет формантів Д. Джонса та Дж. Велса, які вимовляють кардинальні голосні голоси, представлено на рис. 3. Частоти формантів, за їх акустичним визначенням, можна оцінити з частотного спектра звуку, використовуючи спектрограму (представлену на рис. 4) або аналізатор спектру. Для оцінки акустичних резонансів голосового тракту із запису мовлення можна використовувати лінійне прогностичне кодування.

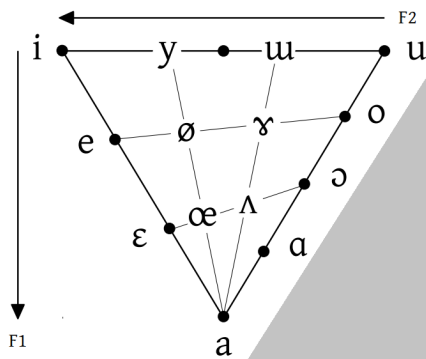


Рис. 3. Сюжет формантів Д. Джонса та Дж. Веллса у вигляді трикутника

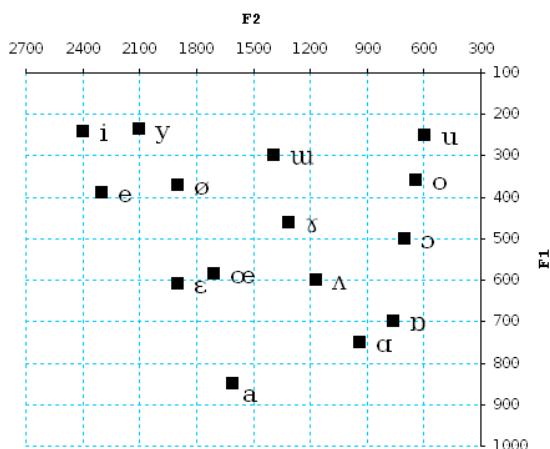


Рис. 4. Спектрограма голосних звуків англійської мови

Розглянута вище математична база синтезу голосу людини знайшла своє практичне застосування. Так, уже в 1982 р. компанія «Mattel Electronix» розробила систему Intellivoice Voice Synthesis, що містила синтезатор мови SP0256 Narrator (http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html; <https://www.porticus.com/en/home/>; <https://en.wikipedia.org/wiki/Vocoder>), який мав 2 КБ пам'яті лише для зберігання бази даних загальних слів, які можна було комбінувати для створення фраз в іграх Intellivision.

Розроблені в 1983 р. компанією «Atari Inc.» персональні комп'ютери 1400XL/1450XL мали систему з інтегрованим в операційну систему синтезатором голосу на базі мікросхеми Votrax SC01. Ці комп'ютери використовували алгоритм Finite State Machine, щоб переводити англійську граматику у звуковий синтез (https://www.porticus.org/bell/belllabs_kitsss.html; https://uk.wikipedia.org/wiki/Синтез_мовлення; Изотов, 2015). У 1984 р. компанія «Apple Computers» випустила програму MacInTalk, засновану на принципі семплінгу звуків.

Сучасні системи під Microsoft Windows використовують компоненти SAPI 4 та SAPI 5 для підтримки розпізнавання і синтезу голосу на основі принципів розпізнавання та відтворення звукового тексту. Серед online-систем з відкритим вихідним кодом можна виділити (http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html; <https://www.porticus.com/en/home/>; <https://en.wikipedia.org/wiki/Vocoder>):

- Festival Speech Synthesi, яка використовує систему дифонного синтезу та деякі більш сучасні техніки;
- eSpeak, що заснована на синтезі формантів;
- gnuspeech, яка використовує артикуляторний синтез.

Кросплатформний програмний продукт, що пропонують, розроблений для відтворення голосу. Він є важливим тим, що за допомогою цього програмного продукту можна оптимізувати розробку програм, якими зможуть користуватися люди, котрі потребують спеціальної підтримки (наприклад, особи з вадами зору).

Невеликі вимоги до ресурсів, операційної системи та максимальна кросплатформність зможуть забезпечити роботу програмного продукту і на простих за конструкцією спеціалізованих пристроях, що спроектовані для людей з особливими потребами.

В основу принципів розробки та функціонування програмного продукту покладено принципи та методи синтезування голосу, що використані для проектування апарату VODER (Voice Operating Demonstrator, демонстратор голосових операцій).

VODER був першою спробою електронного синтезу людської мови. Метод, покладений в його розробку, полягав у тому, що людська мова розбивалася на її найпростіші акустичні компоненти (фонемі).

VODER синтезував людську мову, імітуючи звуки голосового тракту людини. Оператор міг вибрати один з двох основних звуків за допомогою зап'ястної клавіатури. Тон гудіння, що генерується релаксаційним осцилятором, продукував голосні та носові звуки.

На рис. 5 наведено фрагмент програмного коду розробленого програмного продукту.

```
1100 PRINT :PRINT "PITCH = ";:INPUT A
1110 IF A<0 OR A>255 THEN 1100
1120 POKE 39439,A:GOTO 500
1200 PRINT:PRINT "SPEED = ";:INPUT A
1210 IF A<0 OR A>255 THEN 1200
1220 POKE 39438,A:GOTO 500
1300 PRINT :PRINT "THROAT = ";:INPUT A
1310 IF A<0 OR A>255 THEN 1300
1320 POKE 38881,A:SYS 38882:GOTO 500
1400 PRINT :PRINT "MOUTH = ";:INPUT A
1410 IF A<0 OR A>255 THEN 1400
1420 POKE 38880,A:SYS 38882:GOTO 500
1500 PRINT:PRINT " LIGHTS 0:3N= 0:3F=F
n
1510 GET AS:IF AS="" THEN 1510
1520 IF AS="F" OR AS="-" THEN :JLI 0:GOTO
500
1530 IF AS="N" OR AS="/" THEN :JLI 1:GOTO
500
1540 GOTO 1510
1600 JPI 64:JSP 72:JKN 128,128:JLI 0:GOT
0 500
READY.
```

Рис. 5. Фрагмент 1 програмного коду кросплатформного синтезатора голосу

Висновки. Проведене дослідження надає можливість зробити такі висновки.

Аналіз математичного апарату синтезу голосу засвідчив наявні проблеми з відображенням голосу людини та шляхи їх подолання.

Переважна частина програм голосового синтезу має проблеми з точністю відображення голосу людини, розроблена для конкретних типів комп'ютерів чи прив'язана до конкретних операційних систем.

Серед голосових програм не вистачає голосових синтезаторів, що мають таку властивість, як платформна незалежність.

Враховуючи результати проведеного аналізу та необхідність наявності кросплатформних голосових синтезаторів, прийнято рішення щодо розробки голосового синтезатора, визначено його математичне обґрунтування, перспективи кросплатформності.

СПИСОК ПОСИЛАНЬ

Звукова хвиля. [online] Доступно: <https://www.porticus.org/bell/belllabs_kitsss.html> [Дата звернення 29 квітня 2020].

Изотов, М.И., 2015. Механические волны. Звуковые волны. Длина волны. В: *Мне нужно решить задачу.* [online] Доступно: <<http://izotovmi.ru/PRPR/Fizika/Mehanika/zkolm060.htm>> [Дата обращения 26 апреля 2020].

Модуляция и демодуляция. [online] Доступно: <<http://ittrap.com/modemmodulation-and-demodulation/>> [Дата обращения 26 апреля 2020].

Синтез мовлення. [online] Доступно: <https://uk.wikipedia.org/wiki/Синтез_мовлення> [Дата звернення 25 квітня 2020].

History and Development of Speech Synthesis. [online] Available at: <http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html> [Accessed 25 April 2020].

Porticus. [online] Available at: <<https://www.porticus.com/en/home/>> [Accessed 27 April 2020].

Vocoder. [online] Available at: <<https://en.wikipedia.org/wiki/Vocoder>> [Accessed 28 April 2020].

REFERENCES

History and Development of Speech Synthesis. [online] Available at: <http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html> [Accessed 25 April 2020].

Izotov, M.I., 2015. *Mechanicheskie volny. Zvukovye volny. Dlina volny* [Mechanical waves. Sound waves. Wavelength]. In: *Mne nuzhno reshit' zadachu* [I need to solve a problem]. [online] Available at: <<http://izotovmi.ru/PRPR/Fizika/Mehanika/zkolm060.htm>> [Accessed 26 April 2020].

Moduljacija i demoduljacija [Modulation and demodulation]. [online] Available at: <<http://ittrap.com/modemmodulation-and-demodulation/>> [Accessed 26 April 2020].

Porticus. [online] Available at: <<https://www.porticus.com/en/home/>> [Accessed 27 April 2020].

Syntez movlennia [Synthesis of speech]. [online] Available at: <https://uk.wikipedia.org/wiki/Syntez_movlennia> [Accessed 25 April 2020].

Vocoder. [online] Available at: <<https://en.wikipedia.org/wiki/Vocoder>> [Accessed 28 April 2020].

Zvukova khvyliia [Sound wave]. [online] Available at: <https://www.porticus.org/bell/belllabs_kitsss.html> [Accessed 29 April 2020].

UDC 004.451

Tkachenko Olexandr,

*PhD in Physics and Mathematics,
Associate Professor at the department of Software Engineering,
National Aviation University,
Kyiv, Ukraine
aatokg@gmail.com
<https://orcid.org/0000-0001-6911-2770>*

Tkachenko Kostiantyn,

*PhD in Economics,
senior lecturer at the department of Information Technologies,
State University of Infrastructure and Technology,
Kyiv, Ukraine
tkachenko.kostyantyn@gmail.com
<https://orcid.org/0000-0003-0549-3396>*

Shcherbatiuk Vladyslav,

*undergraduate at the department of Information Technologies,
State University of Infrastructure and Technology,
Kyiv, Ukraine
vladscherbatiuk@gmail.com
<https://orcid.org/0000-0002-4592-6400>*

SOME PROBLEMS AND PERSPECTIVES OF CROSSPLATFORM VOICE SYNTHESIZERS DEVELOPMENT

The purpose of the article is to research and consider the general problems and perspectives of voice synthesizer development based on their cross-platform, description and analysis of the development of the author's speech synthesis program, which can support work on the maximum number of systems capable of sound reproduction and independence of operating system.

The research methods are methods of semantic analysis of the basic concepts of the considered subject area (voice synthesizers). The article considers approaches to the development of cross-platform voice synthesizers.

The novelty of the research is the analysis of the mathematical apparatus that can be used for voice synthesis (vowels, different types of consonant sounds) and the description of some aspects of the development of their own author's system that synthesizes the human voice.

Conclusions. The vast majority of voice synthesis programs are designed specifically for certain types of computers or are tied to a specific operating system. The authors proposed a voice synthesizer, determined its mathematical justification and prospects for its cross-platform.

Keywords: voice synthesis; modulator; demodulator; digital modulation; information technology support of voice synthesis; voice synthesizer; cross-platform voice synthesizer.

УДК 004.451

Ткаченко Александр,

*кандидат физико-математических наук,
доцент кафедры инженерии программного обеспечения,
Национальный авиационный университет,
Киев, Украина
aatokg@gmail.com
<https://orcid.org/0000-0001-6911-2770>*

Ткаченко Константин,

*кандидат экономических наук,
старший преподаватель кафедры информационных технологий,
Государственный университет инфраструктуры и технологий,
Киев, Украина
tkachenko.kostyantyn@gmail.com
<https://orcid.org/0000-0003-0549-3396>*

Щербатюк Владислав,

*магистрант кафедры информационных технологий,
Государственный университет инфраструктуры и технологий,
Киев, Украина
vladscherbatiuk@gmail.com
<https://orcid.org/0000-0002-4592-6400>*

НЕКОТОРЫЕ ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ РАЗРАБОТКИ КРОССПЛАТФОРМЕННЫХ ГОЛОСОВЫХ СИНТЕЗАТОРОВ

Целью статьи является исследование и рассмотрение общих проблем и перспектив разработки голосовых синтезаторов на основе их кроссплатформенности, описание и анализ процессов разработки авторской программы синтеза речи, поддерживающей работу на максимальном количестве систем, способных к звуковоспроизведению и независимых от типа операционной системы.

Методами исследования являются методы семантического анализа основных понятий рассматриваемой предметной области (голосовые синтезаторы). В статье исследованы подходы к разработке кроссплатформенных голосовых синтезаторов.

Новизной проведенного исследования является анализ математического аппарата, применяющегося для синтеза голоса (гласных, различных типов согласных звуков), и описание некоторых аспектов разработки собственной авторской системы, синтезирующей голос человека.

Выводы. Большинство программ голосового синтеза разработано специально для определенных типов компьютеров или привязано к конкретной операционной системе. В работе исследован голосовой синтезатор, определено его математическое обоснование и перспективы кроссплатформенности.

Ключевые слова: синтез голоса; модулятор; демодулятор; цифровая модуляция; информационно-технологическая поддержка синтеза голоса; голосовой синтезатор; кроссплатформенный голосовой синтезатор.

10.05.2020