АВТОМАТИЧНИЙ ОЗВУЧУВАЧ УКРАЇНСЬКИХ ТЕКСТІВ НА ОСНОВІ ФОНЕМНО-ТРИФОННОЇ МОДЕЛІ З ВИКОРИСТАННЯМ ПРИРОДНОГО МОВНОГО СИГНАЛУ

Тарас Вінцюк, Микола Сажок,Тетяна Людовик, Руслан Селюх

Міжнародний науково-навчальний центр інформаційних технологій та систем
40 просп. Академіка Глушкова, Київ 03680
Електронна пошта: {vintsiuk, mykola, tetyana_lyudovyk, selyukh}@uasoiro.org.ua

Taras Vintsiuk, Mykola Sazhok, Tetyana Lyudovyk, Ruslan Selyukh. **Automatic Ukrainian Text-to-Speech System Based on Phoneme-Threephone Model Using Natural Spoken Signal.** The text-to-speech system in time domain for Ukrainian is described. The concatenated acoustic elements are chosen in accordance to phoneme-threephone model for speech synthesis. Acoustical data is taken from the speaker voice passeport. The computerized tools for speech synthesis research and development are described.

1 Вступ

Озвучення текстів і сьогодні залишається важливою й актуальною задачею усномовної інформатики. Попри значні успіхи в синтезуванні мовних сигналів, як і в інших напрямках усномовної інформатики [1], озвучення текстів досі не є розповсюдженою технологією. Це зумовлено зокрема тим, що підвищились вимоги до якості та натуральності звучання синтезованої мови. Разом з тим для впровадження синтезу в портативних пристроях повинна задовольнятися вимога на обмеження швидкодії та обсяги пам'яті.

Пропонований озвучувач українських текстів засновується на фонемно-трифонній моделі розпізнавання та синтезу мовних сигналів, синтез сигналу відбувається у часовому просторі, при цьому використовується природномовний акустичний матеріал з усномовного файлу диктора. Це дозволяє максимально зменшити внесення спотворень у згенерований сигнал та значно розвантажити обчислювальний модуль.

2 Фонемно-трифонна модель синтезу мовних сигналів

Поняття фонеми надто абстрактне і погано піддається формалізації в чистому вигляді. Натомість, пропонується формалізм, у якому взято до уваги мінливість сигналу фонеми, зумовлену сусідніми в потоці мовлення звуками або коартикуляцією. Отже, вводимо поняття фонеми-трифона, коли розглядається фонема в контексті з попередньою та наступною фонемами. Поняття фонеми-трифона покладається в основу акустичних моделей як розпізнавання, так і синтезу усної мови.

Фонемно-трифонна акустична модель усної мови дозволяє врахувати явище коартикуляції, що виникає при взаємодії звуків у потоці мовлення. Справді, при відтворюванні послідовності звуків, що відповідають певним фонемам, рухи мовного апарату людини відбуваються з певною інерційністю. Стан мовного апарату перед наступним рухом залежить від попереднього звуку (фонеми), а отже і динаміка рухів мовного апарату при переході до наступної фонеми різна в залежності від попередньої фонеми, а це, зрештою, якісно відбивається на акустичному сигналі. Водночас, весь мовний апарат немов би готується до наступного звуку, і завершує попередній звук у стані, з якого більш вигідно переходити до звуку, який слідує.

Формалізм фонеми-трифона, який пропонується, дозволяє також конкретизувати поняття границь між фонемами, точніше, між фонемами-трифонами. Таким чином, незначне відлуння попереднього звуку та ледь помітна "чутність" наступного є допустимими при розставлянні границь фонем-трифонів. Уцілому, розставляння границь між фонемами-трифонами залишається досить складною задачею, та все ж більш окресленою, ніж для фонем.

Набір фонем-трифонів, як і набір (алфавіт) фонем для кожної мови є унікальним. Фонемно-трифонна транскрипція формується на основі фонетичного тексту за універсальним правилом стикування фонем-трифонів [2].

Виходимо з того, що задано скінченну множину K, куди входять фонеми $k \in K$, які спостерігаються в природній мові [3]. До алфавіту включено також фонему-паузу #. У множині K для української мови розрізняємо наголошені та ненаголошені голосні, м'які та тверді приголосні: $k \in \{$А, О, У, Е, И, І, А1, О1, У1, Е1, И1, І1, Б, Б', В, В', Г, Г', Ґ, Ґ', Д, Д', Ж, Ж', З, З', Й, К, К', Л, Л', М, М', Н, Н', П, П', Р, Р', С, С', Т, Т', Ф, Ф', Х, Х', Ц, Ц', Ч, Ч', Ш, Ш', ДЗ, ДЗ', ДЖ, ДЖ', #$\} \equiv K$ – загалом 57 фонем.

Фонема-трифон $t=uWv$ є фонемою W, яка розглядається під впливом сусідніх фонем u та v. Першою є u, що передує W, а другою – v, що слідує за W. За правилом допустимих сполучень фонем-трифонів [2] допустимими для з'єднання є лише фонеми-трифони вигляду $t_1=uWv$ і $t_2=wVz$ через Wv та wV.

Загальна кількість фонем-трифонів у алфавіті теоретично дорівнює кількості базових фонем у степені три (125000 для алфавіту з 50 фонем). Практично ж можна обмежитися декількома тисячами фонем-трифонів. Відсутні фонеми-трифони замінюються найближчою згідно з фонемно-трифонною ієрархією.

З природи усномовного сигналу випливає, що дзвінкі фонеми можна описувати (транскрибувати) як послідовність одно-квазіперіодичних мікросегментів, що мають певну форму звукової хвилі та довжину (Рис. 1). Такий опис поширюємо також і на глухі (шумні) фонеми і називатимемо акустичною транскрипцією фонеми-трифона. Послідовність хвиль одноквазіперіодичних мікросегментів утворює акустичний образ (прототип) фонеми-трифона. Сукупність всіх акустичних образів усіх фонем-трифонів, властивих певній людині, складає усномовний файл диктора [2].

Розглянемо процес автоматичного озвучення тексту. Спочатку, проводиться розбиття тексту за схемою: абзац—речення—синтагма (інтонаційна група)—ритмогрупа (група наголосу)—фонетичне слово. З абзаців виділяються речення, кожне речення розбивається на синтагми, ті, в свою чергу, – на ритмогрупи. І вже всередині ритмогрупи визначаються фонетичні слова, які або збігаються з орфографічними словами або містять їх декілька, включно зі службовими словами. Потім з використанням фонетичних знань про усну мову орфографічний текст перетворюється на фонетичну та одночасно й на фонемно-трифонну транскрипції за універсальним правилом. Далі згідно розпізнаних типів синтагм будується інтонаційний контур, розраховуються тривалості поточних одноквазіперіодичних мікросегментів та фонем уцілому. При цьому враховується, що кожна синтагма містить лише один основний наголос, який відповідає ядерній ритмогрупі. Решта ритмогруп – початкова, перед'ядерні та післяядерні.

На підставі розрахунків приступаємо до власне формування (синтезу) усномовного сигналу. Для кожної фонеми-трифона з фонемно-трифонної транскрипції обирається акустичний образ (прототип) з бази даних. Цей прототип піддається перетворенням згідно розрахованої тривалості фонеми-трифона та інтонаційного контуру. В результаті цих перетворень маємо отримати визначену перед цим кількість квазіперіодів розрахованої довжини. Перетворені прототипи квазіперіодів і загалом фонем-трифонів об'єднуються, і отриманий в результаті сигнал подається на засоби озвучення.

Визначальною рисою кожної технології синтезу усномовного сигналу є алгоритм змінювання просодичних характеристик прототипів елементів компіляції. Основною метою при зміні просодики, тобто інтонації та темпу, є досягнення якомога кращої якості синтезу за найменших обчислювальних витрат на кожну дискрету синтезованого сигналу.

Як видно з аналізу відповідних алгоритмів [4], у часово-амплітудній області досягається і те й інше. Так, вельми прийнятна якість синтезу за технологією *PSOLA* досягається за досить скромних витрат на обчислення при зміні інтонаційних характеристик прототипу елемента компіляції – до 9 арифметичних дій на одну дискрету. У технології *MBROLA* обчислювальні витрати ще скромніші – 6 арифметичних дій на одну дискрету, і це при кращих показниках якості синтезу, ніж у *PSOLA*. Технології *Unit-Selection* взагалі не передбачають яких-небудь інтонаційних змін прототипу сигналу, хоч при цьому і виникають надмірні витрати пам'яті на зберігання всіх можливих інтонаційних проявів кожного елементу компіляції, що конкатенуються без жодних перетворень сигналу. Це в свою чергу позбавляє певної гнучкості саму систему синтезу усної мови.

Пропонується ще один спосіб компіляції одноквазіперіодичних мікросегментів у амплітудно-часовій області. В основі цього методу закладено модель лінійного прогнозування сигналу, що дозволяє за певною кількістю попередніх відкліків сигналу спрогнозувати наступні з достатньо високою точністю апроксимації:

$$\widetilde{f}_n = -\sum_{s=1}^{s=m} a_s f_{n-s} + \varepsilon_n \,, \qquad (1)$$

де $\widetilde{f}_n$ — відліки прогнозованого сигналу, f_n — відліки спостережуваного сигналу, a_s, $s = 1{:}m$ — параметри передбачення, кількість яких m обирається в межах від 10 до 20, ε_n — похибка прогнозування. Параметри передбачення оцінюються на інтервалі аналізу рівному одному або двом квазіперіодам шляхом, наприклад, мінімізації суми квадратів похибки прогнозу.

Отже, нехай розрахуок тривалостей фонем та довжин квазіперіодів або мікросегментів і їх кількості в кожній фонемі вже виконано. Тоді відповідний прототип фонеми-трифону необхідно піддати темпоральним змінам, тобто привести до розрахованої довжини. Це досягається шляхом доведення кількості квазіперіодів до розрахованої. Таким чином, темпоральні зміни прототипу фонеми-трифону здійснюються шляхом викидання або повторення певних мікросегментів. У свою чергу, інтонаційні зміни сигналу здійснюються внаслідок скорочення або збільшення довжин відповідних квазіперіодів.

При темпоральних змінах, які вимагають подовження або скорочення тривалості прототипу фонеми-трифона до заданої довжини, обчислюється кількість мікрофонем (квазіперіодів), на яку їх необхідно збільшити або зменшити. Збільшення кількості мікрофонем відбувається шляхом повторення деяких квазіперіодів прототипу фонеми-трифону певну кількість разів. Зменшення кількості квазіперіодів здійснюється за рахунок викидання певних квазіперіодів.

Визначення квазіперіодів (мікрофонем), які повторюємо або викидаємо при темпоральних змінах, не є однозначним і потребує додаткових досліджень. Тому маємо право обрати одне з простіших рішень, а саме повторюємо (викидаємо) переважно центральні квазіперіоди.

Як зазначалося, інтонація сигналу регулюється шляхом задання певної довжини квазіперіодів на вокалізованих ділянках синтезованого сигналу. Для того, щоб подовжити окремо взятий квазіперіод згідно моделі лінійного прогнозування, достатньо знати

певну кількість попередніх відліків та значення коефіцієнтів прогнозування на відповідному інтервалі аналізу. Решту відліків, що доповнюють квазіперіод до потрібної довжини, обчислюємо за алгоритмом (1). Скорочення довжини квазіперіоду до заданої здійснюється шляхом видкидання відліків квазіперіоду понад задану довжину.

3 Комп'ютерні засоби дослідження та розроблення фонемно-трифонної моделі синтезу мовлення

Розроблено комп'ютеризовані засоби формування бази даних і знань, які використовуються як для синтезу, так і для пофонемного розпізнавання усної мови. З їх допомогою також проводяться експериментальні дослідження розпізнавання та синтезу.

При формуванні баз даних і знань, опрацюванні методів та алгоритмів, що стосуються усномовної інформатики, розроблені програмні засоби забезпечують виконання таких дій:

- введення фонетичних специфікацій, що включають опис базових фонем та перелік можливих інтонацій для обраної природної мови;
- накопичення навчальної вибірки за заданим текстом;
- автоматична сегментація навчальної вибірки на одноквазіперіодичні мікросегменти та на квазіперіодичні й неперіодичні сегменти;
- автоматизована сегментація навчальної вибірки на фонеми-трифони;
- дослідження та розроблення методів та алгоритмів синтезу усної мови.

Початкова інформація про природну мову складається з алфавіту базових фонем та переліку інтонацій. Задаємо її в режимі діалогу. Спочатку вводимо потрібну мову до списку підтримуваних мов або активізуємо відповідну мову, якщо вона вже є в реєстрі. В окремій секції задається алфавіт базових фонем та опис кожної з них. Передбачено підтримку паралельних алфавітів у кодуванні як кирилицею, так і латинськими літерами. Окреме поле відведене для задання символу, який розділяє фонеми. Нарешті, передбачено секцію, в якій опису-ються можливі інтонації та відповідні їм символи.

Таку початкову інформацію задаємо як для окремої мови, так і для групи мов одночасно у випадку, якщо в поставленій задачі є елементи багатомовності.

Навчальною вибіркою називаємо сукупність наговорених диктором звукових файлів разом з відповідними орфографічними або фонетичними текстами, які є окремими словами або фразами. Звукові файли з навчальної вибірки ще називатимемо реалізаціями слова або фрази.

Накопичення навчальної вибірки за заданим текстом виконується за допомогою вікна накопичення навчальних вибірок. На початку задається ім'я (ідентифікатор) диктора та вводиться текст, орфографічний або фонетичний, за яким проводиться запис навчальної вибірки. Передбачено можливість запису реалізацій слів або фраз в окремі файли для зручності при подальшому обробленні даних.

Запис навчальної вибірки відбувається за такою схемою:

1. Мишкою активізується текст реалізації, з якої розпочинаємо запис.
2. Диктор натискає кнопку запису *Record* і, утримуючи її, промовляє відповідний текст.
3. Після промовляння тексту диктор відпускає кнопку запису.
4. Автоматично активізується наступна реалізація, якщо список не вичерпано. Вручну можна вибрати будь-який інший елемент.
5. Для запису наступної реалізації переходимо до кроку 2. Якщо записуваний текст вичерпано, завершуємо роботу.

Записаний усномовний сигнал прослуховує експерт з метою виявлення бракованих або невиразних реалізацій, які позначаються як "погані" (*bad*). Остаточно, "погані" реалізації перезаписуються за вищенаведеною схемою з тією лише різницею, що автоматично активізуватиметься наступна відбракована реалізація.

В результаті процедури запису навчальної вибірки сформуються файли, що відповідають реалізаціям

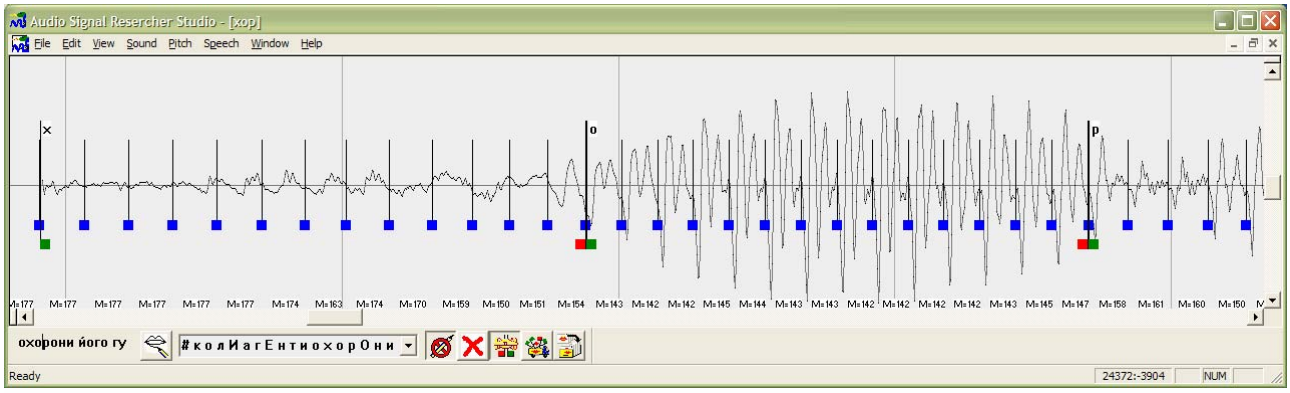

Рис. 1. Приклад зображення ділянки відсегментованої реалізації навчальної вибірки з чоловічого голосу. Сигнал розглядається крізь одноквазіперіодичну сітку, "натягнуту" як на періодичні сегмети, що відповідають фонемам-сонорантам, так і на неперіодичні сегменти, що відповідають глухим та паузі. Цифри знизу сигналу позначають тривалість квазіперіода в дискретах.

навчальної вибірки, яким співставлено ім'я диктора та текст, що було вимовлено. Далі проводимо сегментацію реалізацій навчальної вибірки на фонеми-трифони, а вибірки окремих фонем-трифонів – на одноквазіперіодичні мікросегменти та квазіперіодичні й неперіодичні сегменти. Причому, порядок сегментування може бути і зворотній.

На Рис. 1 показано приклад результату сегментування реалізації з навчальної вибірки. Ділянці усномовного сигналу співставлено відповідні фонеми-трифони, та відображено границі одноквазіперіодичних мікросегментів.

Сегментація навчальної вибірки на фонеми-трифони проводиться за участю експерта. Програмне забезпечення автоматично розставляє границі фонем лише дуже приблизно. Далі експерт, керуючись звуковою та візуальною підказкою уточнює межі фонем.

Пересування границь фонем здійснюється у двох режимах. Перший режим передбачає незалежне пересування границь сусідніх фонем, тобто можливі ділянки сигналу, на яких фонеми "накладаються" одна на одну, або ділянки, які не належать жодній фонемі. Коли активований другий режим, шляхом переведення кнопки *Linked Marks* у натиснутий стан, границі сусідніх фонем пересуваються узгоджено: початок поточної фонеми збігається із закінченням попередньої. В обох випадках за наявності одноквазіперіодичної розмітки границі фонем автоматично синхронізуються з границями квазіперіодів, тобто відбувається автоматичне прив'язування границь фонем-трифонів до одноквазіперіодичної сітки.

При уточнюванні границь фонем експерт використовує режим візуалізації сигналу, в якому зображується авторегресійний спектр, та режим прослуховування (*Sound->Sound Phone*), при якому озвучується фонема, на яку користувач-експерт вказує лівою кнопкою мишки.

Сегментація навчальної вибірки на одноквазіперіодичні мікросегменти та квазіперіодичні й неперіодичні сегменти виконується як на окремих звукових файлах, так і на всій навчальній вибірці в автоматичному режимі. Експерту лише необхідно вказати найбільшу та найменшу допустимі довжини квазіперіодів, а також обмеження на приріст тривалості одноквазіперіодів.

Границі квазіперіодів у вікні дослідження звукового сигналу зображаються у вигляді вертикальних рисок з кольоровими квадратиками-ручками в нижній частині риски. Ручки призначені для пересування границь квазіперіодів (мікрофонем).

Як тільки проведено сегментацію навчальної вибірки, інформація про фонеми-трифони та їх структуру вноситься до реєстру фонем-трифонів, формуючи таким чином усномовний файл диктора. Для цього слід натиснути кнопку *Update Collection*, попередньо виділивши сигнал тих фонем-трифонів, які заносяться до реєстру. Якщо не виділено жоної фонеми-трифона, до бази даних заносяться всі фонеми-трифони з реалізації.

Знання про індивідуальні інтонаційні контури отримуються шляхом оброблення відсегментованих реалізацій фраз, в яких представлені всі типи інтонації.

Сформований усномовний файл диктора зберігається у вигляді файлу, який містить не сам сигнал, а лише посилання на фрагменти звукового файлу, що відповідають фонемам-трифонам. Таким чином, зберігається зв'язок з навчальною вибіркою, і зміни сегментації, проведені на навчальній вибірці, автоматично відображаються в базу даних шляхом виклику команди *Update Collection*.

4 Дослідження синтезованого сигналу

В рамках Студії дослідника усномовного сигналу розроблено стенди дослідження синтезу та розпізнавання усного мовлення [5]. З використанням першого стенду виконуються експериментальні дослідження розбірливості та якості звучання синтезованого усномовного сигналу. На другому стенді проводяться дослідження пофонемного розпізнавання.

Стенд дослідження усномовного синтезу являє собою вікно, представлене діалоговою панеллю з багатьма елементами регулювання та командними елементами, які логічно розбиті на частини, як це показано на Рис. 2.

Поле задання орфографічного або фонетичного тексту, що буде подано на озвучення, дозволяє маніпулювати зі зразками тексту або транскрипції, озвучений сигнал яких планується дослідити. Це поле складається з вікна вибору поточного озвучуваного тексту, текстового вікна та командних елементів – ґудзиків, що спонукають додати введений у текстове вікно текст до списку озвучуваних текстів або видалити непотрібний зразок тексту.

Вікно вибору представляє собою список зразків орфографічного або фонетичного тексту. Перші кільканадцять символів елементу списку допомагають дослідникові обрати орфографічний /фонетичний текст із уведених раніше.

Вибраний зі списку текст одразу ж відображається у текстовому вікні. Це дозволяє бачити текст цілком, редагувати його. Щоб додати новий зразок тексту, орфографічного або фонетичного, слід увести зразок тексту до текстового вікна або відредагувати текст попередньо доданого зразка і виконати команду додання нового зразка тексту шляхом активації відповідного командного ґудзика. Також передбачена можливість видалити зразок тексту.

Щойно додано новий зразок тексту або вибрано з уведених раніше зразків, поле виділення фонем-трифонів відображає фонемно-трифонні транскрипції, що відповідають поточному зразкові тексту. У поточній фонемно-трифонній транскрипції передбачено можливість виділяти окремі фонеми-трифони, які з метою досліджень можна замінювати на інші та/або задавати їм різні просодичні характеристики за допомогою описаних далі полів.

Поле вибору фонем-трифонів дозволяє "підсаджувати" різні фонеми-трифони з інди-відуального усномовного файлу диктора у фонемно-трифонну послідовність озвучуваного тексту.

Основний елемент поля вибору фонем-трифонів є таблиця фонем-трифонів, доступних у індивідуальному усномовному файлі диктора, що містить база даних і знань озвучення текстів. Таблиця показує лише фонеми-трифони-претенденти, тобто ті фонеми-трифони, чия термінальна фонема збігається з термінальною фонемою виділеної фонеми-трифона у полі фонемно-трифонної транскрипції.

У таблиці коротко характеризуються фонеми-трифони-претенденти, показано їх оригінальне фонемне оточення. Щоб замінити виділену фонему-трифон у поточній фонемно-трифонній послідовності озвучуваного тексту, необхідно вибрати фонему-трифон з таблиці та викликати команду *Assign*.

Поле регулювання просодичних характеристик сигналу дозволяє змінювати інтонацію, темп і гучність синтезованого сигналу. Це поле містить список всіх одноквазіперіодичних сегментів (мікросегментів) виділеної фонеми-трифона поточної фонемно-трифонної транскрипції. У сусідніх текстових віконечках відображаються значення довжини вибраного одноквазіперіоду та його гучності. Передбачена можливість розмножувати окремі квазіперіоди.

Шляхом зміни довжини квазіперіодів фонеми-трифона в допустимих межах відбувається зміна інтонаційного контуру фонеми на низькому рівні. Викидання окремих мікросегментів або їх множення призводить відповідно до скорочення або подовження фонеми-тифона вцілому, а отже до зміни його темпоральних характеристик.

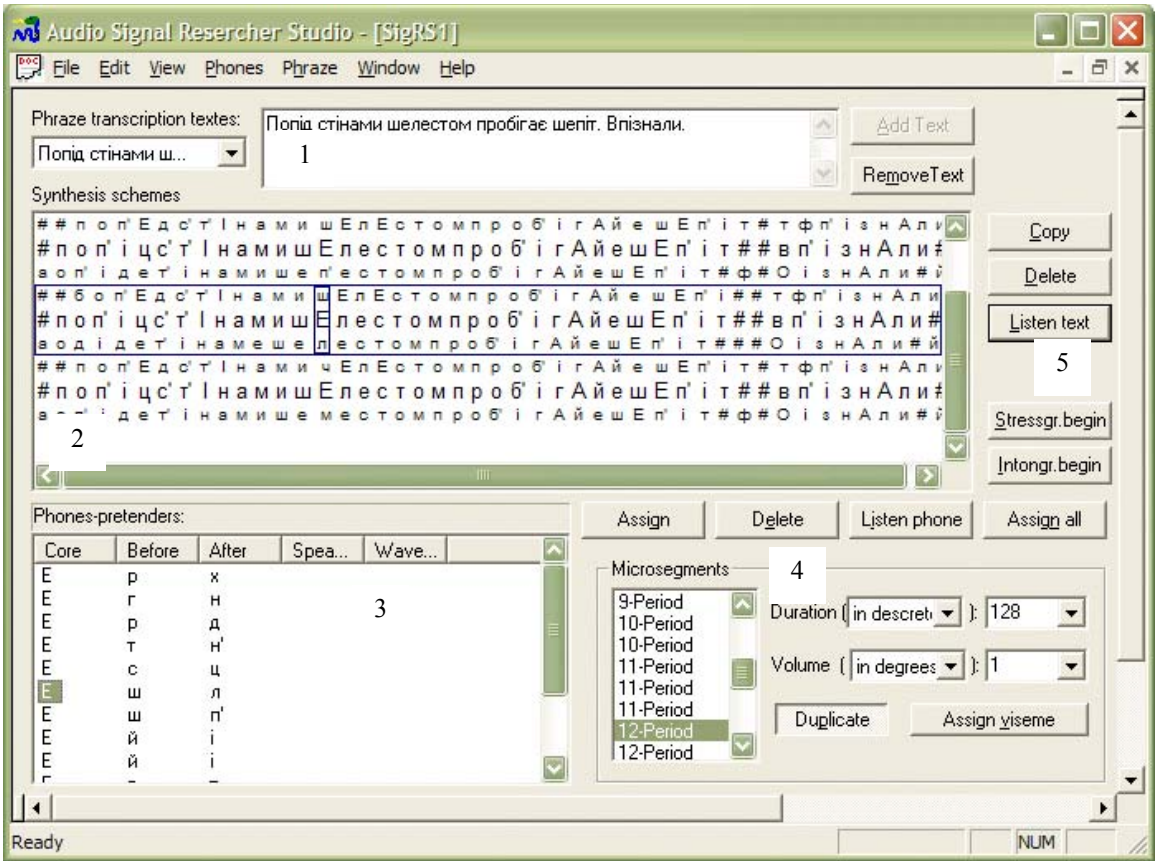

Рис. 2. Вікно дослідження усномовного синтезу представляє собою діалогову панель з елементами регулювання та командними елементами, які підрозділяються на такі розділи:

1) секція задання озвучуваного тексту, орфографічного або фонетичного;
2) інтерактивний список фонемно-трифонних транскрипцій заданого орфографічного (фонетичного) тексту;
3) секція вибору фонем-трифонів з бази даних і знань;
4) секція регулювання просодичних характеристик сигналу;
5) загальні командні елементи.

Передбачена також можливість зміни інтонаційних харктеристик на вищому рівні шляхом задання інтонаційних контурів синтезованого тексту. Відповідна команда у меню (*Phraze->Apply intonation*) викликає діалогове вікно, в якому за спеціальною схемою описуються інтонаційні контури ритмогрупи або синтагматичні. Після введення опису інтонаційних контурів і підтвердження операції, просодичні параметри керування синтезатором можуть бути обчислені за одним з уведених контурів.

Загальні командні елементи містяться як на стенді, окремо від описаних вище секцій, так і в меню.

Після заповнення належним чином полів стенду досліджень усномовного сигналу подається загальна команда озвучення тексту. На самому початку роботи зі стендом необхідно завантажити базу даних і знань для озвучення українських текстів загальною командою з меню. Одночасно можливо працювати як з декількома базами даних, так і з різними стендами дослідження синтезу усної мови.

Загальний порядок роботи зі стендом дослідника усномовного синтезу виглядає наступним чином. Щоб започаткувати нові дослідження, у Студії дослідника усномовного сигналу створюємо новий документ типу Стенд дослідника озвучення текстів (*Speech Synthesis Master*). Далі необхідно викликати файл бази даних і знань озвучення текстів (*Phones->Add speech DB*). Таким чином задається основна конфігурація досліджень: природня мова, фонетичні знання та усномовний файл диктора. При наступному сеансі роботи зі стендом у цій же конфігурації файл бази даних і знань озвучення текстів завантажується автоматично.

Новий озвучуваний текст, наприклад, фонетичний, вводиться у текстовому вікні поля задання озвучуваного тексту, а потім додається супутньою командою *Add*. При цьому автоматично відображається відповідна фонемно-трифонна транскрипція у головному вікні поля виділення фонемно-трифонних транскрипцій.

Далі підсаджуємо ті фонеми-трифони з усномовного файлу диктора, які заплановано дослідити. Для цього за допомогою мишки виділяємо потрібну фонему-трифон зі списку досліджуваних фонемно-трифонних послідовностей, обираємо фонему-трифон, яку заплановано “підсадити”, зі списку можливих фонем-трифонів в усномовному файлі диктора, і закріплюємо цю заміну командою *Assign*.

Зміни інтонації проводимо як на вищому рівні шляхом задання інтонаційних контурів, так і на нижчому, вручну змінюючи довжини квазіперіодів. Останнє здійснюється шляхом зміни значень у віконечку довжини виділеного квазіперіоду. Щойно точка введення перейде до іншого елемента стенду (мишкою виділяємо наступний квазіперіод), змінені значення довжини квазіперіоду запам’ятаються і будуть використані при наступній процедурі синтезу сигналу.

Темпоральні зміни прототипу на низькому рівні здійснюються за допомогою маніпуляцій з мікросегментами фонем-трифонів. Множення виділених мікросегментів за допомогою команди *Duplicate* призводить до подовження окремої фонеми-трифона, а отже і до загального сповільнення темпу. Викидання мікросегментів, тобто задання їм нульової довжини, призводить до скорочення довжини фонеми-трифона, а тому – до прискорення темпу.

Нарешті, командою *Listen text* запускаємо процедуру синтезу, яка моделює оригінальний алгоритм синтезу усномовного сигналу в часово-амплітудній області. При цьому відкривається новий документ типу *Усномовний сигнал*, куди автоматично вставляється синтезований сигнал, який тепер є доступним для прослуховування, аналізу та подальшого дослідження (Рис. 1).

При прослуховуванні експертами синтезованих як окремих слів, так і злитого мовлення, було з’ясовано словесну розбірливість синтезованого сигналу, що виявилася на рівні не менше 90% на перших сенсах, зростаючи за мірою “звикання” до синтезованого голосу.

5 Висновки

Запропонований метод синтезу дозволяє озвучувати українські тексти з доволі прийнятною розбірливістю, натуральністю синтезованого сигналу та збереженням індивідуальності мовця.

Обсяг акустичних данних, що використовуються при синтезі, у розгорнутому вигляді становить від 5 МБ і вище для одного диктора. Обсяг лінгвістичної інформації, що використовується для розставлення наголосів в українських словах становить 4 МБ. Такі обсяги даних, а також вимоги до швидкодії є прийнятними для застосування в реальних комп’ютерних системах та в портативних пристроях на сучасній електронній базі.

Чекають розв’язку проблеми неоднозначності наголосів та точної відповідності типів інтонації.

Література

1. Т.К. Винцюк. *Анализ, распознавание и смысловая интерпретация речевых сигналов*. — Киев: Наукова думка, 1987.
2. Taras K. Vintsiuk, Mykola M. Sazhok: *Speaker Voice Passport for a Spoken Dialogue System*. — Proceedings of the 3rd International Workshop "Speech and Computer" - SPECOM'98, St.-Petersburg, 1998.
3. Taras K. Vintsiuk, Tetiana V. Liudovyk, Mykola M. Sazhok. — *Phonetic Knowledge Base for Ukrainian*.— Proceedings of the 3rd International Workshop "Speech and Computer" - SPECOM'98, St.-Petersburg, 1998.
4. T. Dutoit, H. Leich. — A Comparison of Four Candidate Algorithms in the Context of High Quality Text to Speech Synthesis. — ICASSP'94.
5. Микола Сажок. *Комп’ютерні засоби експериментальних досліджень усномовного сигналу*. — Праці 4-ї Всеукраїнської міжнародної конференції “УкрОбраз-98”, Київ, 1998.