

## РОЗРОБЛЕННЯ МОВЛЕННЄВИХ БАЗ ДАНИХ ДЛЯ СИНТЕЗУ МОВЛЕННЯ ЗА ТЕКСТОМ

*Розглядаються питання, пов'язані із розробленням мовленнєвих баз даних (МБД), які є важливим компонентом систем синтезу мовлення за текстом. Наведено принципи та етапи розроблення МБД, у відповідності до яких було створено 6 індивідуальних мовленнєвих баз даних. Наведено результати тестування синтезованого мовлення.*

### Вступ

На сьогодні найбільш поширеним методом синтезу мовлення за текстом є конкатенативний метод [1–4], який полягає в поєднанні (склеюванні) відрізків мовленнєвих сигналів, що зберігаються у мовленнєвих базах даних (МБД). Елементами МБД найчастіше бувають відрізки мовленнєвих сигналів, що відповідають таким фонетичним одиницям як фонемі (звуки), дифоні (відрізки мовленнєвого сигналу від середини однієї фонемі до середини наступної) та напівфоні (половини дифонів). В системі синтезу українського мовлення за текстом, розробленій у МННЦТіС, елементами МБД є фонемі у контексті, тобто про кожну фонему МБД відомо, якими були попередня та наступна фонемі. Послідовність елементів МБД, які вибираються під час синтезу мовлення, відповідають послідовності фонем тексту, що подається на озвучення. Послідовність фонем тексту разом з їх тривалістю та інтонаційним контуром (контуром частоти основного тону – ЧОТ) визначається під час аналізу вхідного тексту.

Мовленнєві бази даних, що використовуються у системах синтезу мовлення, не є спільними для носіїв тієї чи іншої мови, вони завжди індивідуальні та містять інформацію про особливості голосу та вимови окремих дикторів-донорів. Чим більше обсяг МБД, тим повніше представлено у ній фонемну (звукову), темпоральну та інтонаційну

варіативність мовлення диктора. Як наслідок, синтезоване мовлення звучить більш натурально. Крім того, чим більшим є обсяг МБД, тим меншими є спотворення мовленнєвих відрізків у місцях конкатенації внаслідок необхідної просодичної модифікації, яку застосовують для отримання відповідних тривалості та контура ЧОТ. Інакше кажучи, чим більшим обсяг МБД, тим більше ймовірність того, що в ній знайдеться елемент в необхідному контексті, з необхідною тривалістю та контуром ЧОТ.

В статті розглядаються принципи та етапи розроблення МБД, які використовуються під час настроювання системи синтезу мовлення на вимову диктора та безпосередньо під час синтезу мовлення за текстом.

Усього було розроблено 6 індивідуальних МБД: три чоловічі голоси та три жіночі. В статті в основному розглядаються МБД СВЯТОСЛАВ, ДМИТРО та ВАЛЕРІЙ обсягом відповідно 12000, 8500 та 13000 елементів.

## **1 Загальні принципи розроблення мовленнєвих баз даних**

**Вибір одиниць конкатенації.** В якості одиниць конкатенації обрано відрізки мовленнєвих сигналів, що відповідають одиницям мови – фонемам. На акустичному рівні фонемі спостерігаються не ізольовано, а у контексті з сусідніми фонемами. У мовленні фонемі зазнають змін акустичних характеристик внаслідок впливу артикуляції оточуючих фонем. З метою відображення в одиницях конкатенації змін акустичних характеристик, яких зазнають фонемі у мовленні внаслідок впливу артикуляції оточуючих фонем, елементами мовленнєвих баз даних було обрано фонемі-трифони [2], тобто фонемі у контексті.

У вигляді послідовності фонемі-трифонів слово „люди” буде виглядати так:

#-л'-У л'-У-д У-д-и д-и-#

**Вибір обсягу мовленнєвих баз даних.** Теоретично, якщо кількість фонем у мові дорівнює  $N$ , то кількість можливих фонем-трифонів дорівнює  $N^3$ . Було проведено дослідження мовленнєвого матеріалу, який складався з читання україномовних текстів трьома дикторами. На рис. 1 наведено залежність кількості різних фонем-трифонів від загальної кількості фонем-трифонів за даними мовлення дикторів СВЯТОСЛАВА, ДМИТРА та ВАЛЕРІЯ.

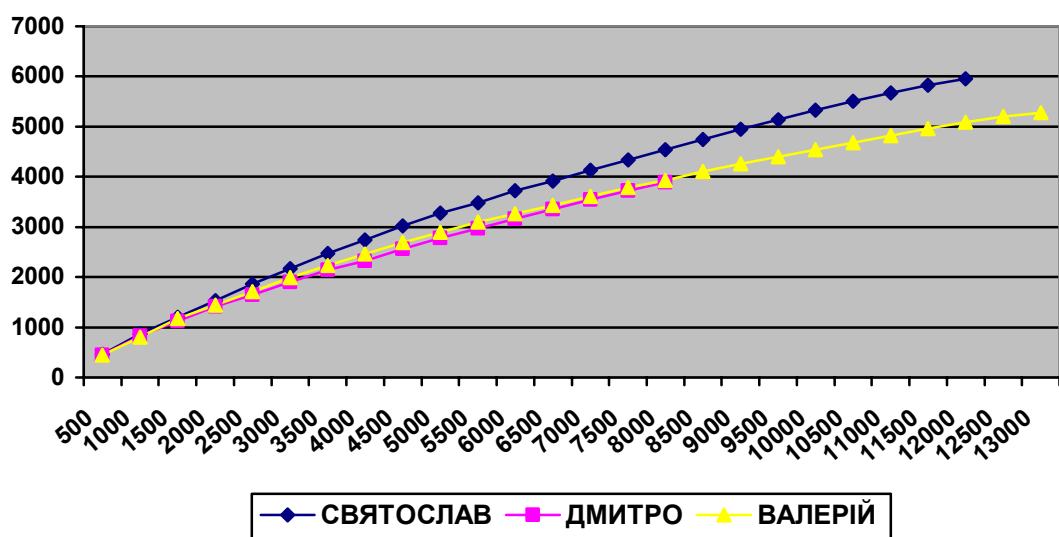


Рис. 1. Залежність кількості різних фонем-трифонів (вісь ординат) від загальної кількості фонем-трифонів (вісь абсцис) за даними мовлення дикторів СВЯТОСЛАВА, ДМИТРА та ВАЛЕРІЯ

Не всі з  $N^3$  теоретично можливих фонем-трифонів зустрічаються у мовленні. По-друге, деякі різні фонемні контексти виявляються схожими між собою з точки зору того впливу, який вони справляють на сусідні фонемі. Це означає, що під час синтезу можна вибрати замість фонемі-трифона іншу, схожу з нею за контекстом.

З другого боку, у мовленні фонемі-трифони зустрічаються у різних просодичних контекстах, і для озвучення текстів також потрібно мати фонемі-трифони у різних просодичних контекстах. Це означає, що для того, щоби під час синтезу була можливість використовувати „готові”

фонем-трифони, а не модифікувати їх, а отже, не спотворювати, у МБД одна і та сама фонема-трифон має бути присутня з різною тривалістю та різними значеннями ЧОТ.

Будемо розрізнявати МБД невеликого (до 3000 фонем-трифонів), середнього (від 3000 до 50000 фонем-трифонів) та великого (більше 50000 фонем-трифонів) обсягу.

Невеликий обсяг МБД призводить до необхідності просодичної модифікації фонем-трифонів, яка під час синтезу негативно впливає на натуральність синтезованого мовлення. Збільшення обсягу МБД – найкоротший та надійніший шлях до покращення якості синтезованого мовлення. Однак, великі МБД складно надійно розробляти, зокрема, сегментувати та анотувати. Окрім цього, записування великого обсягу мовленнєвого матеріалу потребує багато часу. Мовлення одного диктора, записане у різні дні, може відрізнятися за якістю. Нажаль, це залежить не стільки від технічних умов запису та обладнання, скільки від фізіологічних змін у мовленнєвому апараті людини та її психофізіологічного стану. Тому бажано записувати мовлення на протязі одного дня.

Середні МБД можуть бути відсегментовані та анотовані більш надійно. З другого боку, такі МБД містять фонем-трифони у різних просодичних контекстах. Середні МБД дозволяють ручну корекцію сегментації та анотації, як наслідок, анотації МБД детально і точно відображають вимову диктора. За даними анотації можна встановлювати закономірності фонетичних процесів та створювати моделі вимови.

**Відображення стилів мовлення.** Індивідуальна мовленнєва база даних повинна відображати не тільки голос та вимову диктора, а й стиль читання, який залежить від типу тексту, за яким проводиться запис мовленнєвого матеріалу.

**Автоматизація розроблення мовленнєвих баз даних.** Автоматизація прискорює розроблення та наповнення МБД, а також додає

цьому процесу послідовності, що позитивно впливає на роботу алгоритмів синтезу мовлення. На сьогодні етапи розроблення МБД автоматизовано різною мірою. Автоматизація проводиться шляхом формалізації та алгоритмізації дій експертів-фонетистів.

## 2 Етапи розроблення мовленнєвих баз даних

Мовленнєві бази даних було розроблено за наступним планом:

- визначення алфавіту фонем;
- вибір навчальних текстів;
- створення мовленнєвих корпусів;
- фонемне транскрибування мовленнєвих корпусів;
- створення МБД на підставі мовленнєвих корпусів;
- ручна/автоматична сегментація МБД на фонеми-трифони;
- автоматична сегментація фонем-трифонів на квазіперіоди основного тону (ОТ).

**Визначення алфавіту фонем.** Алфавіт фонем української мови містить: 6 наголошених голосних: /А/, /О/, /У/, /Е/, /И/, /І/; 6 ненаголошених голосних: /а/, /о/, /у/, /е/, /и/, /і/; 22 тверді (непалаталізовані) приголосні: /б/, /в/, /г/, /ґ/, /д/, /ж/, /з/, /к/, /л/, /м/, /н/, /п/, /р/, /с/, /т/, /ф/, /х/, /ц/, /ч/, /ш/, /дж/, /дз/; 23 м'які (палаталізовані) приголосні: /б'/, /в'/, /г'/, /ґ'/, /д'/, /ж'/, /з'/, /к'/, /л'/, /м'/, /н'/, /п'/, /р'/, /с'/, /т'/, /ф'/, /х'/, /ц'/, /ч'/, /ш'/, /дж'/, /дз'/, /й/; паузу: /#/.

Наголошені та відповідні ненаголошені голосні вважаються різними фонемами, оскільки мають розбіжні акустичні характеристики. Подвоєні та відповідні неподвоєні приголосні не розрізняються, оскільки фактично єдиною акустичною ознакою, що їх розрізняє, є тривалість.

**Вибір навчальних текстів.** Вибір навчального тексту пов'язаний з забезпеченням сегментного та просодичного покриття, зокрема, типів інтонації та темпу українського мовлення, що відображається у розмаїтті

текстових стилів. Для розроблення МБД було використано навчальні тексти повинні 5 типів:

- ізольовані слова;
- ізольовані фрази;
- довгі уривки з творів художньої літератури;
- тексти радіоноvin;
- інструкції (наприклад, для користувачів голосових диктофонів).

У таблиці 1 наведено розподіл текстового матеріалу за типами текстів.

Таблиця 1.

Типи текстів, на матеріалі яких створено мовленнєві корпуси

МБД	ізольовані слова	ізольовані фрази	тексти радіоноvin	інструкції	уривок з повісті
СВЯТОСЛАВ		+			
ДМИТРО			+		
ВАЛЕРІЙ	+	+	+	+	+

Текстовий матеріал частково перетинався, що було зроблено з метою можливості порівнювати відповідний мовленнєвий матеріал та виводити загальні сегментні та просодичні моделі. Так, частково перетинався текстовий матеріал у дикторів ДМИТРО та ВАЛЕРІЙ (радіоноvinи), а також у дикторів СВЯТОСЛАВ та ВАЛЕРІЙ (ізольовані фрази).

**Створення мовленнєвих корпусів.** Мовленнєвий матеріал диктора СВЯТОСЛАВА було накопичено в звичайному приміщенні. Мовленнєві сигнали записувалися безпосередньо в пам'ять комп'ютера в оцифрованому вигляді у форматі wav. Мовленнєвий матеріал диктора ДМИТРА було взято з Інтернету у вигляді mp3-файлів та перетворено у формат wav. Мовленнєвий матеріал диктора ВАЛЕРІЯ записувався у захищеному від шуму приміщенні з ретельним підбором типу мікрофону.

Слід зауважити, що тільки диктор ВАЛЕРІЙ є професійним диктором, що має досвід роботи у студіях звукозапису. Цей диктор вважається носієм орфоепічних норм української мови.

Мовленнєві файли зберігаються у wav-форматі з частотою дискретизації 22.050 кГц та розрядністю 16 біт. В залежності від типу тексту мовленнєві файли було розбито на окремі файли, що відповідають словам, реченням або абзацам. Ці окремі файли становлять елементи (одиниці) мовленнєвих корпусів. Кожен елемент мовленнєвого корпусу супроводжується описом у вигляді орфографічного тексту та фонемної транскрипції.

**Фонемне транскрибування мовленнєвих корпусів.** Для фонемного транскрибування було використано програму автоматичного розставлення наголосів [5] та програму автоматичного перетворення буква-фонема [6]. Ці засоби дають змогу отримувати нормативну фонемну транскрипцію навчальних текстів. Відповідність отриманої автоматичної транскрипції реальному мовленню було перевірено вручну за допомогою стандартних програм аналізу мовленнєвих файлів та автоматизованих засобів розроблення МБД [5], що дозволяють прослуховувати та бачити на екрані дисплею весь елемент мовленнєвого корпусу або його частини. Отримані в результаті транскрипції відображають особливості вимови дикторів.

У фонемній транскрипції вказані наголоси (наголошені голосні позначаються великими літерами) та паузи. Позначення меж між словами та реченнями відсутнє.

У склад опису елемента мовленнєвого корпусу входять:

- назва звукового файлу, в якому цей елемент зберігається;
- ім'я диктора, що начитував мовленнєвий корпус;
- орфографічний текст елемента мовленнєвого корпусу;
- реальна фонемна транскрипція елемента мовленнєвого корпусу.

Приклад елемента мовленнєвого корпусу:

Акустичний файл – “phrases\_2a\_2.wav”;

Диктор – ВАЛЕРІЙ;

Орфографічний текст – “Якою мовою будете говорити?”;

Фонемна транскрипція – /# й а к О й у м О в о й у # б У д е т е г о в о р И т и #/.

В таблиці 2 наведені дані щодо розроблених мовленнєвих корпусів.

Таблиця 2.

Загальні характеристики мовленнєвих корпусів

Мовленнєвий корпус	Обсяг пам'яті (Мб)	Кількість слів	Загальна тривалість мовленнєвого корпусу (з паузами) (хв.)
ізолювані речення (СВЯТОСЛАВ)	53.036	2211	21.57
тексти радіоновин (ДМИТРО)	26.759	1446	7.93
ізолювані слова (ВАЛЕРІЙ)	46.961	292	8.52
тексти радіоновин (ВАЛЕРІЙ)	70.533	1410	13.19
ізолювані речення (ВАЛЕРІЙ)	7.488	88	1.24
інструкції (ВАЛЕРІЙ)	7.764	94	1.28
уринок з повісті (ВАЛЕРІЙ)	440.803	9575	82.35

**Створення мовленнєвих баз даних на підставі мовленнєвих корпусів.** Будемо розрізняти мовленнєві корпуси та МБД. Різниця між мовленнєвим корпусом та МБД полягає в тому, що елементами мовленнєвих корпусів є слова, речення або абзаци, а елементами МБД є окремі фонем-трифони.

Для створення МБД на підставі мовленнєвих корпусів треба провести, по-перше, сегментацію елементів мовленнєвих корпусів на фонем-трифони, а по-друге, сегментацію фонем-трифонів на квазіперіоди ОТ. У МБД фонем-трифони знаходяться у тому ж порядку, що й у мовленнєвому корпусі.



Сегментація на фонемі-трифони виконується вручну або автоматично з можливим наступним ручним корегуванням. Сегментація фонем-трифонів на квазіперіоди ОТ виконується автоматично з можливим наступним ручним корегуванням. Сегментація мовленнєвого сигналу на квазіперіоди ОТ дозволяє отримати криву зміни ЧОТ.

Було розроблено дві версії МБД СВЯТОСЛАВ та дві версії МБД ВАЛЕРІЙ, відповідно із ручною та автоматичною сегментацією на фонемі-трифони.

### **Ручна сегментація мовленнєвих баз даних на фонемі-трифони.**

Ручна сегментація робилася на підставі фонемної транскрипції за допомогою графічного інтерфейсу, що дозволяє візуалізацію мовленнєвих сигналів, прослуховування їх, розставляння та редагування меж між послідовними фонемами-трифонами. Ця процедура, як і взагалі усі операції, пов'язані з конкатенацією мовленнєвих сигналів, виконується у амплітудно-часовому просторі за осцилограмою сигналу.

Принципи ручної сегментації: слуховий контроль, одноманітне з формальної точки зору віднесення фрагментів перехідних ділянок до тієї чи іншої фонемі-трифона, вибір меж між фонемами-трифонами у точках перетинання “нульової” лінії перед максимальним амплітудним піком.

У сумнівних випадках для більш точного визначення початку або кінця фонемі-трифона цей сигнал розглядається у спектральному просторі за допомогою стандартних засобів візуалізації та аналізу мовленнєвих сигналів WASP [7].

Нижче наводиться фрагмент анотації однієї з МБД (диктор ВАЛЕРІЙ). В анотації містяться результати сегментації. В першій колонці – ідентифікатор фонемі-трифона, у другій – ім'я фонемі-трифона, що складається з трьох частин (попередня фонема – поточна фонема – наступна фонема), у третій – тривалість фонемі-трифона у мс:

798	а-н-А	113.42
799	н-А-л'	219.32
800	А-л'-і	74.69
801	л'-і-з	129.71
802	і-з-#	154.74

### **Автоматична сегментація мовленнєвих баз даних на фонем-трифони.**

При порівняно незначній кількості реалізацій фонем-трифонів (до 1000) у МБД ручна сегментація на фонем-трифони потребує відносно небагато часу (тиждень). Якщо ж кількість реалізацій перевищує десятки тисяч, ручна сегментація стає неприйнятною. З метою прискорення формування МБД пропонується використовувати автоматичні процедури сегментації на фонем-трифони. Для цього будуються акустичні моделі фонем-трифонів та на основі цих моделей уточнюються границі потрібних сегментів. При виборі типу моделей фонем-трифонів спираємось на акустико-статистичний підхід до опису мовленнєвого сигналу, згідно з яким кожна фонема-трифон подається у вигляді станів, в яких опис мовленнєвого сигналу на інтервалі аналізу задається нормальним розподілом, а переходи між цими станами відбуваються з певними імовірностями. Параметри моделей оцінюються шляхом навчання.

При автоматичній сегментації фонем-трифони було наближено монофонами (фонемами-трифонами без контексту), вибрано їх алфавіт (58 штук) і кількість станів у кожній моделі (5 або 6, включно з вхідним і вихідним станами). Мовленнєвий сигнал описувався мел-частотним кепстром і його 1-ю і 2-ю похідними на інтервалі аналізу 25 мс з кроком 10 мс. Приховані параметри моделей монофонів були оцінені за 2 ітерації процедури навчання. З використанням отриманих моделей, були визначені границі між фонемами-трифонами.

Приклад автоматичної та ручної сегментації наведено на рис. 2.

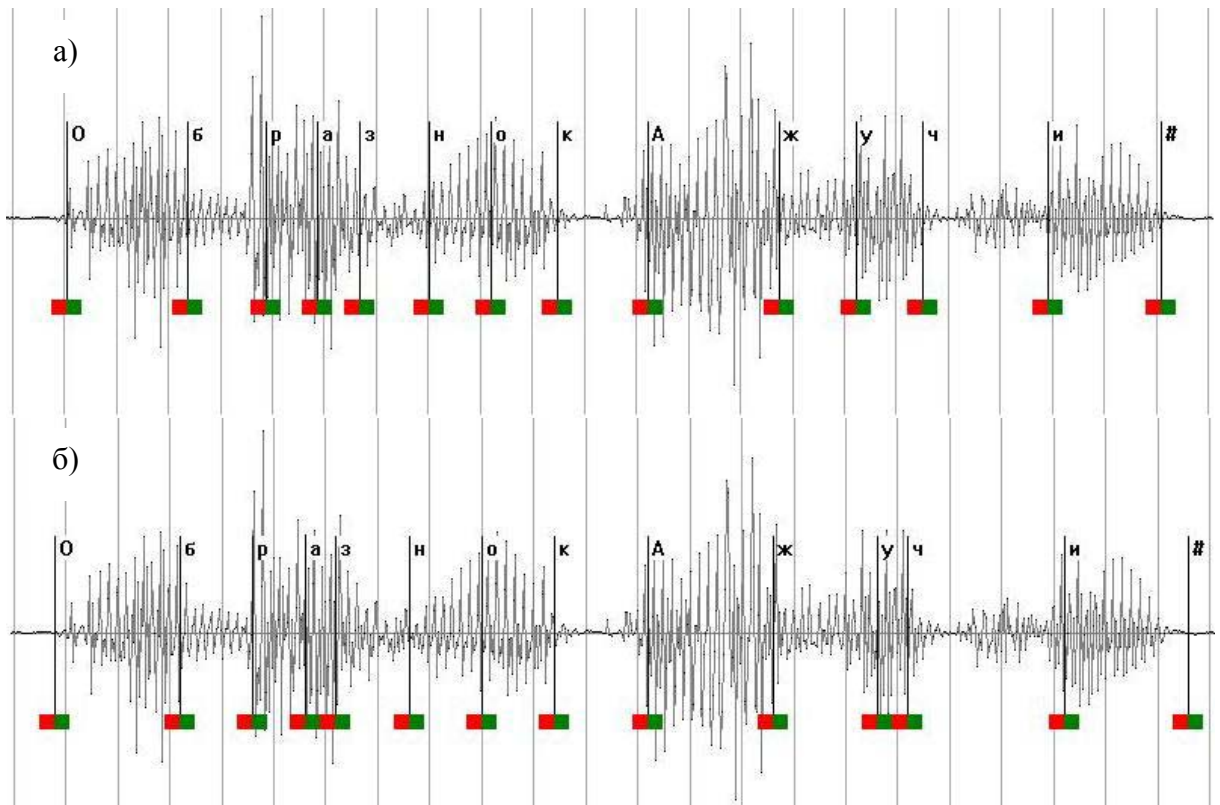


Рис. 2. Результати сегментації фрази «Образно кажучи» на фонемі-трифони: а) сегментацію проведено вручну, б) сегментацію виконано автоматичною процедурою.

### **Автоматична сегментація фонем-трифонів на квазіперіоди ОТ.**

Визначення довжин квазіперіодів ОТ фактично означає виділення послідовності значень ЧОТ, тобто, контура ЧОТ.

Кількість квазіперіодів ОТ кожного ФУСКА та їх довжин визначається автоматично за допомогою програмних засобів, розроблених на основі методу аналізу періодичності сигналу [5, 8].

На квазіперіоди ОТ сегментуються лише голосні та дзвінкі приголосні фонемі-трифони.

Передбачено також можливість ручного корегування помилок автоматичного виділення довжин квазіперіодів ОТ .

Нижче наводиться той самий фрагмент анотації МБД, що отримана вже після сегментації фонем-трифонів на квазіперіоди ОТ. Після тривалості фонемі-трифона наведено довжини її квазіперіодів ОТ в мс:

797	#-а-н	143.27	11.07	10.70	11.11	10.61	10.52	10.48	10.43	10.07	10.07	9.80	9.89	9.34	9.25	9.89
798	а-н-А	113.42	9.07	8.62	8.98	8.53	8.62	8.71	8.44	8.62	8.62	8.80	8.71	8.53	9.16	

799	н-А-л'	219.32	8.80	9.02	10.02	9.43	9.61	9.93	9.98	9.98	10.02	10.07	10.07	9.98	10.16	10.07
			10.07	9.98	10.25	10.16	10.70	10.79	10.52	9.66						
800	А-л'-і	74.69	10.79	10.34	10.07	10.34	10.52	11.88	10.70							
801	л'-і-з	129.71	10.79	10.43	10.61	10.52	10.70	10.70	10.79	10.79	10.98	10.88	11.16	11.29		
802	і-з-#	154.74	11.97	11.88	11.52	11.52	11.70	11.97	11.25	11.61	11.97	12.24	11.88	12.43	12.74	

На рис. 2 та рис. 3 наведено приклади відсегментованих фонем-трифонів, а саме, представлено результати сегментації першого голосного (/о/) фрази “Повторюйте команди після голосової підказки.”, прочитаної дикторами ВАЛЕНТИНОЮ та ВАЛЕРІЄМ. Довжини квазіперіодів ОТ наведено у дискретах.

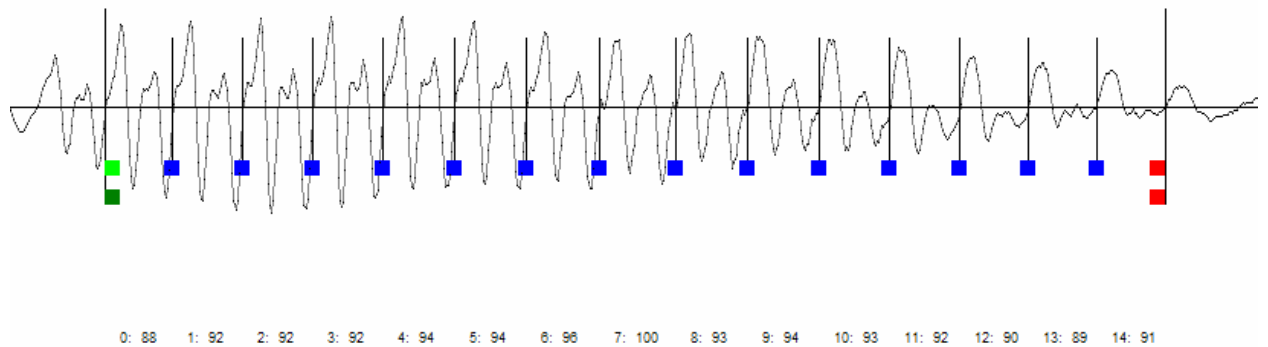


Рис. 2. Результат сегментації фонем-трифона „п-о-ф” у вимові диктора ВАЛЕНТИНИ

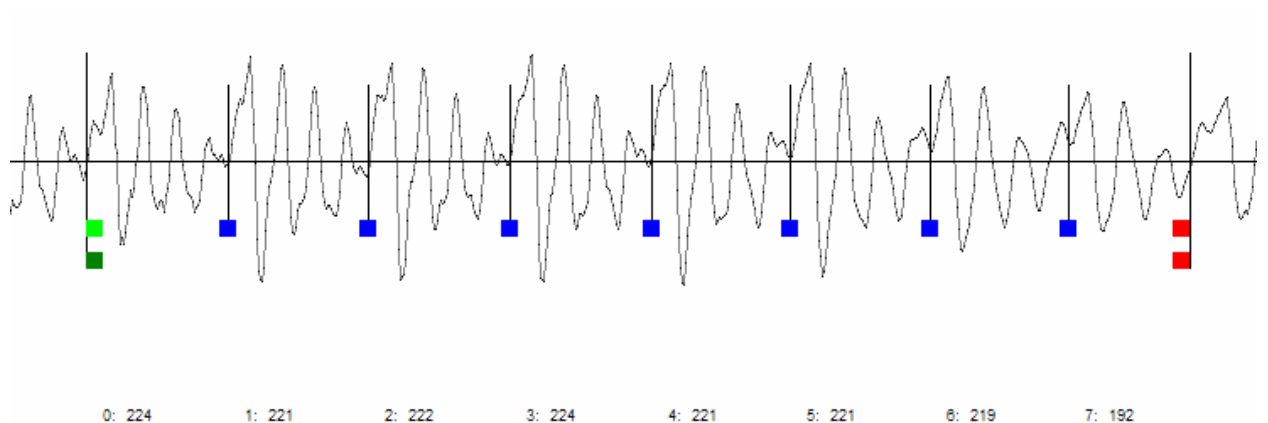


Рис. 3. Результат сегментації фонем-трифона „п-о-в” у вимові диктора ВАЛЕРІЯ

### 3 Анотація мовленнєвих баз даних

Від ступеня детальності та точності анотування мовленнєвої бази даних залежить настроювання лінгвістичного процесора на вимову окремих дикторів, від якого в свою чергу залежать якість та натуральність синтезованого мовлення. Принцип, який покладено в основу анотування МБД, полягає в тому, що анотація повинна бути максимально об'єктивною та детальною.

Анотації МБД використовуються не лише для настроювання лінгвістичного процесора на вимову дикторів, але й безпосередньо під час синтезу мовлення за текстом. Для прискорення цього процесу до інформації, про яку йшлося вище, до анотації додається інформація про кількість квазіперіодів ОТ та середнє значення довжини квазіперіоду ОТ. Ця інформація використовується модулем вибирання елементів, і присутність її в явному вигляді скорочує час обчислення мір схожості.

Нижче наведено фрагмент повної анотації фонем-трифонів мовленнєвої бази даних. Передостання колонка містить кількість квазіперіодів ОТ фонем-трифона, а остання – середнє значення довжини квазіперіоду ОТ.

797	#-а-н	143.27	11.07	...	9.89	14	10.20
798	а-н-А	113.42	9.07	...	9.16	13	8.71
799	н-А-л'	219.32	8.80	...	9.66	22	9.93
800	А-л'-і	74.69	10.79	...	10.70	7	10.66
801	л'-і-з	129.71	10.79	...	11.29	12	10.79
802	і-з-#	154.74	11.97	...	12.74	13	11.88

Використання МБД в процесі синтезу мовлення за текстом описано в [1].

### 4 Результати тестування розбірливості та натуральності синтезованого мовлення

Розбірливість синтезованого мовлення вимірюється процентом правильно розпізнаних слів. Для оцінки сприйнятої натуральності

використовується середня оцінка (MOS – mean opinion score). Було проведено декілька експериментів, пов'язаних із тестуванням синтезованого мовлення. Нижче наводяться результати двох з них.

**Тестування розбірливості синтезованого мовлення.** Перевірялась послівна розбірливість синтезованого мовлення. До тестування було залучено 15 осіб – студентів та викладачів Інституту філології та Лінгвістичного університету (3 чоловіки, 12 жінок, середній вік – 23 роки). Усі аудитори є носіями української мови. Тестові тексти, що подавалися на прослуховування, мали обсяг 364 та 281 слово. Аудиторам було запропоновано записати тексти, які вони почули. Тексти прослуховувалися аудиторами три рази. Розбірливість синтезованого мовлення (процент вірно розпізнаних слів) виявилася рівною 98.35%.

**Тестування натуральності синтезованого мовлення.** До тестування було залучено 15 аудиторів. Тестовий матеріал складав 2 зразки мовлення, синтезованого голосом одного диктора (ВАЛЕРІЯ) за одним і тим самим текстом, але з використанням МБД, отриманих різними способами: з автоматичною сегментацією на фонемі-трифони та квазіперіоди ОТ та з ручною сегментацією на фонемі-трифони і автоматичною сегментацією на квазіперіоди ОТ. Аудиторам було запропоновано прослухати зразки синтезованого мовлення три рази та оцінити їх якість за 5-бальною шкалою. Оцінка сприйнятої якості здійснювалася шляхом вирахування середнього балу. Середня оцінка якості мовлення, синтезованого з використанням МБД, отриманої шляхом ручної сегментації на фонемі-трифони та автоматичної сегментації на квазіперіоди ОТ, становила 3.67, в той час як використання МБД, отриманої за допомогою повністю автоматичної сегментації (на фонемі-трифони та квазіперіоди ОТ), становила 2.6.

## Прикінцеві положення

Якість синтезованого мовлення залежить від обсягу та покриття мовленнєвої бази даних, тобто від розмаїття звукових, темпоральних та інтонаційних варіантів мовленнєвих сигналів, які в ній містяться. Збільшуючи обсяг та покриття МБД, ми збільшуємо можливість знаходження в процесі синтезу мовлення за текстом таких елементів МБД, що мають необхідні характеристики (наприклад, знаходження елемента в потрібному контексті, із потрібною тривалістю та контуром ЧОТ). В результаті синтезоване мовлення звучить якісніше та натуральніше. Іншим важливим фактором, що впливає на якість синтезованого мовлення, є ступінь детальності опису мовленнєвих одиниць.

Важливим є продовження роботи над вдосконаленням автоматичної сегментації як на фонемно-трифони, так і на квазіперіоди ОТ, оскільки автоматизація процесів, що виконуються експертами, дозволяє значно прискорити розроблення нових МБД.

## Література

1. Людовик Т.В., Сажок Н.Н. Использование речевых баз данных большого объема при синтезе речи в системах искусственного интеллекта. // Проблемы управления и информатики, 6, 82-87, 2003.
2. Тарас Вінцюк, Тетяна Людовик, Микола Сажок, Руслан Селюх. Автоматичний озвучувач українських текстів на основі фонемно-трифонної моделі з використанням природного мовного сигналу // Праці 6-ї Всеукраїнської міжнародної конференції “Оброблення сигналів і зображень та розпізнавання образів” – УкрОбраз’2002, Київ, 2002, с. 79–84.
3. Лобанов Б.М., Карневская Е.Б., Левковская Т.В. “Синтезатор речи по тексту как компьютерное средство «клонирования» персонального голоса”, Тр. Международной конференции Диалог-2001, 265-272, М., 2001.
4. Hunt, A. and Black, A. Unit selection in a concatenative speech synthesis system using a large speech database // Proceedings of ICASSP 1996, vol 1, pp. 373-376, Atlanta, Georgia.
5. Микола Сажок. Автоматизовані засоби формування баз даних і знань для озвучення українських текстів. АКД, Київ, 2004, 20 с.
6. T.Vintsiuk, T.Liudovyk, M.Sazhok, "Phonetic Knowledge Base for Ukrainian". - In: "SPECOM'98. Proceedings of the International Workshop "Speech and Computer", St.-Petersburg, Russia, 26-29 October 1998, pp. 279-282.
7. <http://www.phon.ucl.ac.uk/resources/sfs/wasp.htm>.

8. Taras Vintsiuk. Оптимальне розбиття сигналів на квазіперіодичні та неперіодичні сегменти // Праці 3-ї Всеукраїнської міжнародної конференції “Оброблення сигналів і зображень та розпізнавання образів” – УкрОбраз’1996, Київ, 1996, с. 43-47.