

АКУСТИКО-ФОНЕТИЧНІ АСПЕКТИ МОДЕЛЮВАННЯ КОМУНІКАЦІЇ ЛЮДИНИ Й КОМП'ЮТЕРА

Проаналізовано основні підходи до моделювання мовленнєвої комунікації людини й комп'ютера шляхом інтегрування технологій синтезу та розпізнавання мовлення в інтерфейси програмного забезпечення. Розглянуто особливості сучасних автоматичних діалогових систем, які поєднують у собі функції синтезу та розпізнавання мовлення, а також містять моделі знань, необхідні для розгортання релевантного діалогу між людиною й комп'ютером. Систематизовано чинники, що спричиняють додаткові складнощі під час автоматичного синтезу та аналізу мовлення. Запропоновано концептуальну схему акустико-фонетичного аналізу зв'язного мовлення, що інтегрується у загальну систему автоматичного розпізнавання. Проаналізовано компоненти типової діалогової системи, що складається з таких частин: фонетичного аналізатора, детектора фонетичних слів, детектора акустичних слів, синтаксичного аналізатора, інтерпретатора аналізу, програми просодичної сегментації, програми фонетико-орфографічної транскрипції, генератора мовлення, діалогового автомату. Окреслено способи підвищення ефективності автоматичного розпізнавання мовлення шляхом використання досягнень сучасної сигнальної фонетики, яка намагається встановити та описати зв'язки між одиницями мовлення й певними класами матеріальних реалізацій у формі сигналів, що можуть бути акустично виміряні.

Ключові слова: мовлення, автоматичний синтез мовлення, розпізнавання мовлення, комунікація, фонетика, діалог.

Постановка наукової проблеми та її значення. Зростаюча роль мовленнєвого спілкування між людиною та комп'ютером спричиняє переосмислення природи комунікативної поведінки загалом і мовленнєвої комунікації зокрема. Очевидно, що використання мовленнєвих технологій уже вийшло за межі сфери інформаційних технологій та стає звичайним атрибутом повсякденного життя [1; 6; 7]. Свідченням цього є інтегрування автоматичних синтезаторів мовлення в середовище текстових редакторів, їхнє широке застосування в мобільній телефонії, автоматичних системах оголошення інформації та ін. [11, 2].

Лише протягом останнього десятиліття спостерігається революційний розвиток систем автоматичного розпізнавання мовлення: від розпізнавання ізольованих слів, вимовлених одним диктором із обмеженим лексиконом (приблизно 50 слів), до розпізнавання потоку зв'язного мовлення, згенерованого будь-яким мовцем із довільним словником [16, 36]. Реальністю сьогодення є ведення телефонного або письмового діалогу з комп'ютером на задану тему, зокрема отримання інформації щодо розкладу руху поїздів, літаків, наявності квитків та ін.. Забезпечення релевантного функціонування систем автоматичного синтезу та розпізнавання мовлення передбачає розв'язання низки проблем, пов'язаних із моделюванням різних типів знань у самих системах, а також моделювання діалогічної комунікації людини й комп'ютера [1; 2]. Під час аналізу діалогу, що відбувається між абонентом і оператором інформаційної системи, стає зрозуміло, що оператор наділений певними фізичними (слуховими, артикуляційними, зоровими) і когнітивними (сприйняття, розмірковування та ін.) властивостями [9]. Отже, він здатен чути, розуміти й контекстуально інтерпретувати висловлювання абонента, а також повинен здійснювати пошук інформації, координувати процес комунікації та надавати релевантну відповідь на запит абонента. В інтерфейсах багатьох комп'ютерних програм втілено таку модель комунікації. Однак результати її використання свідчать, що релевантний діалог можливий лише за умови обмеження комунікації конкретною предметною галуззю (наприклад, прогноз погоди, розклад руху потягів, редагування тексту та ін.) [5].

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.

Будь-яке моделювання комунікації передбачає, передусім, моделювання взаємодії. *Взаємодія (інтерація)* – це взаємний вплив суб’єктів комунікації один на одного [14]. *Розмова* – це особливий вид взаємодії, тобто взаємодія за допомогою голосу. У дослідженні ми визначаємо *діалог* як вербальну інтерацію, що передбачає обмін висловлюваннями, вимовленими двома людьми або людиною та машиною. Усний діалог із комп’ютером – альтернатива письмово-графічній комунікації за допомогою мовних засобів, представлених в інтерфейсах комп’ютерних програм. Однак обидва способи комунікації користувача й комп’ютера, на нашу думку, інтегруються в дискурсі програмного забезпечення. Тому в межах дискурсу програмного забезпечення доречно розрізняти *усну* та *письмово-графічну* мовленнєві комунікації (див. *схему 1*).

Із зазначеної схеми стає зрозуміло, що автоматична обробка усного мовлення представлена двома напрямками: 1) автоматичний синтез мовлення (від друкованого тексту до усного мовлення); 2) автоматичне розпізнавання мовлення (від усного мовлення до друкованого тексту).



Схема 1. Функціонально-структурні аспекти дискурсу програмного забезпечення

Однак жоден із підходів не забезпечує досягнення основної мети комунікації – результативного обміну інформації між її учасниками. Отже, ми висуваємо припущення про те, що дискурсивна комунікація актуалізується винятково в автоматичних *діалогових системах*, які поєднують у собі функції синтезу та розпізнавання мовлення, а також містять моделі знань, необхідні для розгортання релевантного діалогу між людиною й комп’ютером.

У сучасних діалогових системах реалізуються два традиційних типи ведення усного діалогу:

1) надання людиною команд у вигляді звукових повідомлень і послідовне реагування системи на ці команди;

2) взаємний обмін звуковими повідомленнями між людиною та комп’ютером.

Орієнтація на звукове мовлення під час розробки систем, які забезпечують спілкування людини й комп'ютера, спричиняє диференційований підхід до вирішення завдань автоматичного розпізнавання ізольованих команд та розпізнавання зв'язного мовлення. Розв'язання цих завдань ускладнюється через існування низки чинників, як-от:

- 1) варіативність мовленнєвого сигналу, який подається на автоматичний розпізнавач;
- 2) залежність продуктивності системи від індивідуальних властивостей вимови мовця, його психологічного стану, гучності, темпу та ступеня чіткості вимови;
- 3) характеру акустичних шумів;
- 4) об'єм лексичної, синтаксичної й семантичної інформації, що потребує обробки.

Розпізнавання звукового мовлення в багатьох випадках набуває спільних рис із розпізнаванням графічних зображень. Суть процедури зводиться до наступного: якщо можна подати опис об'єкта, то елементи цього опису можуть використовуватись як ознаки цього об'єкта [8, 1235]. При цьому сам опис виступає фундаментом побудови середовища, у якому розв'язуватимуться власне завдання розпізнавання. Аналіз фрагментів об'єкта дає змогу розробити певний словник термінів, за допомогою якого будуватиметься опис. Розпізнавання звукових іміджів полягатиме в створенні «детермінованого або імовірнісного вирішального правила, яке даватиме змогу відносити нові іміджі, що надходять, до того чи іншого класу об'єктів» [3, 16].

Під час сегментації мовлення інформативності набувають також просодичні характеристики. Стратегія автоматичного розпізнавання мовлення з урахуванням просодичних характеристик базується, як правило, на розподілі зв'язного мовлення на фрази та визначення наголосу в цих фразах. Надійна й достовірна інформація отримується шляхом визначення диференціальних ознак наголошених складів. Кількість та розмежування синтаксичних кордонів і наголошених складів використовуються для відбору «можливих семантичних та синтаксичних структур, у межах яких прогнозуватимуться слова, що частково відповідають тим характеристикам, які були отримані під час проведення акустичного аналізу» [12, 1246].

Система акустико-фонетичного аналізу зв'язного мовлення розглядається як частина загальної системи автоматичного розпізнавання. Звукова хвиля подається на вхід до системи, а на виході одержуємо послідовність фонемоподібних одиниць, які позначаються як акустичні фонетичні елементи (наділені конкретними акустичними параметрами) [13]. Акустичний аналіз починається з обробки вхідного сигналу та визначення параметрів. На цьому етапі проводять спектральний аналіз і застосовують лінійне передбачення. Наступний етап зводиться до проведення детальнішої сегментації та класифікації, які необхідні для подальшого лінгвістичного аналізу [13].

Мова як динамічна, чітко структурована система застосовує синтаксичні, семантичні й прагматичні обмеження задля компенсації помилок та усунення двозначностей, які можуть виникати під час акустичної реалізації висловлювань. Цей факт знаходить підтвердження в дослідженні фізіологічних і ментальних властивостей процесу сприйняття мовлення людиною [15].

У багатьох сучасних системах здійснюються спроби моделювання окремих властивостей діяльності людського мозку в процесі сприйняття мовлення [4]. Характерною рисою цих систем є комбіноване використання даних із метою подвійної інтерпретації висловлювання: з позиції його синтаксичної та семантичної припустимості та з позиції його акустично-фонологічного аналізу. Інколи різні типи інформації координуються за допомогою спеціального пристрою – парзера (parser), що здійснює граматичний розбір речень із метою передбачення послідовності слів у висловлюваннях. Не зважаючи на технологічне й статистично-орієнтоване спрямування сучасних систем автоматичного аналізу мовлення, усі вони застосовують елементи лінгвокогнітивного моделювання мовленнєвої діяльності людини.

У сучасних системах розпізнавання зв'язного мовлення використовують також теорію надмірності, яка притаманна природному зв'язному мовленню. Це проявляється в спробах перенесення процесу прийняття рішень на найвищий рівень інформативності [10]. Прийняття будь-якого рішення на акустичному рівні супроводжується втратою певної кількості інформації. Цей процес не базується на лінгвістичній інформації, застосування якої можливе лише на рівні повної інтерпретації повідомлення. Процес розпізнавання слів таким чином із розпізнавання знаків із наступним контекстуальним аналізом зводиться до власне розпізнавання слів як дискретних одиниць мовлення. Лінгвістичний розпізнавач отримує на вході «відсегментовану стрічку векторів, кожен сегмент якої розглядається як вихідний символ, тобто як термінальний знак граматичної системи, яка, у свою чергу, зумовлює кожне конкретне лінгвістичне висловлювання» [3, 19].

Кінцева мета процесу розпізнавання мовлення – ведення діалогу між людиною й комп'ютером із використанням природної мови. Типова діалогова система складається з таких частин:

- 1) фонетичного аналізатора;
- 2) детектора фонетичних слів;
- 3) детектора акустичних слів;
- 4) синтаксичного аналізатора;
- 5) інтерпретатора аналізу;
- 6) програми просодичної сегментації;
- 7) програми фонетико-орфографічної транскрипції;
- 8) генератора мовлення;
- 9) діалогового автомату.

Фонетичний аналізатор дає змогу перейти від реального мовленнєвого сигналу до його фонетичної інтерпретації. Кожному сегменту, який виділений фонетичним аналізатором, відповідає певна множина впорядкованих висловлювань, які розміщені в базі даних згідно зі зменшенням вірогідності їхнього застосування. Ця множина висловлювань має назву «фонетичний спектр фрази» [18, 699].

Фонетичний аналіз проводиться поетапно: на першому етапі відбувається кодування акустичного сигналу в цифровий формат, яке здійснюється за допомогою вокодера; на другому етапі – сегментація звукового сигналу; на третьому етапі – ідентифікація фонем.

Центральною частиною всієї системи є *діалоговий автомат*, який виконує дві функції [14]:

- 1) визначає, які типи фраз мають найбільшу вірогідність бути вимовленими оператором відповідно до контексту діалогу;
- 2) вирішує, яке спрямування буде мати діалог (запит інформації, відповіді на питання та ін.).

Досвід використання діалогових систем підтверджує гіпотезу про те, що спілкування між людиною й машиною допомагає поетапно виправляти помилки в розпізнаванні.

Важлива проблема, яка спричиняє значні труднощі під час розпізнавання мовлення, – проблема *адаптованості систем до конкретних дикторів*. Нормалізація параметрів або ознак для «усунення відмінностей між індивідуальним дикторським мовленням та адаптацією систем до роботи без орієнтації на конкретного диктора відбувається шляхом запровадження універсальної шкали формантних частот» [3, 21]. Ця шкала успішно застосовується для ідентифікації стаціонарних голосних звуків. Однак її ефективність значно зменшується при різноманітних ефектах коартикуляції. Можливі шляхи розв'язання цієї проблеми передбачають вивчення формантних дистрибуцій для ідентифікації голосних звуків із урахуванням довготи голосного, контекстно-зумовлених часових характеристик і спектральних характеристик консонантного оточення.

У процесі створення систем автоматичного розпізнавання звукового мовлення особливого значення набувають експерименти в галузі сприйняття мовлення [17]. ЕОМ, що розпізнає мовлення, часто копіює не тільки функції людського вуха, що покликані аналізувати звуки, а й окремі функції людського мозку, які відповідають за запам'ятовування та генерування логічних зв'язків між елементами мови. Однак, як відзначають дослідники [6; 7], відсоток помилок під час голосового введення інформації все ще складає 10 %, у той час, як при введенні тексту він у 10 разів менший та складає 0,1 %.

Постає питання про оптимізацію систем розпізнавання. Сучасні їх зразки розглядають мовлення як знакову систему, а комп'ютер, що розпізнає мовлення, виконує функції перетворювача пропускну здатності [8]. Оптимізація роботи системи базуватиметься на аналізі надлишковості мовленнєвих знаків. Надлишковість у цьому контексті розуміють як образи відповідних джерел інформації, які можна виміряти й описати за допомогою акустичних характеристик. Аналіз згаданих образів доводить, що обидва компоненти мовлення – фонетика та просодика – однаково важливі для розпізнавання фонетичних знаків. Збільшення словника призводить до збільшення надмірності в природному мовленні й до зменшення надмірності в закодованому мовленні. Крім того, збільшення словника спричиняє експонентне збільшення помилок у природному мовленні та зменшення помилок у закодованому мовленні. Таким чином, аналіз надмірності знакових систем допомагає оптимізувати процес розпізнавання мовлення, установлює кращий розподіл у процесі кодування адитивних і мисленнєвих функцій.

Висновки та перспективи подальших досліджень. Як уже зазначалось, акустичний аналіз просодичних явищ, навіть виконаний на найвищому рівні, не дає змоги робити висновки про релевантність фонетичних явищ. Виникає необхідність звернення до вивчення перцепції. Дослідники вказують на те, що, зважаючи на той факт, що потенційним розпізнавачем мовлення виступає комп'ютер, «людина, яка генерує мовлення, може запроваджувати механізми внутрішнього контролю та внутрішнього виробництва спеціального мовлення» [8, 1235]. Це спеціальне мовлення вимовлятиметься чіткіше, ніж звичайно, його синтаксичні й прагматичні характеристики відповідатимуть вимогам потенційного перцептора – комп'ютера.

Для підвищення ефективності автоматичного розпізнавання мовлення необхідно також вивчити питання про матеріальні кореляти звуків мовлення. Виявлення та експліцитний опис матеріальних реалізацій звуків мовлення – основні завдання *експериментальної фонетики*. Сферою її дослідження є фізіологічні й акустичні кореляти звуків мовлення. Розвиток теорії сприйняття мовлення привів до визначення певних типів сигналів для реалізації матеріальних аудитивних процесів. Вивченням цих сигналів займається *сигнальна фонетика* [3, 21]. Відомо, що всі матеріальні процеси, які супроводжують конкретний акт комунікації та напряду від головного мозку мовця до головного мозку слухача, можуть бути представлені й описані за допомогою різних типів сигналів.

Сигнальна фонетика намагається встановити та описати зв'язки між одиницями мовлення й певними класами реалізацій у формі сигналів, що можуть бути акустично виміряні. Як відомо, лінгвістичні одиниці не описуються без урахування внутрішньої компетенції комунікантів як соціальних психо-фізичних систем. Знаємо також, що акустичний мовленнєвий сигнал – це, насамперед, власне фізичний феномен.

Дослідженням акустичних мовленнєвих сигналів займається лінгвістична акустика, яка також вивчає можливості розпізнавання й опису тексту, що представлений у звуковому форматі. Перспективними у цьому сенсі вважаємо дослідження, що базуються на твердженні про те, що мовленнєвий сигнал містить достатньо надмірної інформації, яка робить можливим процес звукової мовленнєвої комунікації. Це дасть змогу розширити варіативний потенціал система автоматичного синтезу та розпізнавання мовлення.

Джерела та література

1. Біскуб І. П. Англомовний дискурс програмного забезпечення як модель мовленнєвої взаємодії людини й комп'ютера : монографія / І. П. Біскуб. – Луцьк : Волин. нац. ун-т ім. Лесі Українки, 2009. – 388 с.
2. Вейценбаум Дж. Возможности вычислительных машин и человеческий разум: от суждений к вычислениям / Дж. Вейценбаум. – М. : Радио и связь, 1982. – 368 с.
3. Потапова Р. К. Введение в лингвокибернетику / Р. К. Потапова. – М. : Изд-во Моск. гос. лингв. ун-та, 1990. – 140 с.
4. Beale R. Neural Networks and Pattern Recognition in Human-computer Interaction / R. Beale, J. Finlay. – N. Y. : Ellis Horwood, 1992. – 386 p.
5. Bennacef S. An Oral Dialogue Model Based on Speech Acts Categorization, Workshop on Spoken Dialogue Systems / S. Bennacef, F. Нієл, H. Bonneau-Maynard // ESCA Workshop on Spoken Dialogue Systems. – 1995. – P. 237–240.
6. Carpenter B. Human versus machine: psycholinguistics meets ASR / B. Carpenter // Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding. – Keystone, CO., 1999. – P. 225–228.
7. Cutler A. Response time as a metric for comparison of speech recognition by humans and machines / A. Cutler, T. Robinson // Proceedings of ICSLP. – Banff, Canada, 1992. – P. 189–192.
8. Dusan S. On integrating insights from human speech recognition into automatic speech recognition / S. Dusan, L. R. Rabiner // Proceedings of Interspeech. – Lisbon, Portugal, 2005. – P. 1233–1236.
9. Fink J. Personalised hypermedia information through adaptive and adaptable system features: User modeling, privacy and security issues / J. Fink, A. Kobsa, J. Schreck // Intelligence in Services and Networks: Technology for Cooperative Competition / A. Mullery, M. Besson R. Campolargo, R. Reed (Eds.). – Berlin ; Heidelberg : Springer, 1997. – P. 459–467.
10. Holmes J. Speech synthesis and recognition / J. Holmes, W. Holmes. – London ; N. Y. : Taylor and Francis, 2002. – 298 p.
11. Lippmann R. Speech recognition by machines and humans / R. Lippmann // Speech Communication. – 1997. – No 22 (1). – P. 1–15.
12. Maier V. An investigation into a simulation of episodic memory for automatic speech recognition / V. Maier, R. K. Moore // Proceedings of Interspeech. – Lisbon, Portugal, 2005. – P. 1245–1248.
13. McGuire T. W. Group and computer-mediated discussion effects in risk decision making / T. W. McGuire, S. Kiesler, J. Siegel // Journal of Personality and Social Psychology. – 1987. – No 52. – P. 917–930.
14. Minker W. Speech and Human-Machine Dialog / W. Minker, S. Bennacef. – N. Y. ; Boston ; Dordrecht ; London ; Moscow : Kluwer Academic Publishers, 2004. – 89 p.
15. Moore R. K. Constraints on theories of human vs. machine recognition of speech / R. K. Moore, A. Cutler // Proceedings of the Workshop on Speech Recognition as Pattern Classification / eds. by R. Smits, J. Kingston, T. M. Nearey, R. Zondervan. – Nijmegen : MPI for Psycholinguistics, 2001. – P. 145–150.
16. Nass C. Speech interfaces from an evolutionary perspective / C. Nass, L. Gong // Communications of the ACM. – 2000. – No 43 (9). – P. 36–43.
17. Shi R. Function words in early speech perception / R. Shi, J. Werker, A. Cutler // The proceedings of the 15th International Congress of Phonetic Sciences Casual Products. – Adelaide. – 2003. – [CD-ROM. – 3009–3012].
18. Weintraub M. Linguistic constraints in hidden Markov model based speech recognition / [M. Weintraub, H. Murveit, M. Cohen, P. Price, J. Bernstein, G. Baldwin, D. Bell] // Proc. ICASSP_89. – Glasgow, Scotland, May 1989. – P. 699–702.

References

1. Biskub, Iryna. 2009. *Anhlovovnyi Dyskurs Prohramnoho Zabezpechennia yak Model Movlennievoi Vzaiemodii Liudyny i Kompiutora*. Lutsk: Volynskiy Natsionalnyi Universytet Imeni Lesi Ukrainky.
2. Veitsenbaum, Dzhozef. 1982. *Vozmozhnosti Vychislitelnykh Mashyn i Chelovecheskii Razum: ot Suzhdenii k Vychisleniiam*. Moskva: Radio i Sviaz.
3. Potapova, Rodmonga. 1990. *Vvedeniie v Lingvokibernetiku*. Moskva: Izdatelstvo Moskovskogo Gosudarstvennogo Lingvisticheskogo Universiteta.
4. Beale, R., and Finlay, J. 1992. *Neural Networks and Pattern Recognition in Human-computer Interaction*. New York: Ellis Horwood.
5. Bennacef, S., and Нієл, F., and Bonneau-Maynard, H. 1995. “An Oral Dialogue Model Based on Speech Acts Categorization, Workshop on Spoken Dialogue Systems”. *ESCA Workshop on Spoken Dialogue Systems*, 237–240.
6. Carpenter, B. 1999. “Human versus machine: psycholinguistics meets ASR / B. Carpenter”. *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, 225–228. Keystone, CO.
7. Cutler, A., and Robinson, T. 1992. “Response time as a metric for comparison of speech recognition by humans and machines”. *Proceedings of ICSLP*, 189–192. Banff, Canada.
8. Dusan, S., and Rabiner, L. R. 2005. “On integrating insights from human speech recognition into automatic speech recognition”. *Proceedings of Interspeech*, 1233–1236. Lisbon, Portugal.

9. Fink, J., and , Kobsa, A., and Schreck, J. 1997. "Personalised hypermedia information through adaptive and adaptable system features: User modeling, privacy and security issues". In *Intelligence in Services and Networks: Technology for Cooperative Competition*, edited by A. Mullery, M. Besson R. Campolargo, R. Reed, 459–467. Berlin; Heidelberg: Springer.
10. Holmes, J., and Holmes, W. 2002. *Speech synthesis and recognition*. London ; New York: Taylor and Fransis.
11. Lippmann, R. 1997. "Speech recognition by machines and humans". *Speech Communication*, 22 (1): 1–15.
12. Maier, V., and Moore, R. K. 2005. "Aninvestigation into a simulation of episodic memory for automatic speechrecognition". *Proceedings ofInterspeech*, 1245–1248. Lisbon, Portugal.
13. McGuire, T. W., and Kiesler, S., and Siegel, J. 1987. "Groupand computer-mediated discussion effects in risk decision making". *Journal of Personality and Social Psychology*, 52: 917–930.
14. Minker, W., and Bennasef, S. 2004. *Speech and Human-Machine Dialog*. New York; Boston; Dordrecht; London; Moscow: Kluwer Academic Publishers.
15. Moore, R. K., and Cutler, A. 2001. "Constraints on theories of human vs. machine recognition of speech". In *Proceedings of the Workshop on Speech Recognition as Pattern Classification*, edited by R. Smits, J. Kingston, T. M. Nearey, R. Zondervan, 145–150. Nijmegen: MPI for Psycholinguistics.
16. Nass, C., Gong, L. 2000. "Speech interfaces from an evolutionary perspective". *Communications of the ACM*, 43 (9): 36–43.
17. Shi, R., and Werker, J., and Cutler, A. 2003. "Function words in early speech perception". *The Proceedings of the 15th International Congress of Phonetic Sciences Casual Products*. Adelaide. [CD-ROM. – 3009–3012].
18. Weintraub M., and Murveit, H., and Cohen, M., et al. 1989. "Linguistic constraints in hidden Markov model based speech recognition". *Proc. ICASSP_89*. Glasgow, Scotland, May, 699–702.

Бискуб Ирина. Акустико-фонетические аспекты моделирования коммуникации человека и компьютера. Проанализированы основные приципы моделирования коммуникации человека и компьютера путем интеграции технологий синтеза и распознавания речи в интерфейсы прикладных программ. Рассмотрены особенности современных диалогических систем, сочетающих функции автоматического синтеза и распознавания речи, а также содержат модели знаний, необходимых для развертывания релевантного диалога между человеком и компьютером. Систематизировано факторы, провоцирующие дополнительные трудности во время автоматического синтеза и анали за речи. Предложена концептуальная схема фонетико-акустического анализа связной речи. Проанализированы компоненты типической диалогической системы. Очерчены способы повышения эффективности автоматического распознавания речи, предусматривающие использование достижений современной сигнальной фонетики, устанавливающей связь между речевыми единицами и класами материальных реализаций, имеющих форму сигналов.

Ключевые слова: речь, автоматический синтез речи, распознавание речи, коммуникация, фонетика, диалог.

Biskub Iryna. Acoustic and Phonetic Aspects of Modeling Human-Computer Communication. The article summarizes the main approaches to modeling human-computer speech communication by integrating speech synthesis and recognition technologies into the graphical user interfaces. Modern automatic dialogue systems are in the focus of attention. They combine automatic speech synthesis and recognition with conceptual models of knowledge required for modeling relevant human-computer dialogue. The reasons causing additional complications in speech synthesis and recognition are carefully analyzed. The article offers a principal scheme for the acoustic-phonetic analysis of human speech. The components of a typical automatic dialogue system are singled out and examined using the instruments of modern experimental phonetics. The paper suggests a guide of how to implement the mechanisms of modern signal phonetics in order to improve the performance of automatic dialogue systems.

Key words: speech, automatic speech synthesis, speech recognition, communication, phonetics, dialogue.