Створення моделі урахування взаємовпливів ознак природності звучання мовленнєвих елементів для конкатенативного сегментивного синтезу

Крак Ю.В., Шатковський М.М

Інститут кібернетики ім. В.М. Глушкова НАН України
yuri.krak@gmail.com, mykolash@gmail.com

Abstract

Natural Ukrainian language attributes are analyzed and formalized; natural Ukrainian language features such as positioning, stressing and multi-phoneming are offered; natural Ukrainian language features interrelation model is developed; synthesis objects and elements selection approach are described.

1. Вступ

Засоби аналізу та синтезу звукових голосових мовних сигналів та мовного апарату людини розробляються ще з середини XX ст. в багатьох визначних наукових організаціях світу. Проблемами аналізу та синтезу голосової мовної інформації займаються наукові колективи в Міжнародному науково-навчальному центрі інформаційних технологій і систем [1] [2], Лабораторії систем та обробки сигналів Політехнічного факультету університету Монса [3], Дніпропетровському національному університеті [4], в Київському національному університеті імені Тараса Шевченка [5], в Державному університеті інформатики і штучного інтелекту [6] та у багатьох інших наукових центрах.

Серед найбільш розповсюджених та затребуваних практичних застосувань комп'ютерних усно-мовних технологій є людино-комп'ютерні інтерфейси, засоби IP-телефонії, комп'ютерні call-центри, телефонні системи автоматичного інформування, мовленнєве супроводження систем автоматизації прийняття рішень, системи голосового інформування в транспорті, голосові інструкції з налаштування апаратного та програмного забезпечення, інформаційно-довідкові програмні засоби, озвучування ігор, електронної пошти, субтитрів, читання електронних книг, системи комп'ютерного навчання, додатки для людей з вадами зору, системи комп'ютерного оповіщення, голосове озвучення банківських та інших комерційних даних тощо.

Перспективним напрямком використання систем комп'ютерного синтезу мовлення є керування складними пристроями та рухомими механізмами. В таких системах засоби людино-комп'ютерного природно-мовного діалогу дозволяють користувачу задіяти слух в якості додаткового дублюючого каналу контролю за ситуацією.

В синтезі мови, останні десятиліття найбільшого використання набув конкатенативний синтез мови.

Розповсюдженість даної концепції полягає в тому, що в основі такого синтезу лежать природні, вимовлені професійним диктором, корінним носієм мови, звукові елементи мови. Це призводить до досить високого рівня природності звучання синтезованого мовлення.

Системи комп'ютерного конкатенативного синтезу генерують вихідний акустичний сигнал на основі конкатенації (послідовного додавання) необхідних елементів синтезу. Процес конкатенації визначається структурою бази даних елементів синтезу, оскільки, безпосередньо залежить від природних даних, тим самим, забезпечуючи високу природність звучання синтезованої мови. Тому, при підвищенні природності звучання синтезованої мови буде зростати і розмірність елементної бази синтезу. Існує кілька стандартних підходів до вибору концепції формування мінімальних елементів синтезу – вибір фонів, алофонів, дифонів, складів і фонем-трифонів тощо. Природність та якість звучання такого підходу пояснюється тим, що в системах конкатенативного синтезу мовлення елементами синтезу є реальні природні мовленнєві сигнали, вимовлені професійними дикторами, корінними носіями мови.

2. Проблема урахування природності звучання мовлення в системах конкатенативного синтезу

Основною сучасною проблемою створення та вдосконалення засобів комп'ютерного конкатенативного синтезу українського мовлення є природність звучання синтезованого мовленнєвого сигналу та врахування просодичних властивостей природного мовлення. Особливу актуальність набувають дві взаємопов'язані проблеми дослідження мовлення – аналіз природного голосового мовлення та синтез характеристик природності його звучання.

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. При читанні вголос (і, як наслідок, в розмовній мові) послідовність фонем, що відповідає слову, може не співпадати з послідовністю літер, відповідних фонемам, тобто, відрізнятися від буквальної конкатенації фонем (фонів чи алофонів).

Для визначення просодичних параметрів при синтезі мови за текстом можуть бути використані різноманітні просодичні моделі, зокрема, автосегментна метрична модель, акустично-фонетична суперпозиційна модель, модель IPO, модель Tilt.

Проте, досвід практичного створення систем голосового озвучення текстової інформації українською мовою показав, що комп'ютерна модифікація властивостей природних мовленнєвих сигналів далеко не завжди покращує натуральність звучання синтезованого мовлення, хоча і являє собою цікаву та перспективну наукову проблему.

2.1. Аналіз природного українського мовлення

Вимова ізольованих голосних (і багатьох приголосних також) поділяється на три фази: екскурсія, витримка і рекурсія. В той час як, структура вимови неізольованого голосного інша. Артикуляція голосних звуків змінюється і в залежності від їх положення в словах та від характеру складів. Коли звуки входять до складу звукосполук і слів, то їх вимова, на відміну від окремої вимови тих самих звуків, стає зв'язною. Якщо, наприклад, вимовляються підряд один за одним кілька звуків, то в цьому випадку рекурсія приголосного здійснюється не до індиферентного положення, а зливається в один спільний момент з екскурсією наступного голосного [7].

Дослідження природного українського мовлення та аналіз властивостей його звучання показали необхідність врахування великої кількості взаємопов'язаних факторів, які впливають на акустичну, фонетичну та фонологічну структуру природних мовленнєвих елементів. Так, при аналізі природних мовленнєвих елементів та формуванні штучних елементів синтезу, серед інших властивостей, слід враховувати і такі:

- позиційну обумовленість елементів синтезу,
- наголошеність голосних звуків,
- пом'якшеність та подвоєність приголосних звуків,
- властивості коартикуляційних переходів між приголосними та голосними звуками,
- роздільність звучання приголосних та голосних звуків, збіги приголосних звуків та ін.

2.2. Вибір характеристик природності звучання українського мовлення

Виходячи з результатів проведених фонетичних досліджень українського мовлення, можемо навести такі ключові ознаки природності звучання українського мовлення:

- G_1 – ознака впливу приголосних звуків на голосні. Визначається багатьма фонетичними факторами, основними з яких вважатимемо – зв'язність вимови звуків, що входять до складу звукосполук та слів, обумовлення взаємного пристосування артикуляції звуків за рахунок взаємного накладання кінцевої та початкової фаз сусідніх звуків, виділення перехідних та основної частини голосних, лабіалізація приголосних перед [у], [о]. Дана ознака зумовлює використання складів в якості основних сегментів саме через коартикуляційні переходи між фонемами.
- G_2 – ознака твердості чи м'якості приголосних. В залежності від пом'якшення приголосного в слові за допомогою м'якого знаку чи пом'якшуючого голосного, звучання приголосного змінюється.
- G_3 – ознака роздільності звучання приголосних та голосних звуків. Наприклад, відкриті склади типу «приголосна»«'»«голосна».
- G_4 – ознака подвоєння і подовження приголосних. Подовжені приголосні мають характерне звучання, штучне створення якого являє собою складну та неоднозначну задачу з коартикуляційної модифікації приголосних.

- G_5 – ознака спрощення приголосних при вимові. Найчастіше спрощення відбуваються в суфіксах слів при збігові декількох приголосних.
- G_6 – ознака відмінності звучання наголошених та ненаголошених голосних. Зумовлює створення як наголошених, так і ненаголошених сегментів.
- G_p – ознака позиційності сегменту. Характеризує розташування сегменту в слові – на початку слова, в середині слова чи в кінці слова.

2.3. Взаємовплив ознак природності звучання українського мовлення

Згідно проведеного аналізу, звуки українського мовлення характеризуються високим рівнем взаємовпливів, наведемо деякі з них:

- на рекурсію приголосних накладається екскурсія голосних, в тому числі пом'якшуючи приголосні.
- при вживанні апострофу приголосні не пом'якшуються наступною голосною.
- коартикуляція наголошених голосних змінює просодичні та акустичні властивості приголосних інакше ніж ненаголошених голосних – збільшується амплітуда та кількість періодів основного тону приголосних (у випадку вокалізованих приголосних)
- екскурсія наголошених голосних накладається на рекурсію приголосних інакше ніж ненаголошених голосних, тим самим, змінюючи просодичні та акустичні властивості коартикулюючих звуків – збільшується амплітуда та кількість періодів основного тону приголосних (у випадку вокалізованих приголосних)
- подвоєні приголосні пом'якшуються наступними голосними.
- у випадках спрощення чи заміни фонем при вимові, часто відбувається пом'якшення набутих приголосних.
- у наголошених та ненаголошених випадках при спрощенні (чи заміни) приголосної в групі приголосних відбуватимуться різні коартикуляційні явища тощо.

На рис.1 наведено схему взаємовпливів всіх описаних ознак звучання елементів природного українського мовлення, окрім ознаки позиційності (зв'язок з якою є актуальним для всіх інших ознак):

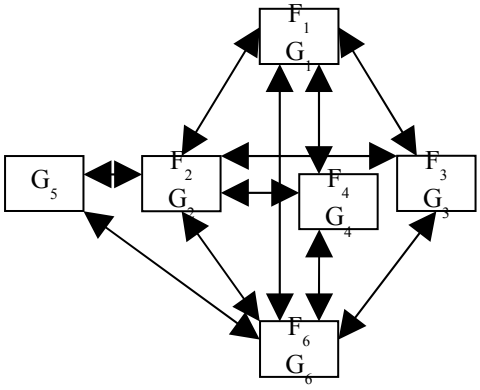

Рисунок 1: Схема ознак природності звучання українського мовлення.

Враховуючи 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 присутня в усіх відкритих складах української мови. Проте, G_2 враховується в коартикуляційних властивостях міжфонемних переходів. Іншими словами G_2 є важливою, але залежною від ознаки G_1 коартикуляційного впливу приголосних на голосні.

3. Модель урахування взаємовпливів ознак природності звучання українського мовлення

Ознаки G_1, G_2, G_3, G_4 та G_5 визначають властивості та взаємозв'язки звучання фонем українського мовлення. Таким чином ці ознаки формують властивості мультифонемності звучання українського мовлення. Тоді, рис.1 можна зобразити у вигляді рис.2, де G_s – ознака наголошеності голосного, а G_m – сукупність властивостей ознак G_1 – G_5:

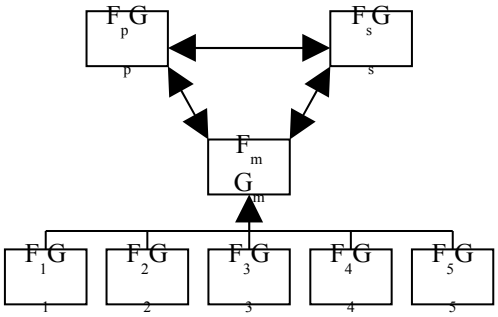

Рисунок 2: Схема взаємовпливів ознак природності звучання українського мовлення для конкатенативного сегментивного синтезу

Таким чином, згідно запропонованої схеми кожний сегмент одночасно визначається трьом ознаками природності звучання:

1. Позиційністю – початкові, внутрішні та кінцеві сегменти.
2. Наголошеністю – наголошені та ненаголошені сегменти.

3. Мультифонемністю – фонеми, склади, складно-синтезовані сегменти, морфологічно обумовлені сегменти.

Ознаки природності звучання сегментів не протиставляються в звучанні сегментів, а взаємодоповнюються, як наведено в прикладі в таблиці 1:

Таблиця 1: Взаємозв'язок властивостей звучання мовленнєвих сегментів

№	Сегмент	Позиційність	Наголошеність	Мульти-фонемність
1	#се	початковий	ненаголошений	складовий
2	г	внутрішній	ненаголошений	фонемний
3	ме_нт^	кінцевий	наголошений	мульти-фонемний
4	#по	початковий	ненаголошений	складовий
5	зи	внутрішній	ненаголошений	складовий
6	ці_	внутрішній	наголошений	складовий
7	й	внутрішній	ненаголошений	фонемний
8	ність^	кінцевий	ненаголошений	мульти-фонемний
9	#на	початковий	ненаголошений	складовий
10	го	внутрішній	ненаголошений	складовий
11	ло_	внутрішній	наголошений	складовий
12	ше	внутрішній	ненаголошений	складовий
13	ний^	кінцевий	ненаголошений	мульти-фонемний
14	#мульти	початковий	ненаголошений	мульти-фонемний
15	фо	внутрішній	ненаголошений	складовий
16	не_	внутрішній	наголошений	складовий
17	м	внутрішній	ненаголошений	фонемний
18	но	внутрішній	ненаголошений	складовий
19	го^	кінцевий	ненаголошений	складовий

Запропонована модель здійснює перехід від зазначених властивості звучання природного українського мовлення до ознак природності звучання вже комп'ютерно синтезованого мовлення. Та дозволяє здійснити урахування позиційних та коартикуляційних взаємовпливів цих ознак.

4. Вибір об'єктів та елементів синтезу на основі ознак природності звучання українського мовлення

Об'єктом синтезу розглядатимемо слово в зв'язку з тим, що мінімальним, семантично повним, носієм інформації є саме слово. Роль слова як центральної одиниці мови зумовлюється, також, залежністю інших одиниць мови (фонем, морфем, словосполучень, речень) від слова та місця в ньому і системними відношеннями зі словом.

Елементами синтезу розглядатимемо сегменти – певні морфологічно, фонетично та просодично обумовлені одиниці синтезу. База даних сегментів покриває всю множину слів української мови, що робить можливим голосовий синтез довільних слів української мови. Сегментами можуть бути фонеми, літери, склади, морфологічні одиниці чи їх частини, певні сполучення фонем

Слово розкладається на три частини, згідно з ознакою позиційності ~~Fp~~G_p – початкову, внутрішню та кінцеву. Причому, в серед сегментів містяться як наголошені сегменти, так і не наголошені, згідно ознаки наголошеності G_s.

Сегменти вибираються якнайбільшими (тобто такими, що складаються з якнайбільшої кількості літер/звуків), згідно ознаки мультифонемності G_m. Таким чином зменшується кількість конкатенацій сегментів і підвищується природність звучання синтезованої мови.

Зв'язуючими елементами синтезу є внутрішні сегменти – позиційно обумовлені фонеми та склади. Просодично характеризуючими, тобто такими, що підвищують природність звучання синтезованої мови, є початкові, кінцеві та мультифонемні сегменти – позиційно обумовлені морфологічні одиниці чи їх частини, фонеми, літери, певні сполучення фонем та склади.

В табл. 2 наведено приклади розкладу слів на сегментовані об'єкти синтезу:

Таблиця 2: Перехід від слів до об'єктів та елементів синтезу

№	Слово	Об'єкт синтезу	Сегменти
1	сегмент	#сегме_нт^	#се, г, ме_нт^
2	позиційність	#позиці_йність^	#по, зи, ці_, й, ність^
3	наголошений	#наголо_шений^	#на, го, ло_, ше, ний^
4	мультифонемного	#мультифоне_много^	#мульти, фо, не_, м, ного^

При створенні моделі врахування ознак природності звучання українського мовлення та побудові наборів сегментів було узагальнено три рівні лінгвістичного аналізу української мови.

Фонетичний рівень лінгвістичного аналізу української мови – аналіз найпростіших елементів мови – фонем та алофонів, їх акустично та перцептивно контекстно залежних реалізацій. В результаті до наборів сегментів були додані фонеми, з урахуванням їх положення в слові та наголошеності голосних.

Фонологічний рівень лінгвістичного аналізу української мови – аналіз можливих комбінацій фонем та складів. В результаті до множини сегментів були додані склади, з урахуванням їх положення в слові, пом'якшеності приголосних, йотованості та наголошеності голосних.

Морфологічний рівень лінгвістичного аналізу української мови – аналіз складів та морфем. В результаті до множини сегментів були додані складні морфологічно обумовлені початкові та кінцеві сегменти та мультифонемні сегменти.

Саме дослідження, узагальнення та опрацювання перших трьох рівнів лінгвістичного аналізу української мови в рамках комп'ютерного конкатенативного синтезу, дозволяє врахувати ознаки природності звучання слів української мови та конкатенувати мовні сегменти з їх подальшим природним, як структурно так і якісно, звуковим синтезом.

На основі запропонованої моделі розроблено модель подання текстової інформації для конкатенативного сегментивного синтезу та об'єктно-елементну модель конкатенативного сегментивного синтезу українського мовлення. Базуючись на підході, зазначеному в цих моделях, створено бази даних текстових сегментів, які покривають множину слів української мови. Також, створено тестові бази даних звукових сегментів використовуючи різних дикторів. Проведено ряд експериментів з комп'ютерного конкатенативного сегментивного синтезу слів української мови. Дані

експерименти продемонстрували високій рівень розбірливості та натуральності звучання синтезованих слів та показали важливість урахування зазначених ознак природності звучання.

5. Висновки

В роботі розглянуто проблему формалізації та врахування ознак природності звучання українського мовлення для задач комп'ютерного конкатенативного синтезу слів української мови. Наведено ознаки природності звучання українського мовлення та обґрунтовано їх вибір.

Проведено аналіз взаємовпливу ознак природності звучання українського мовлення. На основі якого, запропоновано модель врахування ознак природності звучання українського мовлення, яка дозволяє врахувати коартикуляційні та просодичні ознаки звучання українського мовлення та їх взаємовпливи з метою підвищення натуральності та розбірливості комп'ютерно синтезованих слів українського мовлення.

Обґрунтовано вибір об'єктів та елементів синтезу. Описано та продемонстровано перехід від слів до об'єктів синтезу та сегментів. Проведено стислий лінгвістичний аналіз проблеми врахування ознак природності звучання українського мовлення.

6. Література

[1] Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк — Київ: Наукова думка, 1987. — 261 с.

[2] Сажок М.М., Селюх Р.А., Юхименко О.А. Адаптація акустичних моделей фонем для пофонемного розпізнавання ізольованих слівукраїнської мови. // Штучний інтелект. – 2009. – №4 – С. 230–233.

[3] Dutoit T. MBR-PSOLA: Text-to-speech synthesis based on an MBE resynthesis of segments database / T. Dutoit, H. Leich // Speech Communication. — 1993. — № 13. — PP. 435—440

[4] Карпов О.Н. Технология построения устройств распознавания речи / О.Н. Карпов — Днепропетровск: Изд-во Днепропетр. ун-та. — 2001. — 184 с.

[5] Кривонос Ю.Г. Структура, свойства, характеристики объектов и элементов синтеза речи / Ю.Г. Кривонос, Ю.В. Крак, Н.Н. Шатковский // Компьютерная математика. — 2006. — № 1. — С.61—69.

[6] 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. — 2006. — № 4. — С.679—690.

[7] Ющук І. Фонетичні закони й орфоепія / І. Ющук // Українське мовознавство. — К.: Національний університет імені Тараса Шевченка. — 2004. — № 32–33. — С. 16—22.