О. Володкевич*
Київський національний лінгвістичний університет (Київ)
УДК 81'322.6

СТВОРЕННЯ Й ВИКОРИСТАННЯ БАЗИ ДАНИХ СКЛАДІВ АНГЛІЙСЬКОГО МОВЛЕННЯ ДЛЯ БЛОКУ ЛІНГВІСТИЧНОЇ ОБРОБКИ ТЕКСТІВ У СИСТЕМІ КОНКАТЕНАТИВНОГО СИНТЕЗУ

The article focuses on the problem of English speech synthesis in the System of Concatenation Synthesis. The core principles and difficulties in creating the database for the synthesis are under our analysis. The article presents the structure of the database and the corresponding concepts of its usage in the application. The database strategies underline the natural sounding of the synthesized speech and saving the individual features of a Speaker.

Імітація людського мовлення машиною цікавить вчених із давніх давен, але з XVIII ст., епохи активного розвитку механіки, збереглися два унікальні пристрої. Пневмопристрій Х. Кратценштейна [1], завдяки збудженню резонансів, міг утворювати голосні звуки, а „мовляча" машина В.фон Кемпелена [2], яка зовні нагадувала піаніно, могла імітувати звучання простих складів, міняти тривалість та інтенсивність звучання. Рівень сучасних електронно-обчислювальних машин і можливості програмування необхідних умов синтезу дозволили перетворити зусилля окремих дослідників у потужну технологічну галузь [4]. Сьогодні мовленнєві технології після випробування формантних синтезаторів, які працювали за правилами фільтрації спектральної картини з максимально збагаченого акустичного сигналу, перейшли до нового етапу й активно впроваджують конкатенативний синтез [3].

Конкатенативний, або компілятивний синтез мовлення – це синтез, що базується на конкатенації (з'єднанні) акустичних сегментів певної фонетичної розмірності, дібраних і проанотованих на матеріалі записів індивідуального мовлення диктора. Вважається, що саме цей вид синтезу генерує найбільш природне мовлення. Одиницями такого синтезу можуть бути алофони, дифони, трифони, склади та навіть цілі слова. Щодо просодії (інтонації), то вона залишається найскладнішим питанням при синтезі мовлення. Зазначені синтезатори алофонного й дифонного синтезу (ті, що моделюють інтонацію) поки що не генерують просодію необхідної якості. Мовлення звучить неприродно, що в решті решт спричинює низький рівень сприйняття тексту. І хоча найпопулярнішими та найефективнішими вважаються дифонні синтезатори штучного мовлення, Лабораторією комп'ютерної лінгвістики КНЛУ було поставлено за мету створення системи складового синтезу англійського мовлення.

За своїм фонетичним виявом склад є найменшою вимовною одиницею. Він поєднує в собі сегментну і просодичну природу. Артикуляційні роботи, виконувані для реалізації такого цілісного вимовного зусилля як склад, одразу налаштовані на скоординовану якість поєднуваних у ньому сегментів – приголосних і голосних, на особливості реалізації наголошеності й ненаголошеності для всіх сегментів складу. Тон, інтенсивність, тривалість мають тенденцію до зміни в посладовому ритмі або в кількох складах поспіль. У результаті, при складовому синтезі меншу проблему становлять інтонація синтезованого мовлення та правила асиміляції сегментів.

Зазначені властивості складу стали першим лінгвістичним мотивом вибору цієї вимовної одиниці для конкатенативного синтезу англійського мовлення. Другим технічним мотивом стало міркування про те, що складовий синтез дає кращу розбірливість мовлення за рахунок меншої кількості точок конкатенації (у порівнянні з алофонним і дифонним синтезом).

Таким чином, одиницею синтезу було обрано склад, але склад не в системі мови, а в системі мовлення, де він може реалізувати себе як в межах словоформ, так і на міжсловесному шві, включати у свою структуру одну вершину (голосний сегмент) чи декілька вершин з підпорядкованими одному центру якостями тощо. Тому, порівнюючи зі складом у мові, його можна кваліфікувати як **квазісклад** і всі „ненормативно" реалізовані квазісклади включити до бази даних, що значно поповнює десятитисячний список складів англійського мовлення. Відповідний обсяг начитаних диктором текстів має становити 4–6 годин з включенням повторів для різних інтонаційних моделей.

Ураховуючи, що вибір та анотування такої великої кількості одиниць для бази даних на початковому етапі не автоматизовані і можуть гальмувати реалізацію інших модулів синтезатора, було вирішено опрацювати всі модулі на малому обсязі (550 одиниць) з можливістю поступового долучення приблизно 60 000 квазіскладів.

Як відомо, якість перетворення тексту різної складності й обсягу на синтезоване мовлення в системі будь-якого компілятивного синтезу, і зокрема складового, визначається переважно модулем лінгвістичного оброблення інформації, тобто прямопоропорційна якості й наповненню бази даних, тому що залежить від повноти аналізу різних характеристик текстової графічної інформації. Крім того, рівень розбірливості звучання синтезованого мовлення визначається точністю попередньої сегментації на обрані одиниці, що вносяться як елементи до бази даних, та власною якістю їх звучання.

База даних для аналізованого складового синтезу, що використовується програмою безпосередньо для генерації мовлення, складається з двох частин: інформація, що описує текстове представлення квазіскладу, та дані, що репрезентують його звуковий відповідник. База даних містить таку інформацію:

1. Квазісклад: текстове (орфографічне + пунктуаційне) представлення одиниці, інформація про словесний наголос та паузацію);
2. Позиція конкретного квазіскладу у фонетичному слові або ритмічній групі;
3. Лексичне слово або словоформа, з яких взятий конкретний квазісклад: прив'язка до лексеми чи словоформи є обов'язковою, оскільки саме за допомогою лексичного слова визначається звучання квазіскладу;
4. Фонетичне слово або ритмічна група, з яких взято конкретний квазісклад: вміщує інформацію про логічний наголос;
5. Назва звукового файлу, з якого береться звуковий інтервал конкретного квазіскладу;
6. Початок квазіскладу у файлі (у дискретах);
7. Кінець квазіскладу у файлі (у дискретах);
8. Тривалість квазіскладу (у дискретах).

Цієї інформації достатньо для неускладненого синтезу „текст на озвучення", а обмежена база даних дозволяє використовувати лише ті лексеми, словоформи і фонетичні слова, які представлені в її словнику.

У подібній базі даних складового синтезу англійського мовлення просодія частково вже закладена в самому складі та його позиціонуванні, тому не моделюється спеціально, а саме підбирається потрібний варіант квазіскладу для кожного конкретного випадку на основі реалізованих диктором інтонаційних моделей. На даному етапі інтонаційні моделі диктора враховують три основні типи речень: розповідне, питальне, спонукальне, – а також звертання й вигуки. Квазісклад з необхідними для синтезу просодичними параметрами обирається на основі таких притаманних йому ознак: наявність або відсутність словесного наголосу, наявність або відсутність логічного наголосу, позиція у фонетичному слові й ритмічній групі.

Процес добору необхідних складів для синтезу мовлення передбачає автоматичний складоподіл за законами мовленнєвої реалізації та маркування наголосу на відповідному складі слова з вхідного тексту. Цей перший етап нормалізації вхідного тексту спирається на словник слів із наголосами, що узгоджений з кількістю складових елементів. Отже, здійснюється пошук у базі відповідного текстового представлення складу та визначається його роль в конкретному лексичному слові чи словоформі. Якщо відповідника за цими двома вимогами не знайдено, на цьому етапі склад не відтворюється (на етапі поповнення бази виводиться повідомлення про відсутність інформації про конкретну одиницю у базі даних, а при необхідності прослуховування пошук повторюється для наступного слова без відтворення попереднього чи його частини). Якщо ж необхідний склад знайшовся, потрібна одиниця добирається за такими характеристиками: словесний наголос, позиція у фонетичному слові, логічний наголос. При синтезі інтонаційна модель вхідного тексту узгоджується з формальними ознаками пунктуаційних знаків, що супроводжують текстове представлення складу в базі даних. В ідеальному варіанті передбачається відтворення одиниці з наявністю всіх вище згаданих характеристик, але передбачається, що в багатьох випадках точного відповідника вхідного складу не знайдеться. Тому програма зорієнтована на пошук найкращого варіанту за послідовністю характеристик. З поступовим нарощуванням кількості елементів, які матимуть як мінімум по п'ять

варіантів реалізації, передбачено більше можливостей для створення слів, відсутніх у базі даних, і більш точного їх інтонування.

Окремим програмним етапом синтезу є конкатенація сегментів, маркування додаткових пауз та безпосереднє звукове відтворення послідовності складів.

Оскільки на даному етапі програма синтезу вміщує невелику базу даних, це унеможливлює відтворення власних назв, скорочень, абревіації, відсутніх у начитаному диктором тексті. Через обмеженість інтонаційних моделей можуть виникати проблеми із синтезом деяких видів питань та емоційно чи стилістично забарвлених речень. Має місце також самостійна проблема омонімії. Так, при синтезі мовлення необхідно розрізняти звучання омонімічних слів. Наприклад, англійське слово *read* матиме різне звучання в залежності від вживання у формах теперішнього чи минулого часу. Слово *breathe* може бути іменником або дієсловом і в залежності від цього вимовлятися по-різному. Частотність таких омонімів є достатньо високою.

У цьому синтезаторі, як і в інших відсутній модуль семантичного аналізу тексту. Варто зазначити, що проблема семантичного аналізу тексту може значно поліпшити синтез мовлення. Наприклад, якщо до звичайного розповідного речення '*My sister goes to school every day*.' поставити різні запитання:

What does your sister do every day?
Who goes to school every day?
When does your sister go to school?,

то відповідь кожного разу матиме різну інтонацію в залежності від зміни позиції логічного наголосу. Без блоку семантичного аналізу в багатьох випадках неможливо автоматично визначити місце логічного наголосу у фразі чи реченні, а отже, генерувати просодію, що відповідатиме семантиці природного мовлення. З цією метою необхідним вбачається етап семантичного аналізу тексту, тобто орієнтація на значення.

Література

1. Копелевич Ю. Х., Цверава Г. К. Христиан Готлиб Кратценштейн. 1723 – 1795. – Л., 1989.
2. http://offline.computerra.ru/1999/293/3579/
3. Dutoit T. An Introduction to Text-To-Speech Synthesis . – Springer, 1996. – 316 p.
4. Mark Tatham, Katherine Morton. Developments in Speech Synthesis. – John Wiley and Sons, 2005. – 328 p.

І. Біскуб, к. філол. н.[*]
Волинський державний університет імені Лесі Українки (Луцьк)
УДК 81'322.5

ЛІНГВІСТИЧНА ПАРАМЕТРИЗАЦІЯ ТЕКСТІВ ДЛЯ ТРЕНУВАННЯ КОРИСТУВАЧІВ У СИСТЕМІ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ АНГЛІЙСЬКОГО МОВЛЕННЯ DRAGON NATURALLY SPEAKING

The report deals with the complex analysis of the training session texts taken from Dragon Naturally Speaking program for automatic speech recognition. The new type of software – SEGMIX© is introduced in order to provide automatic text segmentation into words with statistics frequency calculations attached. The results of the frequency calculations are a analyzed from the point of view of artificial text vocabulary simplification which helps to optimize man-machine interaction.

Завдяки науково-технічному прогресу людство отримало автоматизовані системи управління, керування якими здійснюється за допомогою природної мови. З цією метою використовуються як усний так і письмовий варіанти мови. Керування різноманітними системами управління за допомогою природної мови є однією з актуальних проблем сучасної лінгвістики. Особової ваги набувають дослідження об'єктів на усіх мовних рівнях, оскільки інформація про інвентар та структуру цих об'єктів використовується у багатьох кібернетичних системах, що здійснюють взаємодію між людиною та електронно-обчислювальними машинами (ЕОМ).

Поява швидких ЕОМ значно збільшила потенціал інтелектуальної діяльності людини. ЕОМ довели свою здатність заміняти обчислювальну працю сотень і тисяч людей, що спричинило появу нових засобів і методів, які отримали назву кібернетичних.