

«ЗАСОБИ ОЗВУЧЕННЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ НА ОСНОВІ  
КОНКАТЕНАТИВНОГО СЕГМЕНТИВНОГО СИНТЕЗУ СЛІВ  
УКРАЇНСЬКОЇ МОВИ»

Інститут кібернетики імені В.М. Глушкова НАН України

К.т.н. Шатковський М.М.

[www.mykolash.com](http://www.mykolash.com)

[mykolash@gmail.com](mailto:mykolash@gmail.com)

Дослідженням проблем аналізу та синтезу мовлення займалися визначні вчені.

Серед них хочу відзначити таких науковців з України як:

- **Вінцюк Т.К., Сажок М.М.** – Міжнародний науково-навчальний центр інформаційних технологій і систем
- **Карпов О.М.** – Дніпропетровський національний університет
- **Шелєпов В.Ю.** – Державний університет інформатики і штучного інтелекту
- **Крак Ю.В.** – Київський національний університет імені Тараса Шевченка

Сучасні системи комп'ютерного синтезу мовлення є невід'ємною складовою засобів людино-комп'ютерного інтерфейсу інформаційних 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, IBM, Microsoft, Google, Ford, українська компанія PeopleNet та ін.);

- бібліотечні, довідникові, енциклопедичні комп'ютерні системи та web-системи (Texthelp Systems, ReadSpeaker, Ivona та інші);

- системи голосового виведення інформації для людей зі спеціальними потребами (Sakrament, Synapse, Freedom Scientific та ін.);

- підсистеми озвучення текстової інформації web-джерел та телекомунікаційні системи IP-телефонії (компанії Polycom, Adirac, Avaya, Cisco, Grandstream Networks та інші).

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

– **математичний (артикуляторний) синтез** – згідно якого створення штучних мовленнєвих сигналів відбувається на основі моделювання мовного апарату людини;

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

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

**Основною сучасною проблемою** створення засобів комп'ютерного конкатенативного синтезу мовлення є підвищення природності звучання синтезованих сигналів, якого можна досягти ґрунтуючись на аналізі мовленнєвих даних та врахуванні коартикуляційних та просодичних властивостей природного мовлення.

Огляду деякий таких проблем присвячено доповідь.

Специфіка української мови полягає у тому, що орфографічне подання та фонетичне (звукове) відтворення мови є досить наближеними. З іншого боку, фонемі 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 коартикуляційними переходами.

Така будова української мови дає змогу, за умови виділення та врахування певних фонетичних процесів, синтезувати мовлення не конкатенацією транскрибованих фонем (з їх подальшою просодично штучною модифікацією, наприклад, PSOLA-методами), а конкатенацією коартикуляційно та просодично природно зумовлених конструкцій, зокрема, сегментів, забезпечуючи тим самим натуральність та розбірливість звучання синтезованих слів без використання їх фонетичної транскрипції.

## Проблеми конкатенативного синтезу мовлення

Вибір одиниць синтезу (одиниць конкатенації).	→	Вибір фонем, алофонів, дифонів, складів, фонем-трифонів тощо.
Проблема стиків між елементами синтезу та генерування відповідних коартикуляційних переходів.	→	Зменшення кількості елементів синтезу (і, як наслідок, стиків) та штучний синтез переходів.
Врахування особливостей українського мовлення для проблеми конкатенативного синтезу.	→	Дослідження фонетичної складової української мови, вибір властивостей звучання.
Вибір об'єктів синтезу.	→	Слова, словосполучення, тексти.
Створення баз даних елементів синтезу та об'єктів синтезу.	→	Побудова, звукозапис та сегментація корпусу слів.
Конкатенація вихідного звукового потоку.	→	Створення комп'ютерних засобів конкатенації та озвучення мовленнєвого сигналу.
Аналіз звучання синтезованого сигналу.	→	Проведення експертного оцінювання природності звучання синтезованих об'єктів.

## **Виконання аналізу та врахування особливостей природного українського мовлення.**

1. Високий рівень співпадання орфографічного та фонетичного подання мовлення.
2. Позиційна обумовленість фонем у слові:
  - на початку слова,
  - в середині слова,
  - в кінці слова.
3. Високий рівень впливу приголосних на наступні голосні.
4. Відмінність звучання наголошених та ненаголошених голосних та складів
5. Твердість та м'якість звучання приголосних
6. Збіги приголосних при вимові
7. Роздільність приголосних з голосними.
8. Подвоєння/подовження приголосних
9. Наближення частки голосних звуків у мовному потоці до 50%

Таблиця 1. Ознаки природності звучання українського мовлення

Ознака	Опис ознаки
G1	Ознака впливу приголосних звуків на голосні. Визначається багатьма фонетичними факторами, зумовлює використання складів як основних сегментів через коартикуляційні переходи між фонемами
G2	Ознака твердості чи м'якості приголосних. Залежно від пом'якшення приголосного в слові за допомогою м'якого знака чи пом'якшувального голосного звучання приголосного змінюється
G3	Ознака роздільності звучання приголосних та голосних звуків. Наприклад, відкриті склади типу «приголосна»«'»«голосна»
G4	Ознака подвоєння приголосних. Подовжені приголосні мають характерне звучання, штучне створення якого являє собою складне та неоднозначне завдання з коартикуляційної модифікації приголосних
G5	Ознака спрощення приголосних при вимовлянні. Найчастіше спрощення відбуваються в суфіксах слів у разі збігу декількох приголосних
G6	Ознака відмінностей звучання між наголошеними та ненаголошеними голосними. Зумовлює створення як наголошених, так і ненаголошених сегментів
Gr	Ознака позиційності сегмента. Характеризує розміщення сегмента в слові – на початку, в середині чи в кінці



На рис.1 зображено схему взаємозв'язків різних ознак звучання елементів природного українського мовлення.

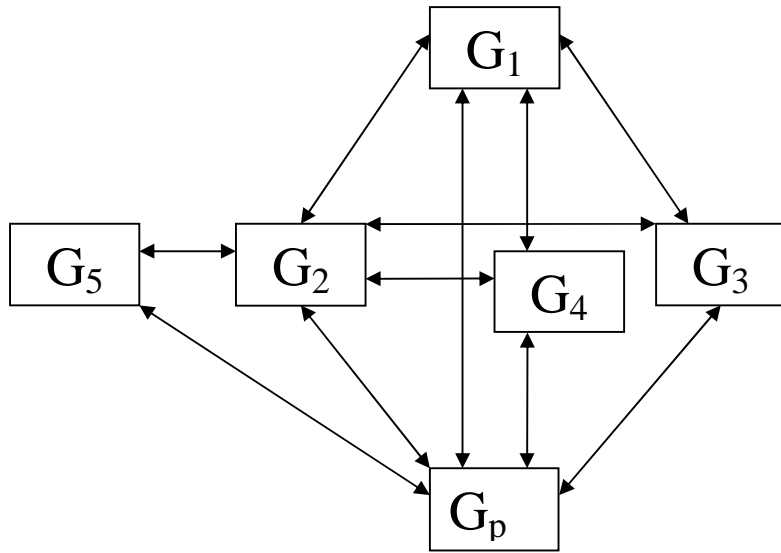


Рис. 1. Схема взаємозв'язків ознак природності звучання українського мовлення

Ознаки  $G_1 - G_5$  визначають властивості та конкатенативного сегментив-взаємозв'язки звучання фонем українського мовлення, тобто ці ознаки формують властивості мультифонемності звучання.

Звідси схему ознак природності звучання можна зобразити у вигляді рис.2, де  $G_s$  – ознака наголошеності голосного, а  $G_m$  – сукупність властивостей ознак  $G_1 - G_5$ :

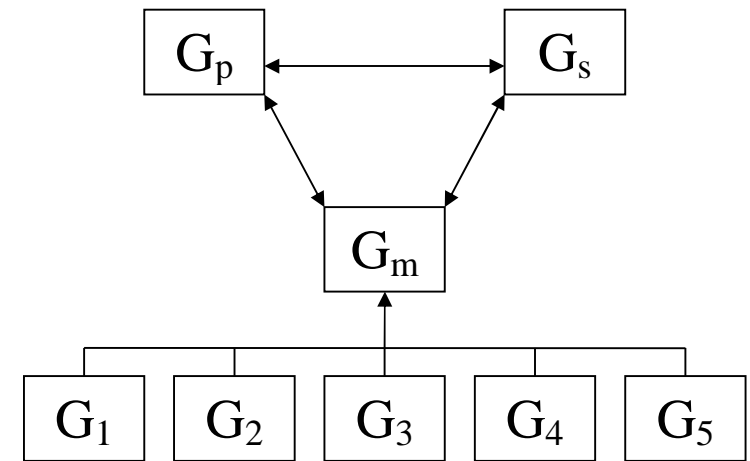


Рис. 2. Схема ознак природності звучання українського мовлення для

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, фонемі-трифони тощо, так і запропоновані в моделі сегменти спеціальної структури):

$$w_j = p_k i f_m,$$

**Об'єкт синтезу** складається із:

- текстового подання слова,
- послідовного набору сегментів  $(p_k, i, f_m)$
- кількості цих сегментів (рейтингу слова)  $R_j = |p_k, i, f_m|$ .

**Елемент синтезу** складається із:

- текстового сегменту,
- неупорядкованого набору слів  $(w_1, \dots, w_r)$  української мови
- кількості цих слів (рейтингу сегменту)  $r_t = |(w_1, \dots, w_r)|$ .

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. 3.

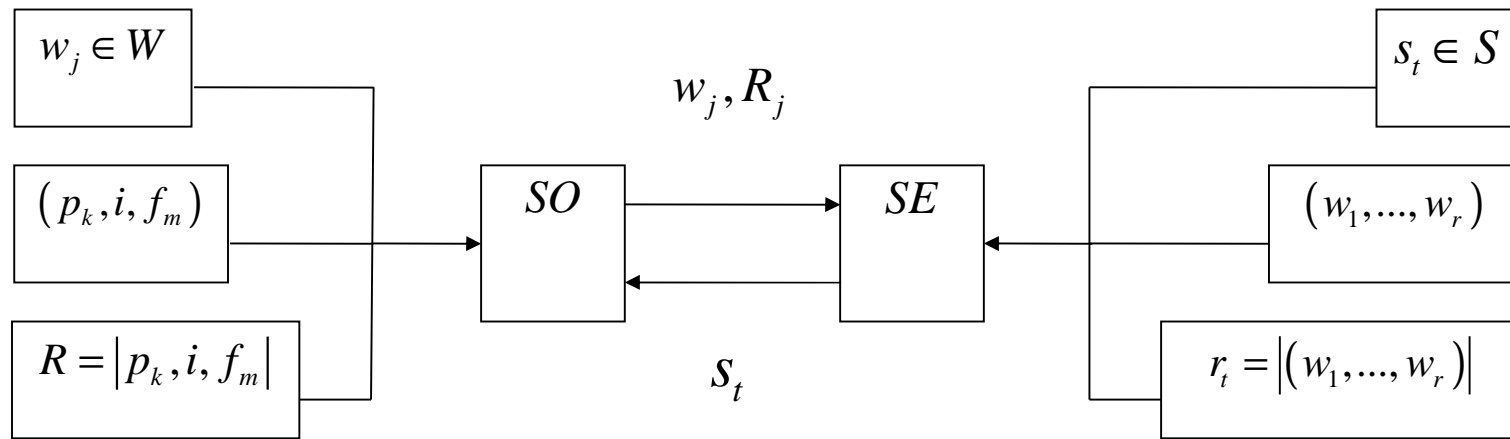


Рис. 3. Схеми об'єктно-елементної моделі подання інформації

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.

Розбірливість звучання базується на чіткому звучанні мовленнєвих сегментів, особливо, позиційно обумовлених складів.

Засобом підвищення розбірливості синтезованих слів виступають початкові та кінцеві сегменти.

В середині слів злитність забезпечується врахуванням коартикуляційних переходів між звуками в складах і мультифонемних сегментах.

Спектральних розривностей вдається уникати за рахунок вибору сегментів в моменти найбільшого амплітудного затухання мовленнєвих сегментів та використання мультифонемних сегментів.

Дякую за увагу!