УДК 004.934.5

М.М. Шатковський
Національний авіаційний університет

Моделі комп'ютерного подання українського мовлення для проблеми конкатенативного сегментивного синтезу

.

В статті наведено моделі комп'ютерного подання мовленнєвої інформації для створення засобів озвучення текстової інформації на основі конкатенативного сегментивного синтезу слів українського мовлення.

В статье приведены модели компьютерного представления речевой информации для создания средств озвучивания текстовой информации на основе конкатенативного сегментивного синтеза слов украинской речи.

The paper presents computer representation models of speech information with a view to create resources of textual information insonification on the basis of concatenative segmentative synthesis of Ukrainian speech words.

Ключові слова: українського мовлення, природність звучання, ознаки природності звучання мовлення, модель подання текстової інформації

Вступ

Розроблення методів, моделей, алгоритмів та програмно-інженерних засобів для систем комп'ютерного розпізнавання та відтворення мовних образів є одним з основних завдань систем штучного інтелекту. Вирішення цих завдань є важливим і актуальним напрямом досліджень з розроблення мультимедійних технологій для систем штучного інтелекту, навчальних та віртуальних середовищ з елементами штучного інтелекту, засобів та систем інтелектуалізації комп'ютерних інтерфейсів.

Проблеми створення таких засобів досліджують у багатьох наукових організаціях світу, до яких належать, зокрема, Лабораторія комп'ютерних наук та штучного інтелекту Масачусетського технологічного інституту, Лабораторія Белла компанії AT&T, Лабораторія систем та оброблення сигналів Політехнічного факультету університету Монса, Відділення теоретичної та прикладної лінгвістики Московського державного університету ім. Ломоносова, Лабораторія розпізнавання та синтезу мовлення Об'єднаного інституту проблем інформатики Національної академії наук Білорусі [1], Державний університет інформатики і штучного інтелекту [2], Міжнародний науково-навчальний центр інформаційних технологій і систем [3], Інститут кібернетики ім.В.М. Глушкова НАН України [4,5] та ін.

Сучасні системи комп'ютерного синтезу мовлення є невід'ємною складовою засобів людино-комп'ютерного інтерфейсу інформацій-них технологій, високоінтелектуальних гіпермедійних технологій, навчальних програм і віртуальних середовищ з елементами штучного інтелекту та мають практичні застосування, зокрема:

– системи мовленнєвого діалогу в апаратних та програмних комплексах (компанії Apple [6], IBM [7], Microsoft [8], Google [9], Ford [10], [11] та ін.);

– бібліотечні, довідникові, енциклопедичні комп'ютерні системи та web-системи (компанії Texthelp Systems [12], VoiceCorp [13] та ін.);

– системи голосового виведення інформації для людей зі спеціальними потребами (компанії Synapse [14], Freedom Scientific [15] та ін.);

– підсистеми озвучення текстової інформації web-джерел і телекомунікаційні системи IP-телефонії (компанії Polycom [16], Addpac [17], Avaya [18], Cisco [19], Grandstream Networks та ін.).

Описані засоби забезпечують зручність (а деколи і єдину можливість) використання програмної складової сучасної комп'ютерної техніки саме завдяки мовленнєвому озвученню інформації. Потреба в них виникає з економічних та соціальних причин, що й зумовлює актуальність створення сучасних засобів комп'ю-терного синтезу мовлення природної якості.

Підходи до комп'ютерного синтезу

мовлення

Системи синтезу мовлення можна класифікувати за способами створення мовного сигналу; вирізняють три основні напрями:

– артикуляторний синтез – створення штучних мовленнєвих сигналів на основі моделювання мовного апарату людини;

– формантний синтез – створення штучних мовленнєвих сигналів на основі акустичних моделей і динамічної зміни таких параметрів, як значення частот основного тону формант та зашумленості;

– конкатенативний синтез – створення вихідного акустичного сигналу на основі конкатенації (послідовного додавання) необхідних елементів синтезу.

Процес конкатенації визначається структурою бази даних елементів синтезу, оскільки безпосередньо залежить від природних даних та забезпечує високу природність звучання синтезованої мови. Тому з підвищенням природності звучання синтезованої мови зростатиме і розмірність елементної бази синтезу.

Є кілька стандартних підходів до вибору концепції формування мінімальних елементів синтезу – фонів, алофонів, дифонів, складів, фонем-трифонів тощо. Природність та якість звучання синтезованих мовленнєвих сигналів пояснюється тим, що в системах конкатенативного синтезу мовлення елементами синтезу є реальні природні мовленнєві сигнали, зазвичай вимовлювані професійними дикторами, корінними носіями мови.

Специфіка української мови, на відміну від багатьох інших мов, полягає у тому, що орфографічне подання та фонетичне (звукове) відтворення мови є досить наближеними. З іншого боку, фонеми в українському мовленні не є ізольованими і внаслідок взаємодії між фонемами модифікуються різні артикуляційні ознаки одних і тих самих звуків з кожним промовлянням. До того ж в українському мовленні наближення частки голосних звуків у мовному потоці до 50%, за умови їх рідкісного збігу, можна трактувати як перевагу відкритих складів «приголосний-голосний» з відповідними міжфонемними коартикуля-ційними переходами.

Основною сучасною проблемою створення методів комп'ютерного синтезу мовлення є підвищення природності звучання синтезованих сигналів, на основі аналізу мовленнєвих даних та врахуванні просодичних властивостей природного мовлення.

Дану статтю присвячено опису створених нових засобів озвучення текстової інформації та моделей конкатенативного синтезу української мови, спрямованих на підвищення розбірливості та натуральності комп'ютерно-синтезованих слів з урахуванням ознак природності звучання українського мовлення.

Конкатенативний підхід до комп'ютерного синтезу мовлення

Технологія конкатенативного синтезу дає змогу комп'ютерам перетворювати довільний текст інформацію в мовлення для надання текстової інформації людям за допомогою голосових повідомлень. Ключова мета text-to-speech додатків в системах зв'язку полягає в поданні голосом текстових повідомлень [20]. Останнім часом робиться дедалі більше спроб її вирішення, більшість досягнутих результатів пов'язані з концепцією конкатенативного синтезу.

Системи конкатенативного синтезу оперують мінімальними усномовними образами – елементами синтезу, які конкатенуються. Принциповим є вибір елементів синтезу, від яких залежатиме природність звучання, розривність та розбірливість синтезованої мови [21]. Популярність цієї концепції полягає в тому, що в основу такого синтезу покладено природні, вимовлені професійним диктором, корінним носієм мови, звукові елементи мовлення. Це є однією з необхідних умов, які допомагають досягати високого рівня природності звучання синтезованої мови.

У системах конкатенативного синтезу (відомий і термін «компілятивний») синтез здійснюється шляхом послідовного додавання (конкатенації) потрібних одиниць з наявної мовленнєвої бази даних. На цьому принципі побудовано велику кількість систем, у яких використовуються різні типи одиниць і різні методи конкатенації. У таких системах потрібно застосовувати оброблення сигналу для зведення частоти основного тону, енергії і тривалості одиниць до тих, якими має характеризуватися синтезована мова. Структура модуля конкатенації визначається обсягом бази даних природних звукових мовленнєвих елементів, оскільки вона безпосередньо залежить від природних даних, тим самим забезпечується висока природність звучання синтезованої мови. Тому з підвищенням природності звучання синтезованої мови буде збільшуватися і елементна база синтезу.

Акустичний процесор генерує та озвучує акустичні дані (звуковий файл), що відповідають вхідному тексту. Генерація звукового файлу з мовленнєвим сигналом полягає у конкатенації вибраних елементів мовленнєвої бази даних.Системи конкатенативного синтезу

маніпулюють мовними сигналами як сукупностями елементів мовлення. Генерація просодичних особливостей та їх варіативність у синтезованому мовленні є дуже складною проблемою [22]. Тому для підвищення рівня природності звучання згенерованих сигналів необхідно розробити таку структуру сегментації/конкатенації природних елементів мовлення, яка б враховувала та використовувала ознаки природності звучання мовних сигналів. Операції, що містяться в модулі цифрового оброблення сигналів, є комп'ютерним аналогом динамічного контролю артикуляторних м'язів та вібруючої частоти голосових зв'язок таким, що вихідний сигнал підбирає вхідні умови. Для цього модуль оброблення цифрового сигналу повинен певним чином враховувати обмеження, оскільки для розуміння фонетичні переходи важливіші від сталих станів [23].

Відповідно до виконаного аналізу, необхідні звукові елементи синтезу (еквівалентні текстовим) зазнають, в разі виникнення потреби, модифікації методами оброблення звукових елементів. Згадані звукові елементи синтезу мовлення мають бути завчасно, згідно з відповідною концепцією синтезу, звукозаписані та виокремлені з загального записаного голосового потоку відповідним маркуванням загального звукового файлу чи в окремі звукові файли.Оброблені сигнали надходять на блок озвучення, де й відбувається створення вихідного звукового сигналу, з наступними діями з його озвучення та/чи запису у вихідні звукові файли

Стислий огляд систем комп'ютерного синтезу мовлення

Система синтезу мовлення Оратор – синтезатор російської мови відповідно тексту

Являє собою набір інструментальних засобів, призначений для впровадження системи синтезу мови в інші програмні продукти. Складається з ряду бібліотек, заголовних файлів, прикладів підключення й документації. Додатково з бібліотекою поставляється один синтезований чоловічий голос. Ядро системи синтезу мови «ЦРТ», із STC TTS Engine 1.5 розроблено і реалізовано відповідно до рекомендації Microsoft Speech API 5.1 [24] .

Система синтезу мовлення The Festival Speech Synthesis System

Комп'ютерна система синтезу мовлення The Festival Speech Synthesis System – дослідницька програма, призначена для вивчення процесу комп'ютерного синтезу мовлення. Проект розпочато в 1996 році в Центрі вивчення мовленнєвих технологій університету Единбурга (The Centre for Speech Technology Research University of Edinburgh). Festival пропонує загальну оболонку для створення систем синтезу мови та приклади різних модулів. У цілому він пропонує всі засоби text-to-speech системи у вигляді декількох API. Festival – мультимовна система синтезу, розроблена мовою C++, використовує Единбурзьку бібліотеку мовного інструментарію. Система підтримує читання різних текстів, врахування інтонації, читання слів з нестандартною вимовою. Можливе підключення бази MBROLA [25].

Система синтезу мовлення Microsoft Speech API SDK

Microsoft SAPI SDK – програмний продукт, створений корпорацією Microsoft спеціально для роботи з усною мовою. SAPI (Speech Application Programming Interface) – цей програмний продукт, оформлений у вигляді набору COM інтерфейсів, надає інтерфейсні функції (у цьому випадку для мовного введення/виведення) іншим додаткам.

Microsoft безкоштовно надає програмний продукт Microsoft SAPI™ та SAPI SDK, що являє собою набір засобів, описів і прикладів, потрібних розробникам для застосування мовних технологій у своїх додатках, що працюють у середовищі Microsoft Windows [26].

Система синтезу мовлення «Агафон»

В основу цієї системи синтезу мови покладено ідею сполучення методів конкатенації й синтезу за правилами.Елементи синтезу в більшості випадків відповідають сегментам фонемної розмірності і є тим самим алофонними реалізаціями традиційних фонем. Мікрофрагменти, що відповідають частинам фонемних сегментів, є тільки для вибухових приголосних (типу /п/, /б/, /п'/, /б'/ і т.д.) і вібрантів (/р/ та /р'/). Головна ж відмінність від традиційних фонетичних подань полягає в тому, що для отримання природної мови враховується багато розбіжностей, обумовлених контекстними фонетичними впливами.

Так, уже в першій версії «Агафон» акустичний інвентар включав 688 одиниць: 158 для приголосних і 530 для голосних. У сучасній версії для жіночого голосу в інвентарі міститься 200 приголосних і близько 1100 голосних алофонів [27].

Система синтезу мовлення Sakrament TTS Engine

Sakrament TTS Engine – система синтезу мови, створена силами компанії Sakrament на

основі унікальної акустичної бази з використанням власних алгоритмів оброблення звуку. Sakrament TTS Engine Home Edition може озвучувати текст п'ятьма різними голосами дикторів і має сервіс, що дозволяє розширити словник наголосів. Система синтезу мови правильно відтворює числа, час, дати, цифрові комбінації, URL, e-mail, індекси, адреси, телефони, обробляє скорочення й абревіатури. Наприклад, «к. ф.-м. н.» вимовляється як «кандидат фізико-математичних наук», «і т.п.» – «тощо» [28].

Система синтезу мовлення SVOX TTS technology

Система синтезу SVOX text-to-speech technology мовлення генерується із тексту з використанням двокрокового методу – аналізу тексту та синтезу голосу. На етапі аналізу тексту абревіатури, специфікації дати та часу й інші послідовності спеціальних символів конвертуються в «читабельний» текст. Після цієї нормалізації кожне слово вхідного тексту аналізується та розкладається на менші одиниці. Потім, на основі проведеного аналізу слів, визначається структура кожного речення. Шаблони постановки фразових наголошень витягуються зі структури речення. Фонетична лексика застосовується для аналізу слова, таким чином, цей процес виконує також і фонетичне подання кожного слова.

Розроблено понад 30 голосів, які покривають передусім європейські, американські та азіатські мови [29].

Система синтезу мовлення Nuance Vocalizer

Система синтезу Nuance Vocalizer створює високоякісне синтезоване мовлення на основі безшовного змішування аудіозаписів наповненої мовленнєвої бази даних. Модульна архітектура програмного забезпечення дозволяє швидко інтегрувати та просто оновлювати мовленнєві дані. Синтезоване на основі text-to-speech підходу мовлення, доступне багатьма мовами, підвищує цінність пристроїв, у яких воно застосовується [30].

Система синтезу мовлення IVONA Reader

Система синтезу IVONA Reader конвертує довільну орфографічну текстову інформацію на персональному комп'ютері у вимовлені слова та дозволяє комп'ютеру прочитати їх вголос. Може використовуватись для text-to-speech читання книжок, інтернет–контенту та документів, створення аудіокниг, вивчення іноземних мов тощо [31].

Особливості перетворення літер у фонеми на вимові

Орфографічне написання відображує, зазвичай, вимовну форму мови в період становлення писемності, на відміну від фонетичного запису, який повинен відображати норму вимови на поточний момент часу. Тому, при читанні орфографічного тексту, читач використовує та застосовує знання про цілий ряд виключень та особливостей вимови на додаток до знання загальних правил перетворення «літера-фонема» [32]. Фонетика – область лінгвістики, що вивчає артикуляцію та сприйняття звуків людського мовлення. Фонологія – область лінгвістики, що вивчає шаблони мовленнєвих звуків [33]. Звуки будь-якої мови можуть бути категоризовані на два типи відомі як силабічні та несилабічні звуки. До перших відносяться голосні, плавні та назальні звуки. До других – приголосні та глайди. Всі мови людства можуть бути широко транскрибовані стандартним міжнародним фонетичним алфавітом визначеним Міжнародною фонетичною асоціацією [34].

Висхідним фонетичним аналізом розглядається три одиниці фонологічного подання звуків відомі як – диференційна ознака, фонема та склад [35]. Слово – найменша вільна форма та базовий компонуючий блок мови [Спенсер]. Просодика відіграє важливу роль при сприйнятті мовлення людиною. За інтонацією визначаються комунікативна направленість висловлювання, логічний зміст, виділення головного та загального, здійснюється виокремлення семантично зв'язаних проміжків мовлення та об'єднання мовленнєвих елементів всередині цих проміжків [32].

Аналіз українського мовлення

Одним з основних завдань дослідження української мови для вибору і створення елементів конкатенативного синтезу є аналіз голосового мовлення з метою виділення властивостей природності звучання та властивостей мовлення для комп'ютерного конкатенативного синтезу. У процесі аналізування необхідно дослідити такі властивості природного українського мовлення:

– частоту визначених буквосполучень у базі даних слів української мови;

– частоту складів у базі даних слів української мови;

– кількості слів з апострофами в базі даних слів української мови;

– кількості слів з різними складносинтезованими властивостями.

Створення мовленнєвих баз даних для комп'ютерного конкатенативного синтезу мовлення є значною проблемою, для вирішення якої необхідно розв'язати ряд кібернетичних задач. Якісні властивості звучання вихідного синтезованого сигналу принципово залежать від двох взаємодоповняльних факторів – способу вибору елементів синтезу та розміру мовленнєвої бази даних. На основі виконаного аналізу пропонуються виділити ключові властивості звучання природного українського мовлення, які суттєво впливають на якість синтезованого мовлення. Для цього досліджено фонетичні ознаки українського мовлення з метою виділення характеристик природності звучання, виконано їх аналіз.

Дослідження голосових мовних сигналів показали, що однакові фонеми в різних контекстних, морфологічних та коартикуляційних умовах мають різні фонетичні характеристики – різняться за амплітудою, частотою основного тону, кількістю періодів основного тону, іншими характеристиками. Ознаки природності звучання українського мовлення для конкатенативного синтезу наведено в табл. 1.

Таблиця 1. Ознаки природності звучання українського мовлення

Ознака	Опис ознаки
G_1	Ознака впливу приголосних звуків на голосні.
G_2	Ознака твердості чи м'якості приголосних.
G_3	Ознака роздільності звучання приголосних та голосних звуків.
G_4	Ознака подвоєння приголосних.
G_5	Ознака спрощення приголосних при вимовлянні.
G_6	Ознака відмінностей звучання між наголошеними та ненаголошеними голосними.
G_p	Ознака позиційності сегмента.

Згідно з виконаним аналізом звуки українського мовлення характеризуються високим рівнем взаємовпливів.

Аналіз сприйняття природного українського мовлення

Найменшою смисловою одиницею, але найважливішою, є слово. Так, згідно з працею [36] мінімальна структурно-семантична одиниця мови, яка виражає своїм звуковим складом поняття про предмети, процеси, явища дійсності, їхні ознаки чи відношення між ними, вільно відтворюється в мовленні і служить для побудови висловлювань; згідно з працею [37] слово – граничний складник речення, здатен безпосередньо співвідноситися з предметом думки; згідно з працею [38] слово – найважливіша одиниця мови, яка позначає явища дійсності та психічного життя людини і зазвичай однаково розуміється колективом людей, які розмовляють однією мовою.

Роль слова як центральної одиниці мови зумовлюється залежністю інших одиниць мови (фонем, морфем, словосполучень, речень) від слова та місця в ньому і системними відношеннями зі словом. При цьому слова – це послідовність фонем, об'єднаних за правилами певної мови. Фонеми можуть мати такі ознаки для їх класифікації [39]:

– дзвінкість і глухість;
– твердість і м'якість;
– взривність і фрикативність;
– назальність чи її відсутність;
– передньоязичність чи задньоязичність.

Інтонація мови – варіативність висоти основного тону голосу, самих квазіперіодів частоти основного тону (пітчів) та їх кількості, гучності мови та її темпу.

Словесний наголос – зміна сили, висоти і тривалості одного чи двох складів у багатоскладових словах за допомогою засобів інтонації.Згідно з працею [40] людське вухо сприймає звукові хвилі довжиною від 1,6 до 22 см, що відповідає частотному діапазону 16–22000 Гц. Що стосується людської мови, то її частотний діапазон становить 300–4000 Гц. У процесі вимови активні органи мови (голосові зв'язки, язик, губи, м'яке небо, язичок, задня спинка зева та нижня щелепа) виконують рухи, що називаються артикуляцією, та формують спектр звуків. Артикуляція складається із приступу, витримки та відступу. Пасивні органи мови лише визначають форму внутрішніх порожнин мовних органів, що впливають на резонансні властивості порожнин [40]. Дедалі більше лінгвістів визнають правильним твердження про те, що саме слово є основною одиницею мови [41].

На вимову впливає, наприклад, плавність, яка залежить від глибини вдиху та раціональної витрати повітря. Можливе навіть добирання повітря на обумовлених мовленнєвих відрізках,

які можуть бути змістовно та орфографічно обумовленими. Завдяки такій властивості природного промовляння забезпечуються неозривність та плавність звучання природної вимови на певних ділянках мовлення.

Українська мова наймелодійніша й найголосніша поміж усіма слов'янськими мовами з великими музикальними можливостями. Відомо було досі, що українська мова своєю доброзвучністю займає одне з перших місць (може, третє або четверте) між усіма європейськими мовами. Цю думку дуже легко пояснити фактами фонетики [42].

У словниках і наукових працях милозвучність, що ототожнюється з евфонією, має, по суті, два термінологічні значення.

По-перше, вона визначається як добре, приємне з погляду фонетичних і лексико–стилістичних норм певної мови звучання окремих мовних елементів – звукосполучень, слів і словосполучень [36]. По-друге, милозвучність — це звукові засоби підсилення виразності мови художнього твору внаслідок досягнення гармонійного добору звуків у тексті.

Це, зокрема, такі засоби, як звукові повтори різних видів, алітерації, асонанси, ритмічність мови, а також уникнення важких для вимови чи неприємних для слуху сполучень звуків у фразі [36]. Неважко помітити, що перше тлумачення милозвучності стосується мови в цілому й увага акцентується на тому, як сприймається її звучання: позитивно–негативно, приємно–неприємно і передбачає порівняння спостережуваної мови з іншою (іншими). Друге тлумачення характеризує милозвучність як стилістичний прийом, побудований на використанні звукових засобів у конкретному художньому творі певної мови. Вони прив'язані до його змісту, покликані увиразнити його. Крім того, милозвучність у першому розумінні визначається, або, що правильніше, регулюється, орфоепічними нормами як неодмінною ознакою української літературної норми, у другому розумінні — зумовлюється бажанням і потребою автора увиразнити текст твору [43].

Милозвучність української мови забезпечується насамперед, декількома факторами – виразною повнозвучною вимовою голосних і приголосних у сильних і слабких позиціях, порівняно невеликою кількістю збігів кількох приголосних, плавною акцентно-ритмічною структурою слова, наспівною мелодикою тощо [43].

Аналіз сприйняття синтезованого мовлення

Технологія синтезу мови має давні корені. Однак лише донедавна такі системи синтезу не знаходили широкого застосування. Однією з причин тому – не достатня якість синтетичної мови. Її роботизованість і неприродність робили не придатною для широкого використання систем синтезу [44].

Можливості синтезованої мови залежать від галузі застосування. Можливе озвучення обмеженої кількості голосових фраз чи речень – систем озвучення з обмеженим словником. Проте можливий і повний синтез довільної мовної інформації – створення синтезаторів з необмеженим словником. Для характеристики якості мови зазвичай використовують такі поняття, як «природність звучання», «фонетична розбірливість», «комфортність сприйняття» і «час звикання» [45].

Природність звучання характеризує близькість синтезованого звуку до людської мови. Перші синтезатори вирізнялися металічним призвуком, відсутністю інтонаційного розподілу фрагментів мови, різкістю звучання чи, навпаки, надто затягнутими голосними звуками. Фонетична розбірливість характеризує, наскільки слухачеві легко або важко розібрати фонеми, вимовлені синтезатором та семантично розпізнавати синтезовану мову. Комфортність сприйняття і час звикання до синтезованої мови відображають суб'єктивну оцінку слухачем якості синтезованої мови. Довге прослуховування синтезованої мови не повинно викликати надмірного стомлення, а час звикання має бути обмеженим.

Проблема створення засобів аналізу мовної інформації для синтезу українського мовлення

Ключовим засобом передачі семантичної та просодичної складових мови в конкатенативному синтезі мовлення є елементи синтезу.

Саме на основі конкатенації звукових елементів синтезу і функціонує конкатенативний синтез. Для вибору та створення елементів синтезу необхідно провести ряд досліджень українського мовлення та створити відповідні комп'ютерні засоби аналізу мовленнєвих даних.

Одним з основних завдань є аналіз голосового мовлення з метою виділення властивостей природності звучання та властивостей мовлення для комп'ютерного конкатенативного синтезу. У процесі аналізування необхідно дослідити такі властивості природного українського мовлення:

– частоту різних визначених буквосполучень у базі даних слів української мови;

– частоту складів у базі даних слів української мови;

– кількості слів з апострофами в базі даних слів української мови;

– кількості слів з різними складносинтезованими властивостями.

Створення мовленнєвих баз даних для комп'ютерного конкатенативного синтезу мовлення є значною проблемою, для вирішення якої необхідно розв'язати ряд кібернетичних задач. Необхідно створити базовий простір системи комп'ютерного конкатенативного синтезу мовлення – множину текстових елементів та відповідних їм звукових елементів синтезу. Текстові елементи синтезу використовуються для сегментації вхідних текстових даних, а звукові елементи синтезу – для конкатенації у вихідний звуковий синтезований мовленнєвий сигнал.

Якісні властивості звучання вихідного синтезованого сигналу принципово залежать від двох взаємодоповняльних факторів – способу вибору елементів синтезу та розміру мовленнєвої бази даних. Необхідною умовою створення мовленнєвих баз даних є узгодження акустичних характеристик (знаходження елемента в потрібному контексті з потрібною тривалістю та контуром частоти основного тону тощо) звукових елементів синтезу між собою для зменшення кількості розривностей вихідного синтезованого мовленнєвого сигналу

Модель взаємозв'язків ознак природності звучання українського мовлення – позиційності, наголошеності та мультифонемності

Для врахування зазначених взаємозв'язків ознак природності звучання та покращення натуральності звучання для вибору елементів синтезу розроблено модель взаємозв'язків ознак природності звучання українського мовлення – позиційності, наголошеності та мультифонемності.З огляду на важливу роль ознаки позиційності мовленнєвих сегментів ключовими ознаками є G_1, G_6, G_p. Саме ці ознаки і формують основні властивості природності звучання українського мовлення. Варто зауважити, що ознака G_2 пов'язана з усіма іншими ознаками, оскільки властивість м'якості чи твердості приголосних наявна в усіх відкритих складах української мови.

Проте G_2 враховано в коартикуляційних властивостях міжфонемних переходів. Інакше кажучи, G_2 є важливою, але залежною від ознаки G_1 коартикуляційного впливу приголосних на голосні.

Ознаки $G_1 – G_5$ визначають властивості та взаємозв'язки звучання фонем українського мовлення, тобто ці ознаки формують властивості мультифонемності звучання. Звідси схему ознак природності звучання можна зобразити у вигляді рис.1, де G_s – ознака наголошеності голосного, а G_m – сукупність властивостей ознак $G_1 – G_5$:

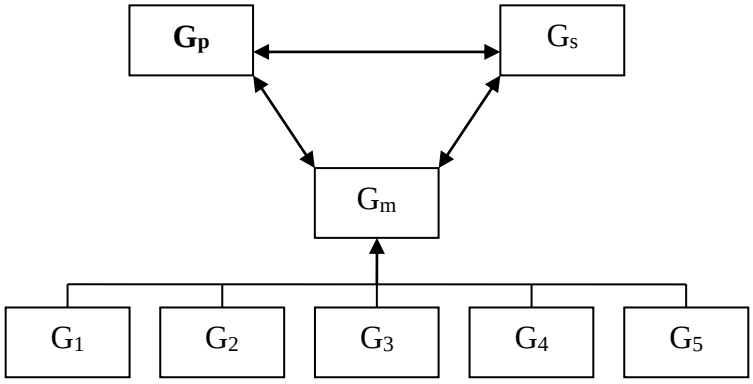

Рис. 1. Модель взаємозв'язків ознак природності звучання українського мовлення – позиційності, наголошеності та мультифонемності

Удосконалена модель подання вхідної текстової інформації для конкатенативного сегментивного синтезу українського мовлення

Для врахування зазначених ознак на основі моделі взаємозв'язків ознак природності звучання українського мовлення запропоновано удосконалену модель подання вхідної текстової інформації для конкатенативного сегментивного синтезу українського мовлення, що дозволило здійснювати сегментацію та конкатенацію мінімальних елементів синтезу (це можуть бути як класичні елементи – фони, дифони, фонеми-трифони тощо, так і запропоновані в моделі сегменти спеціальної структури):

$$w_j = p_k if_m, \quad (1)$$

де w_j – слово із загального набору слів української мови W, $w_j \in W$, $j \in [1,|W|]$, $|W|$ – кількість слів української мови; p_k – початковий (префіксний) сегмент із набору початкових сегментів P, $p_k \in P$, $k \in [1,|P|]$, $|P|$ – кількість початкових сегментів; f_m – кінцевий (суфіксний) сегмент із набору кінцевих сегментів F, $f_m \in F$, $m \in [1,F]$, $|F|$ – кількість кінцевих сегментів. Слово розкладається на три частини відповідно до ознаки позиційності – початкову, внутрішню та кінцеву.

Початкова та кінцева частини є сегментами. Внутрішня частина може бути порожньою, містити один сегмент чи складатися з декількох сегментів. Внутрішня частина i слова w_j може бути відсутня (наприклад в двоскладових словах) – тоді подання слова набуватиме вигляду $w_j = p_k f_m$; внутрішня частина i слова w_j може складатись із одного сегмента $i = i_l$ – тоді подання слова набуватиме вигляду $w_j = p_k i_l f_m$; внутрішня частина слова може складатись з декількох сегментів $i = (i_1,...,i_n)$ – тоді подання слова набуватиме вигляду $w_j = p_k (i_1,...,i_n) f_m$. В усіх сегментах враховується ознака наголошеності, тобто необхідні сегменти є наголошеними. Згідно з ознакою мультифонемності сегменти вибирають якнайбільшими, враховуючи коартикуляційні та просодичні властивості слова – зменшуючи кількість конкатенацій сегментів і підвищуючи природність звучання синтезованої мови. Врахування наголошеності мовного потоку є принциповою складовою розроблюваного підходу. Сегменти, що містять голосні фонеми чи ними є, розглядаються в обох варіантах – наголошеному і ненаголошеному. Довільне слово української мови подається як ключова одиниця мовлення, що спричинено залежністю елементів синтезу від місця розташування в слові, його коартикуляційних властивостей, фізичних та акустичних характеристик, і визначається складністю зв'язків різних елементів слова під час вимовляння. В запропонованому підході завдання синтезу зводиться до озвучення заданих відповідно до вхідного тексту послідовних наборів об'єктів синтезу.

Об'єктно-елементна модель конкатенативного сегментивного синтезу українського мовлення

Щоб отримати можливість підбирати сегменти для озвучення конкретних об'єктів синтезу та щоб зафіксувати й пов'язати набори сегментів і об'єктів синтезу, розроблено об'єктно-елементну модель конкатенативного сегментивного синтезу українського мовлення.

Згідно з цією моделлю об'єктам синтезу ставляться у відповідність послідовні набори елементів синтезу (сегментів) з урахуванням натуральності та розбірливості звучання. Такий підхід дозволяє контролювати якість звучання конкретних синтезованих слів зокрема та підвищити рівень природності звучання синтезованого мовлення взагалі.

Схематично об'єктно-елементну модель конкатенативного сегментивного синтезу українського мовлення зображено на рис. 2.

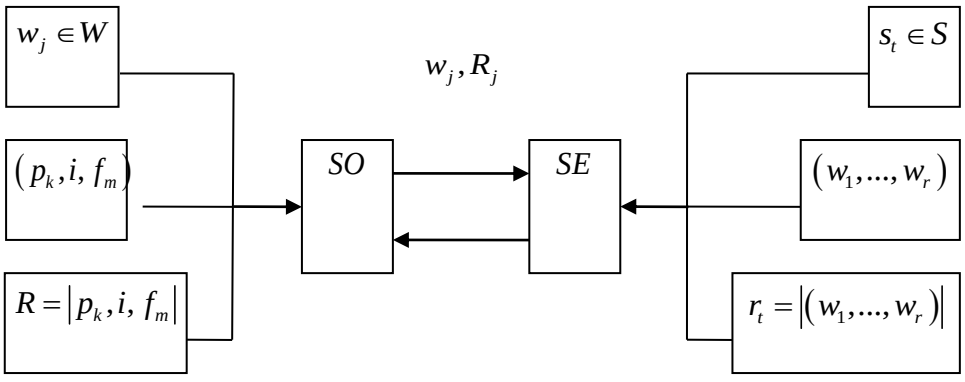

Рис. 2. Схема об'єктно-елементної моделі подання інформації

Крім зазначеного, запропонована модель дозволяє розробити засоби побудови мінімального корпусу слів, на основі чого, після звукозапису, стає можливим вирішувати таке практичне завдання, як створення бази даних звукових сегментів.

Висновки

В статті описано ряд моделей подання мовленнєвої інформації для створення засобів озвучення текстової інформації на основі конкатенативного сегментивного синтезу українського мовлення, а саме:

– запропоновано модель взаємозв'язків ознак природності звучання українського мовлення – позиційності, наголошеності та мультифонемності, використання якої дає можливість врахувати зазначені взаємозв'язків ознак природності звучання та покращити натуральність звучання елементів синтезу;

– наведено удосконалену модель подання текстової інформації для конкатенативного синтезу українського мовлення, що дозволило врахувати ознаки позиційності, наголошеності та мультифонемності звучання мовленнєвих сегментів, наведені в моделі взаємозв'язків ознак природності звучання українського мовлення;

– описано об'єктно-елементну модель конкатенативного сегментивного синтезу, що дозволяє зафіксувати та пов'язати набори об'єктів синтезу та елементів синтезу (сегментів) для конкатенативного сегментивного синтезу, отримати можливість підбирати конкретні сегменти для озвучення конкретних слів української мови та підвищувати рівень природності звучання синтезованого мовлення.

Список посилань

1. Лобанов Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск: Білорус. наука. – 2008. – 316 с.

2. Шелепов В.Ю. Структурная классификация слов русского языка. Новые алгоритмы сегментации речевого сигнала, распознавания фонем и их классов / В.Ю. Шелепов, А.В. Ниценко // Искусственный интеллект. – Донецк: ИПИИ МОН и НАН Украины. – 2006. – № 4. – С. 679–690.

3. Вінцюк Т.К. Автоматичний озвучувач українських текстів на основі фонемно-трифонної моделі з використанням природного мовного сигналу / Т.К. Вінцюк, Т.В. Людовик, М.М. Сажок, Р. Селюх // Пр. 6-ї Всеукр. міжнар. конф. «Оброблення сигналів і зображень та розпіз-навання образів» (УкрОбраз'2002). – К.: УАсОІРО. – 2002. – С. 79 – 84.

4. Кривонос Ю.Г. Структура, свойства, характеристики объектов и элементов синтеза речи / Ю.Г. Кривонос, Ю.В. Крак, Н.Н. Шатковский // Компьютерная математика. – К.: Институт кибернетики им. В.М. Глушкова НАН Украины. – 2006. – №1. – С. 61–69.

5. Шатковський М.М. Об'єктно-елементна модель подання текстової інформації для задачі конкатенативного сегментивного синтезу української мови / М.М. Шатковський // Штучний інтелект. – Донецьк: ІПШІ МОН і НАН України. – 2008. – № 4. – С. 796–802.

6. Веб-сторінка компанії Apple [Електронний ресурс]. – Режим доступу: http://www.apple.com/pro/tips/text_speech.html

7. Веб-сторінка компанії IBM [Електронний ресурс]. – Режим доступу: http://www.research.ibm.com/tts/

8. Веб-сторінка компанії Microsoft [Електронний ресурс]. – Режим доступу: http://www.microsoft.com/speech

9. Веб-сторінка компанії Google[Електронний ресурс]. – Режим доступу: http://www.google.com.

10. Веб-сторінка компанії Ford [Електронний ресурс]. – Режим доступу: http://www.google.com.

11. LaMonica M. Ford brings digital comforts to cars [Електронний ресурс] / M. LaMonica // CNET. – 2010. – Режим доступу до журналу: http://ces.cnet.com/8301-31045_1-10428339-269.html

12. Веб-сторінка компанії Texthelp Systems [Електронний ресурс]. – Режим доступу: http://www.texthelp.com

13. Веб-сторінка компанії VoiceCorp [Електронний ресурс]. – Режим доступу: http://www.voice-corp.com/

14. Веб-сторінка компанії Synapse [Електронний ресурс]. – Режим доступу: http://www.synapseadaptive.com/

15. Веб-сторінка компанії Freedom Scientific [Електронний ресурс]. – Режим доступу: http://www.freedomscientific.com/

16. Веб-сторінка компанії Polycom [Електронний ресурс]. – Режим доступу: http://www.polycom.com/

17. Веб-сторінка компанії Addpac [Електронний ресурс]. – Режим доступу: http://www.addpac.com/

18. Веб-сторінка компанії Avaya [Електронний ресурс]. – Режим доступу: http://www.avaya.com/

19. Веб-сторінка компанії Cisco [Електронний ресурс]. – Режим доступу: http://www.cisco.com/

20. Cox R.V. Speech and language processing for next-millenium communication services / R.V. Cox, C.A. Kamm, L.R. Rabiner, J. Schroeter, J.G. Wilponl // Proc. Of the IEEE. – 2000. – V. 88. – N. 8. – P.1314–1337

21. Сажок М.М. Автоматизовані засоби дослідження синтезу українського мовлення на основі фонемно-трифонної моделі / М.М. Сажок // Автоматизовані системи управління та прогресивні інформаційні технології. – К.: МННЦІТіС. – 2003. – Вип.. 1. – С. 101–113.

22. Strom V. From Text to Prosody without ToBI / V. Strom // Proc. International Conf. on Spoken Language Processing. – Denver, 2002. – P. 2081–2084.

23. Dutoit T. MBR-PSOLA: Text-to-speech synthesis based on an MBE resynthesis of segments database / T. Dutoit, H. Leich // Speech Communication. – 1993. – № 13. – P. 435 – 440.

24. Веб-сторінка опису синтезатора «Оратор» [Електронний ресурс]. Режим доступу: http://www.speechpro.ru/rus/company/tech/tech-orator/

25. Веб-сторінка опису проекту MBROLA [Електронний ресурс]. – Режим доступу: http://tcts.fpms.ac.be/synthesis/

26. Веб-сторінка опису проекту MS Speech SDK [Електронний ресурс]. – Режим доступу: http://msdn.microsoft.com/en-us/library/ms723627(VS.85).aspx

27. Зиновьева Н.В. Программный синтез русской речи (синтезатор «АГАФОН») / Н.В. Зиновьева, О.Ф. Кривнова, Л.М. Захаров // Труды Междунар. семинара по компьютерной лингвистике и ее приложениям. – М: Наука. – 1995. – С. 146–153.

28. Веб-сторінка опису синтезатора компанії «Sakrament» [Електронний ресурс]. – Режим доступу: http://www.sakrament.com/main.php?Lang=ru&TopId=20&Category=1

29. Веб-сторінка опису синтезатора компанії «SVOX» [Електронний ресурс]. – Режим доступу: http://www.svox.com/TTS-Technology.aspx

30. Веб-сторінка опису синтезатора компанії «Nuance» [Електронний ресурс]. – Режим доступу: http://www.nuance.com/for-business/by-solution/contact-center-customer-care/cccc-solutions-services/vocalizer/index.htm

31. Веб-сторінка опису синтезатора компанії «IVONA» [Електронний ресурс]. – Режим доступу: http://www.ivona.com/reader.php

32. Лобанов Б.М. Проблемы и решения компьютерного «клонирования» персонального голоса и речи / Б.М. Лобанов // Проблемы и методы экспериментально-фонетических исследований. – СПб: СПГУ. – 2002. – С. 301 – 308.

33. Yingxu Wang Software engineering foundations: a software science perspective 2008

34. The International Phonetic Alphabet, http://www.arts.gla.ac.uk/IPA/ipa.html

35. Gleason, J.B. The Development of Language, Introduction to Descriptive Linguistics, 4th ed., Allyn and Bacon, Boston, MA.

36. Ганич Д. І. Словник лінгвістичних термінів / Д.І. Ганич, І.С. Олійник. – К.: Вища шк. – 1985. – 127 с.

37. Ахманова О.С. Словарь лингвистических терминов / О.С. Ахманова. — М.: Сов. энцикл. – 1969. – 608 с.

38] Будагов Р.А. Очерки по языкознанию / Р.А. Будагов. – М.: Ин-т языкознания АН СССР. – 1953. – 300 с.

39. Радван Д.В. Психологічний напрямок української фонології: погляд у контексті новітніх досліджень / Д.В. Радван // Актуальні проблеми української лінгвістики: теорія і практика: зб. наук. праць. – К.: Вид. – поліграф. центр «Київський університет». – 2003. – Вип. VII. — С.128 –135.

40. Физический энциклопедический словарь; [гл. ред. А.М. Прохоров; редкол. Д.М. Алексеев, А.М. Бонч-Бруевич, А.С. Боровик-Романов и др.]. – М.: Сов. энцикл. – 1983. – 928 с.

41. Вихованець І. Р. Частини мови в семантико-граматичному аспекті. / І. Р. Вихованець. — К.: Нац. ун-т імені Тараса Шевченка. – 1988. – 256 с.

42. Самійленко В. Твори: у 2 т. / В. Самійленко. – К.: Дніпро. – 1958. – Т. 2. – 365 с.

43. Мосенкіс Ю.Л. Проблема милозвучності української мови: теоретичні й методичні аспекти / Ю.Л. Мосенкіс // Наукові записки НауКМА: Філологічні науки. – К.: Вид-во НауКМА. – 2002. – Т.20. – С. 23–25.

44. Веб-сторінка опису проблеми синтезу мовлення компанії Speechpro [Електронний ресурс]. – Режим доступу: http://www.speechpro.ru/techno/synthesis

45. Веб-сторінка опису проекту JAWS [Електронний ресурс]. – Режим доступу: http://jaws.tiflocomp.ru/synths

Відомості про автора

Шатковський Миколай Миколайович – к.т.н., доцент кафедри інженерії програмного забезпечення факультету комп'ютерних наук Національного авіаційного університету. Коло наукових інтересів – засоби і системи інтелектуалізації комп'ютерних інтерфейсів, методи, моделі, алгоритми та програмні засоби для систем комп'ютерного аналізу, обробки та синтезу мовленнєвих образів, високоінтелектуальні мульти- та гіпермедійні технології і засоби, в тому числі, для систем штучного інтелекту, концептуальні, математичні та програмні засоби для побудови інтелектуальних систем розробки та керування людиноподібними роботами та робототехнічними комплексами.
E-mail: mykolash@gmail.com

Стаття надійшла до редакції 11.03.2011 р.