



ISSN 2450-6486

www.ehs-ss.pl

DOI: <https://doi.org/10.38014/ehs-ss.2020.3-2.08>



Наталія ЗАДОРІЖНА

Просодія синтезованого мовлення цифрових асистентів та їх функції у сучасному світі

NATALIA ZADORIZHNA. **Synthetic speech prosody of digital assistants and their functions in the modern world.** *The article is devoted to the speech prosody analysis of digital assistants and their functions in the modern world. Peculiarities of their functioning in different spheres, in particular in educational, are considered and as well as tasks which they perform. Synthetic speech prosody is investigated on the example of voice assistants which speak in a female, male and genderless voice. Characteristics of melody (melodic scales and terminal tones), FF and intensity range, tempo and pause features, voice quality were taken into account. In the course of comparative study some identical and distinctive features of assistants' speech were singled out. The results of the experiment demonstrate variability of prosodic parameters, used by digital assistants. Voice differentiation and different degrees of expressiveness depend on the sphere of application, users' preferences and communication context.*
Keywords: *virtual* voice assistant, functions of assistants, synthetic speech prosody, speech differentiation, variability of prosodic parameters.

Сучасні глобалізаційні процеси здійснюють вагомий вплив на усі сфери суспільства. Події, що відбулися у 2020 році, сколихнули увесь світ. Запровадження дистанційної комунікації через пандемію у сфері бізнесу, освіти та науки призвели до віртуальної революції. В результаті роль інформаційних технологій значно зросла. Один з найбільш революційних цифрових технологій сьогодення – штучний інтелект, широко запроваджується в сучасне життя.

Обробка природньої мови за допомогою цифрових рішень відкриває нові можливості для користувачів з різними запитами. Інтерпретація живого мовлення у вигляді тексту стала можливою завдяки голосовим помічникам, які **набувають усе більшої популярності**. Серед основних можливостей цифрових асистентів, що постійно зростають, можна виділити наступні: пошук рецептів та озвучення покрокової інструкції, озвучення електронних книг, прослуховування музики, замовлення піци та пошук найближчих ресторанів, оплата рахунків, новини, інформація щодо кіноафіш та спортивних подій, зачитування листів, що надходять до персональної електронної пошти, здійснення покупок тощо. Такі гаджети мають режим шепітного мовлення та можуть сприймати шепітне мовлення мовців. Вони також здатні розпізнавати базові позитивні та негативні емоції.

Функції віртуальних асистентів невпинно розширюються, що дозволяє використовувати їх в освітній сфері. Поряд з інтерактивними дошками, цифровими плакатами, електронними тренажерами, електронними підручниками, цифрові помічники як зразки технологій мультимедіа можуть використовуватись в навчальному процесі, зокрема у школах та університетах [8] не менш ефективно. Так цифровий асистент Алекса успішно використовувався під час шкільних занять з навчальною метою учнями різного віку [4]. Такі інтердисциплінарні проекти дозволять зробити інтерактивні заняття більш захопливими та різноманітними. Проте функції такого засобу як робот-асистент в навчальному процесі все ще потребують вивчення та уточнення. Використання подібних засобів викладачами потребує підготовчої роботи й педагогічного моделювання, інтеграції інформаційного, технічного та педагогічного знання [2, с.215-216].

Цифрові асистенти або роботи-вчителі, що діють за встановленою програмою, в навчальному процесі можуть реалізовувати певні функції для зниження навантаження як на викладача, так і на студентів, зокрема інформативну, адаптивну, компенсаторну, управлінську,

інтегративну, інтерактивну, мотиваційну. (енциклопедія) Вони покликані: 1) демонструвати задані презентації та web-ресурси, 2) супроводжувати представлений контент озвученою голосом інформацією і відповідати на запитання, 3) демонструвати медіаконтент, працювати з «хмарними» додатками й сервісами тощо, 4) здійснювати контроль знань [2, с. 219] та перевірку на плагіат, 5) озвучувати аудіокниги 6) виступати в ролі репетиторів з вивчення іноземної мови, [4, с. 18-19, 26]. Використання цифрових асистентів з метою вивчення мови можливе як в межах навчального процесу, так і поза ним. Штучний інтелект також уможливорює більш персоналізоване навчання, відстежуючи прогрес свого власника та адаптуючи матеріали до конкретних потреб користувача [3].

Віртуальних помічників здебільшого описують як таких, що належать до жіночої статі. Це відображено у рекламі, відгуках споживачів, блогах та маркетингових матеріалах. Асистентам присвоюють жіночі імена і встановлюють жіночий голос за замовчуванням. Такі гендерні маркери пов'язані з певними асоціаціями, що виникають стосовно функцій асистента. Їх основним завданням є надання допомоги. І це здебільшого жіноча роль. Превалювання жіночих голосів у цифрових асистентах можна також пояснити більшою емпатичністю та експресивністю жіночих голосів. Зазвичай жінки краще виражають емоції та більш сприйнятливі до почуттів оточуючих, ніж чоловіки.

Смарт-динаміки або розумні колонки відповідають на запити користувачів, допомагають з організаційними питаннями, налаштуванням нагадувань. Цифрові асистенти на основі хмарних сервісів успішно використовуються у системах «Розумний дім». Цікаво, що у віртуальному просторі чат-боти також представлені молодими особами жіночої статі, оскільки робота з клієнтами переважно виконується молодими жінками.

Вибір голосу залежить від сфери використання та не є випадковим. Компанії аналізують уподобання своїх клієнтів перед налаштуванням голосу помічника. Загалом жіночий голос у цифрових асистентах використовується у більш сервісно-орієнтованих сферах (навігація, прогноз погоди, курс валют). Він звучить як більш привітний, розуміючий. Чоловічий голос зустрічається в більш авторитетних ролях, таких як мобільні додатки у банківській та страховій сфері, тобто там, де потрібна порада експерта. Низький чоловічий голос асоціюється зі статусом, авторитетністю й звучить гучніше, чіткіше,

переконаливіше. Також чоловічі голоси частіше використовуються для об'яв, пов'язаних з небезпекою.

Серед найбільш популярних віртуальних голосових асистентів з елементами штучного інтелекту, розроблених відомими компаніями можна виділити такі : Alexa (Amazon), Siri (Apple) Cortana (Microsoft), Google Assistant. На російськомовному просторі Яндекс запропонував голосову помічницю Алісу, тоді як група Mail.ru запропонувала голосового асистента Марусю. Перелічені цифрові помічники говорять жіночим голосом за замовчуванням та виконують схожі функції з більшою чи меншою вправністю. Лише у Великобританії персональний асистент Сірі компанії Apple розмовляє чоловічим голосом на фоні переважної фемінізації даного сектору віртуальних послуг.

Відомо, що перші моделі віртуальних асистентів звучали чоловічими голосами, проте користувачі їх сприймали скоріше як наукових співробітників, бібліотекарів або інформаційних менеджерів, а не особистих секретарів [6]. Команди простіше давати жінкам через їхню традиційно підлеглу роль у суспільстві [7]. Тому сьогодні чоловічий голос є також доступним у багатьох голосових помічниках, проте не за замовчуванням. Користувачі краще сприймають м'який жіночий голос, інтегрований у віртуальних помічників, оскільки він сприяє створенню атмосфери довіри, турботи та комфорту у спілкуванні.

Така голосова диференціація базується на гендерних стереотипах, посилює їх та призводить до дискусій. Саме цей факт посприяв розробці гендерно-нейтрального голосу «Q» для систем штучного інтелекту [9]. Голос був розроблений креативним агентством Virtue Nordic, а також соціологом Джулі Карпентер. Остання зазначила, що соціальна репрезентація чи її відсутність у засобах масової інформації має важливу роль для особистості і неминуче впливає на соціальні цінності. Те, що штучний інтелект може бути чоловіком чи жінкою лише укріплює вже застарілу традицію гендерних норм. Також виникло питання щодо доцільності асоціації штучного інтелекту з гендером.

Розробники проекту розпочали експеримент із запису декількох реальних голосів, які не ідентифікувалися як суто чоловічі чи жіночі [9]. Записи голосів було об'єднано та створено голос, що коливається між 145 Hz і 175 Hz, що є самим оптимальним показником між нормативним голосом чоловіків і жінок. В ході експерименту було

визначено особливості мелодики, стиль мовлення, формантні характеристики. Метою розробників було створення голосу, що буде звучати природно. Для розуміння того, якою є експресивність гендерно-нейтрального голосу необхідно було вивчити та порівняти звукову палітру жіночих та чоловічих голосів. В ході порівняльного аналізу та налаштування мовлення шляхом звукових модуляцій вдалося розробити гендерно-нейтральний голос, максимально наближений до природнього звучання.

Безгендерний цифровий помічник, створений на основі штучного інтелекту, є універсальним рішенням, яке відповідає духу часу та вирішує потреби певної аудиторії. Гендерно-нейтральний голос може використовуватись на платформах різних компаній, а також в транспорті, в іграх, інструкціях, театрах, супермаркетах та інших локаціях. Безгендерна мода та стиль також набирають обертів і скоріш за все стануть більш популярними у майбутньому.

Поруч із бажанням уникнути гендерної упередженості спостерігається й інша тенденція, що акцентує увагу на гендерній варіативності. Проте, незважаючи на те, що всі голоси дуже своєрідні і компанії намагаються запровадити голосові помічники, які відрізнятимуться від інших зразків штучного мовлення, можна виділити певні гендерні особливості комунікативної поведінки голосових помічників. Чоловічому мовленню притаманна асертивність, професіоналізм, тоді як жіноче мовлення сприймається як менш формальне з превалюванням розмовного стилю. Відмінною є й експресивність мовлення різних голосових помічників.

Метою даного доробку є аналіз просодії синтезованого мовлення голосових асистентів. В ході порівняльного дослідження вдалося виокремити спільні та відмінні риси мовлення аналізованих асистентів. До уваги бралися різні просодичні параметри, серед яких чільне місце посідають діапазон ЧОТ та інтенсивності, тембральні характеристики та темп. Всі ці параметри взаємопов'язані, що дозволяє розробникам створювати синтетичні голосові портрети, що відрізняються просодично. Зразки мовлення персональних асистентів, відібрані для аудитивного та акустичного аналізу є відповідями на конкретні запитання/пошукові запити, реакціями на репліки у бесіді або відповіді на накази.

Гендерно-нейтральний голос вирізняється хрипким тембром, що є маркером впевненості, авторитетності, середнім темпом з тенденцією до уповільнення, оскільки фонаційна тривалість

голосних, реалізованих хрипким голосом є довшою у порівнянні з рівним голосом, а також, маловаріативною мелодикою, вузьким діапазоном ЧОТ та інтенсивності. Найчастотнішими в актуалізаціях є поступово низхідні шкали, низькі низхідні або рівні тони. Середні максимальні значення ЧОТ складають 206 Hz, тоді як середні мінімальні тональні значення дорівнюють 116 Hz. Середні максимальні значення інтенсивності складають 45,3 дБ, а середні мінімальні показники – 42,2 дБ.

Мовлення Гугл Асистента в жіночому варіанті характеризується м'яким тембром, що передає доброзичливе ставлення, пришвидшеним темпом, варіативною мелодикою завдяки використанню хвилеподібних шкал у поєднанні з високими низхідними або висхідно-низхідними тонами. Діапазон ЧОТ є (200 Hz) розширеним з середніми максимальними показниками 347 Hz та мінімальними середніми показниками 147 Hz. Діапазон інтенсивності є середнім (15,9 дБ). Середні максимальні показники складають 56,4 дБ, тоді як середні мінімальні показники – 40,5 дБ. Така комбінаторика просодичних параметрів надає мовленню Гугл Асистента значущості і переконливості.

Перцептивно найбільш емоційним сприймається мовлення персонального асистента Сірі. Мовлення Сірі вирізняється пришвидшеним темпом, розширеним діапазоном ЧОТ (199 Hz), при середніх максимальних показниках у 301 Hz та середніх мінімальних показниках у 151 Hz. У мелодиці переважають хвилеподібні шкали та низькі низхідні тони. Останні є маркерами вірогідності, актуальності, впевненості. Проте діапазон інтенсивності є розширеним (18,9 дБ) при середніх максимальних показниках у 61,7 дБ та середніх мінімальних показниках у 42,8 дБ. Мовлення Сірі відрізняється підкресленою паузацією. Поряд із мелодикою, яка виконує функцію актуального членування та допомагає детермінувати статус інформації, логічні паузи розбивають фрази на смислові відрізки та посилюють смислове навантаження ключових слів. У мовленні Сірі зустрічаються заповнені та незаповнені паузи хезитації для більш природнього звучання. Використання пауз хезитації асистентом має місце тоді, коли асистент не може надати відповіді чи потребує уточнень, озвучуючи додаткові запитання.

Алекса демонструє найбільшу варіативність мелодики й послуговується як висхідно-низхідними, так і низькими низхідними тонами. Це залежить від інформації, яку озвучує асистент. Наприклад,

інформація стосовно погоди озвучується більш емоційно, тоді як дані, стосовно вартості акцій чи режиму роботи супермаркету, озвучуються більш нейтрально. Середній тональний діапазон складає 188 Hz, з середніми максимальними показниками ЧОТ 341 Hz та середніми мінімальними показниками ЧОТ 153 Hz. Середні максимальні показники інтенсивності дорівнюють 58,2 дБ, тоді як середні мінімальні показники складають 44,5 дБ. Середній діапазон інтенсивності є звуженим (13,7 дБ). Темп мовлення асистента, як і в більшості асистентів, є пришвидшеним. Він сприймається слухачами як ознака належної компетентності та упевненості в озвучуваній інформації. Тембр Алекси перцептивно сприймається як світлий, доброзичливий. Він передає бадьорість, піднесеність настрою.

В ході дослідження було проаналізовано мовлення асистента Біксбі, що розмовляє чоловічим голосом. Мелодика даного асистента є ідентичною та не залежить від контексту озвучуваної інформації. Домінують у мовленні Біксбі рівні та поступово низхідні мелодичні шкали у поєднанні з низьким низхідним тоном. Діапазон ЧОТ є звуженим (134Hz), де максимальні тональні показники складають 224 Hz, а мінімальні – 90 Hz. Діапазон інтенсивності можна охарактеризувати як середній (16,6 дБ), де середні максимальні показники дорівнюють 66,7 дБ, тоді як середні мінімальні показники складають 50,1 дБ. Мовлення асистента реалізується у середньому регістрі пришвидшеним темпом. Завдяки такій комбінаториці просодичних параметрів актуалізується приємний, привітний, теплий тембр.

Проаналізувавши синтезоване мовлення найпопулярніших голосових асистентів, зокрема Google Assistant, Siri, Alexa, Bixbi та безгендерного голосу "Q", ми дійшли висновку, що кожен з них має свої особливості, які роблять кожного з них унікальним. Якщо голоси перших моделей віртуальних асистентів можна охарактеризувати як пафосні, комп'ютерні, металеві, то сучасні розробки в цій сфері демонструють голоси, максимально наближені до природнього мовлення. Незважаючи на те, що метою розробників асистентів є спрямування на широке коло споживачів, користувачі можуть знайти серед них те поєднання технічних характеристик та звукової палітри, яке підійде саме їм для створення максимального комфорту у спілкуванні з віртуальним асистентом.

Експресивне синтезоване мовлення є бажаною характеристикою взаємодії між людиною та штучним інтелектом у різних сферах

життя. Синтетичний голос має відповідати уподобанням користувача щодо віку, статі, акценту та характеру. Розробка експресивного синтетичного мовлення є вельми актуальною й вимагає дослідження значних масивів мовлення різної приналежності для ідентифікації різних голосових портретів [7, с.3335], та їхнього аналізу для подальшого синтезу мовлення у різних стилях. Варіативність штучного мовлення допоможе настроювати штучні голоси в різних додатках за бажанням та обирати голосовий портрет, який буде відповідати перевагам користувачів та контексту комунікації.

References:

1. Российская педагогическая энциклопедия: в 2 т. / под редакцией В. В. Давыдова. – Москва: Большая Российская энциклопедия, 1993. – Т. 2. – 607 с.
2. Тихонова Л. П. Моделирование функциональных возможностей цифрового ассистента и условий его реализации в вузовском учебном процессе // Вестник Череповецкого государственного университета. – 2019. – № 4 (91). – С. 215–225.
3. Clark, D. How AI will reboot language learning [Electronic source].-2017.- Access mode: <https://thedigitalteacher.com/blog/how-ai-will-reboot-language-learning>
4. Hales P., D., Anderson M., Christianson T., Gaspar A., Meyer B. J., Nelson B., Shilvock K., Steinmetz M., Timmons M., and Weerd M. V. Alexa?: Possibilities of Voice Assistant Technology and Artificial Intelligence in the Classroom. Empowering Research for Educators: Vol. 3: Iss. 1, Article 4. Available at: <https://openprairie.sdstate.edu/ere/vol3/iss1/4>
5. Hern, A. Adios, Alexa: why must our robot assistants be female? [Electronic source].-2019.- Access mode: <https://www.theguardian.com/technology/shortcuts/2019/mar/04/adios-alex-why-must-our-robot-assistants-be-female>.
6. Hester, H. Technically Female: Women, Machines, and Hyperemployment [Electronic source].-2016.- Access mode: <https://salvage.zone/in-print/technically-female-women-machines-and-hyperemployment>.
7. S  ekely E., Cabral J. P., Abou-Zleikha M., Cahill P., and Carson-Berndsen J. Evaluating expressive speech synthesis from audiobook corpora for conversational phrases," in Proc. of LREC, Istanbul, 2012.- P. 3335-3339.
8. Terzopoulos G., Satratzemi M. Voice Assistants and Artificial Intelligence in Education. BCI'19: Proceedings of the 9th Balkan Conference on Informatics September 2019 Article No.: 34 Pages 1–6.
9. The first genderless voice "Q" [Electronic source].- Access mode: <https://www.genderlessvoice.com>.

Translation of References:

1. Russian pedagogical encyclopedia in 2 vol.; ed. by V. V. Davydov. Moscow: Bol'shaia Rossiiskaia entsiklopediia, vol. 2, 1993. 607 p.
2. Tikhonova L. P. Modeling functional capabilities of the digital assistant and the conditions for its implementation in the university educational process. Bulletin of the Cherepovets State University, 2019, no. 4 (91), pp. 215–225.
3. Clark, D. How AI will reboot language learning [Electronic source].-2017.- Access mode: <https://thedigitalteacher.com/blog/how-ai-will-reboot-language-learning>
4. Hales P., D., Anderson M., Christianson T., Gaspar A., Meyer B. J., Nelson B., Shilvock K., Steinmetz M., Timmons M., and Weerd M. V. Alexa?: Possibilities of Voice Assistant Technology and Artificial Intelligence in the Classroom. Empowering Research for Educators: Vol. 3: Iss. 1, Article 4. Available at: <https://openprairie.sdstate.edu/ere/vol3/iss1/4>
5. Hern, A. Adios, Alexa: why must our robot assistants be female? [Electronic source].-2019.- Access mode: <https://www.theguardian.com/technology/shortcuts/2019/mar/04/adios-alexa-why-must-our-robot-assistants-be-female>.
6. Hester, H. Technically Female: Women, Machines, and Hyperemployment [Electronic source].-2016.- Access mode: <https://salvage.zone/in-print/technically-female-women-machines-and-hyperemployment>.
7. S  kely E., Cabral J. P., Abou-Zleikha M., Cahill P., and Carson-Berndsen J. Evaluating expressive speech synthesis from audiobook corpora for conversational phrases," in Proc. of LREC, Istanbul, 2012.- P. 3335-3339.
8. Terzopoulos G., Satratzemi M. Voice Assistants and Artificial Intelligence in Education. BCI'19: Proceedings of the 9th Balkan Conference on Informatics September 2019 Article No.: 34 Pages 1–6.
9. The first genderless voice "Q" [Electronic source].- Access mode: <https://www.genderlessvoice.com>.

The Author

Zadorizhna Natalia
*PhD in Philology,
 Assistant Professor,
 Department of English Philology
 and Intercultural Communication,
 The Institute of Philology,
 Taras Shevchenko National University of Kyiv,
 Kyiv, Ukraine
 E-mail: zadorizhna@gmail.com*

Abstracts

НАТАЛІЯ ЗАДОРІЖНА. Просодія синтезованого мовлення цифрових асистентів та їх функції у сучасному світі. *Статтю присвячено аналізу просодії мовлення цифрових асистентів та їхнім функціям у сучасному світі. Розглянуто особливості їх функціонування у різних сферах, зокрема в освітній, та завдання, які вони вирішують. Просодію синтетичного мовлення досліджено на прикладі голосових асистентів, що розмовляють жіночим, чоловічим та гендерно-нейтральним голосом. До уваги бралися характеристики мелодики (мелодичні шкали та термінальні тони), діапазон ЧОТ та інтенсивності, темпорально-паузальні характеристики та тембр. В ході порівняльного дослідження вдалося виокремити спільні та відмінні риси мовлення асистентів. Результати експерименту свідчать про варіативність просодичних параметрів, використовуваних цифровими асистентами. Голосова диференціація та різний ступінь експресивності мовлення залежать від сфери застосування, переваг користувачів та контексту комунікації.*

Ключові слова: віртуальний голосовий асистент, функції асистентів, просодія синтетичного мовлення, голосова диференціація, варіативність просодичних параметрів.

НАТАЛИЯ ЗАДОРИЖНАЯ. Просодия синтетической речи цифровых ассистентов и их функции в современном мире. *Статья посвящена анализу просодии речи цифровых ассистентов и их функциям в современном мире. Рассмотрены особенности их функционирования в разных сферах, в частности, в образовательной, и задачи, которые они решают. Просодию синтетической речи исследовано на примере голосовых ассистентов, которые разговаривают женским, мужским и гендерно-нейтральным голосом. Во внимание принимались характеристики мелодики (мелодические шкалы и терминальные тоны), диапазон ЧОТ и интенсивности, темпорально-паузальные характеристики и тембр. В процессе сравнительного анализа удалось выделить общие и отличительные особенности речи ассистентов. Результаты эксперимента свидетельствуют о вариативности просодических параметров, используемых цифровыми ассистентами. Голосовая дифференциация и разная степень экспрессивности речи зависят от сферы применения, предпочтений пользователей и контекста коммуникации.*

Ключевые слова: виртуальный голосовой ассистент, функции ассистентов, просодия синтетической речи, голосовая дифференциация, вариативность просодических параметров.

NATALIA ZADORIŻNA. Prozodia mowy syntetycznej asystentów cyfrowych i ich funkcje we współczesnym świecie. Artykuł poświęcony jest analizie prozodii mowy asystentów cyfrowych i ich funkcji we współczesnym świecie. Uwzględniono specyfikę ich funkcjonowania w różnych dziedzinach, w szczególności w dziedzinie edukacyjnej oraz rozwiązywane przez nie zadania. Prozodię mowy syntetycznej badano na przykładzie asystentów głosowych, które mówią głosem żeńskim, męskim i neutralnym płciowo. Uwzględniono charakterystykę melodii (skale melodyczne i tony końcowe), zakres częstotliwości tonu podstawowego i intensywność, charakterystykę pauzy czasowej oraz barwę. W toku badań porównawczych udało się wyodrębnić wspólne i wyróżniające cechy mowy asystentów. Wyniki eksperymentu wskazują na zmienność parametrów prozodycznych używanych przez asystentów cyfrowych. Zróżnicowanie głosu i różne stopnie ekspresji mowy zależą od zakresu, preferencji użytkownika i kontekstu komunikacji.

Słowa kluczowe: wirtualny asystent głosowy, funkcje asystentów, prozodia mowy syntetycznej, różnicowanie głosu, zmienność parametrów prozodycznych.