

## МОДЕЛЮВАННЯ ПРОСОДИЧНИХ ХАРАКТЕРИСТИК ДЛЯ СИНТЕЗАТОРА УКРАЇНСЬКОГО МОВЛЕННЯ

Синтезатор українського мовлення, розроблений у відділі розпізнавання та синтезу звукових образів МННЦ ІТіС, призначений для озвучення довільних орфографічних текстів українською мовою. Він має засоби генерування просодичних, тобто тональних (інтонаційних) та темпоральних (ритмічних) характеристик слів, речень та текстів, що синтезуються. В цій статті розглядаються питання вибору просодичних моделей та алгоритми обчислення просодичних характеристик в рамках цих моделей.

### Вступ

Просодика (інтонація, наголос, ритм, темп, гучність та паузація) є важливим засобом формування висловлювань та виявлення їх смислу. Тому будь-який синтезатор мовлення за довільним текстом повинен мати засоби генерування просодичних характеристик.

Синтезатор українського мовлення, що розглядається у цій статті, складається з чотирьох головних модулів:

- лінгвістичного процесора;
- усномовних баз даних;
- модуля вибору оптимальних елементів з бази даних;
- акустичного процесора.

Лінгвістичний процесор, в свою чергу, складається з

- блоку нормалізації тексту;
- акцентно-інтонаційного процесора;
- фонетичного транскриптора;
- блоку обчислення просодичних характеристик.

В статті розглядаються питання, що стосуються моделювання просодики, головним чином, інтонації та ритміки як найважливіших її компонентів, що суттєво впливають на натуральність синтезованого

мовлення. Розглядаються моделі ритміки та інтонації та обґрунтовується підхід, що базується на використанні просодичної інформації, яку містять розроблені усномовні бази даних [1].

Дослідження різних баз даних дозволяє провести розподіл просодичних закономірностей на універсальні (загальні для українського мовлення) та індивідуальні, властиві окремим носіям української мови та окремим стилям вимовляння. Універсальні закономірності покладено в основу просодичних моделей, а індивідуальні служать змінними даними, що використовуються під час застосування універсальних правил.

Далі описуються алгоритми просодичного оформлення висловлювань, а саме:

а) обчислення тривалостей алофонів (реалізацій фонем у потоці мовлення), що зумовлює ритм (чергування наголошених та ненаголошених голосних та складів), темп та паузацію (пауза розглядається як окрема фонема);

б) обчислення інтонаційних (мелодичних, тональних) контурів, або контурів частоти основного тону (ЧОТ).

Наприкінці надається приклад фонетико-просодичної транскрипції тексту, яка потрапляє у модуль акустичного процесора для озвучення.

## **1 Просодичні моделі**

В синтезаторах мовлення складність просодичних моделей, кількість та зміст просодичних правил залежать як від практичних можливостей лінгвістичного процесора, так і від того, що розуміється під просодичною структурою тексту. Суттєву роль також відіграє доступний для аналізу фонетико-акустичний матеріал.

З погляду просодики текст складається з інтонаційних фраз (груп). Інтонаційна фраза (ІФ) — це речення або частина речення, визначена

інтонаційно, тобто має фразовий наголос та один з інтонаційних контурів (типів), властивих українському мовленню.

Кожна ІФ складається з однієї або більше акцентних груп. Акцентна група — це слово або кілька слів, що об'єднані словесним наголосом тонально-темпоральної природи, тобто наголошений склад вирізняється як більшою тривалістю, так і наявністю руху тону.

Акцентна група складається з фонетичних слів. Фонетичне слово — це одно або кілька орфографічних слів, що об'єднані словесним наголосом лише темпоральної природи, тобто наголошений склад вирізняється лише більшою тривалістю, а зміни тону на ньому немає.

Ритм та темп вимовляння відображаються у тривалості окремих сегментних одиниць — алофонів фонем. У потоці мовлення на тривалість алофонів впливають різні контекстні та позиційні фактори, а також тип інтонаційної фрази.

Гучність, впливаючи на якість синтезованого мовлення, не відіграє такої важливої ролі, як інтонація та ритміка. Як правило, на початку розповідних ІФ гучність більша, наприкінці — менша. У питальних ІФ максимум гучності припадає на питальне слово.

В найпростіших випадках моделюються лише членування на інтонаційні фрази, тривалості сегментних одиниць та три типи інтонації: завершеність (відповідає крапці у тексті), незавершеність (відповідає комі), та питання (відповідає знаку питання). Сегментна якість синтезованого мовлення у цьому випадку може бути дуже доброю, але “бідна” просодика значно псує натуральність звучання.

Складні і досить ефективні просодичні моделі використовуються у синтезаторах російського мовлення [2, 3]. До цього ж рівня відносяться і моделі, що використовуються у нашому синтезаторі українського мовлення [4]. Головна відмінність полягає у тому, що розроблені

просодичні моделі українського мовлення враховують лише ті явища та закономірності, які можна досліджувати за існуючими фонетико-акустичними базами даних різних дикторів. Внаслідок є можливість моделювати ритмічні та інтонаційні особливості вимовляння різних людей.

У фонетико-акустичних базах даних відсутні відомості про членування на слова, однак вони містять достатньо просодичної інформації: для кожного елемента (фонем-трифона, тобто алофона фонемі із вказівкою на його ліве та праве оточення) відомо його загальну тривалість (для усіх фонем-трифонів) та кількість і тривалість усіх квазіперіодів основного тону (для голосних та дзвінких приголосних). Тривалість послідовних квазіперіодів основного тону є ні чим іншим як тональним контуром сегментної одиниці. Для моделювання ритміки використовується також середнє значення тривалості квазіперіодів основного тону фонем-трифона, оскільки воно корелює із загальною тривалістю фонем-трифона. За цим показником можна визначити тип ІФ: грубо кажучи, у тональному центрі розповідного типу квазіперіоди довші, а питального — коротші. Місця пауз у базі даних свідчать про межі інтонаційних фраз.

Розглянемо окремо моделі ритміки та інтонації.

## **2 Модель ритміки**

Моделювання ритміки використовує базові (середні) тривалості алофонів та правила їх зміни (подовження або скорочення відносно базової) в залежності від багатьох факторів. Правила темпорального оформлення сформульовано окремо для голосних та приголосних.

Для усіх голосних факторами впливу на тривалість є:

- відкритість або закритість складу;
- тип інтонаційної фрази.

Окремо враховуються:

А) для наголошених голосних:

- позиція в інтонаційній фразі (чи остання голосна фрази);
- фразовий наголос.

Б) для ненаголошених голосних:

- позиція по відношенню до початку ІФ;
- позиція по відношенню до кінця ІФ.

Фактори, що подовжують тривалість приголосних:

- позиція абсолютного початку або кінця інтонаційної фрази;
- позиція перед наголошеним голосним.

Фактори, що скорочують тривалість приголосних:

- позиція перед іншим приголосним;
- позиція перед ненаголошеним голосним.

Вважається, що короткі та довгі приголосні розрізняються лише тривалістю.

### **3 Модель інтонації**

Розрізняються 4 типи акцентних груп (АГ) у складі ІФ: початкова, перед'ядерна, ядерна та позаядерна.

Кожна ІФ обов'язково має ядерну АГ, тонально-темпоральний наголос якої співпадає з фразовим. Наголошений голосний ядерної АГ є тональним (мелодичним) центром ІФ. В більшості випадків ядерною є остання АГ в ІФ. Однак в деяких випадках тональний центр ІФ зміщується, наприклад, ним стає питальне слово у питальній ІФ.

Визначення тонального центру та типу АГ є складним завданням, оскільки у тексті часто не буває ніяких ключів, які б дозволили зробити це автоматично.

Залежно від місця ядерної АГ в ІФ інші акцентні групи цієї фрази розглядаються як початкова, перед'ядерна або позаядерні. Розглянемо речення, що складається з двох інтонаційних фраз (усі приклади далі наведено з реальних усномовних баз даних):

Сьогодні ж Софія | продовжує занепадати.

Якщо в ІФ («Сьогодні ж Софія») перша АГ є ядерною («Сьогодні ж»), решта є позаядерними («Софія»). Якщо в ІФ ядерною є друга АГ («продовжує занепадати»), то перша АГ є початковою.

Теоретично, якщо уявити довгу ІФ, то будемо мати:

ПОЧАТКОВА ПЕРЕД'ЯДЕРНА ... ПЕРЕД'ЯДЕРНА **ЯДЕРНА** ПОЗАЯДЕРНА ... ПОЗАЯДЕРНА

З іншого боку, довжина ІФ є обмеженою і в загальному випадку зумовлюється темпом та стилем вимовляння.

Моделювання контурів ЧОТ інтонаційних фраз базується на використанні моделі, яку можна віднести до класу лінійних інтонаційних моделей [2]. Мається на увазі, що інтонаційний контур складається з послідовності тональних цільових точок (мішеней), що мають бути досягнені у певних місцях ІФ. Цільові тональні точки з'єднуються лінійними переходами. Вважається, що кількість точок невелике. Тональні зміни можуть бути плавними або різкими.

Інтонаційні контури відрізняються діапазоном ЧОТ (вузкий, середній та широкий), а також напрямком руху (висхідний, спадний тощо). Інтонаційні фрази, як правило, відокремлюються одна від одної паузами різної тривалості в залежності від типу ІФ.

Як і в [2, 3], синтез інтонації орієнтовано на декілька базових інтонаційних контурів (ІК), що в загальному випадку співвідносяться із прийнятим поділом речень на розповідні, питальні, спонукальні, окличні тощо, тобто за загальним комунікативним призначенням. Однак між типами контурів та комунікативними типами немає взаємно-однозначної відповідності. Наприклад, переліку можуть відповідати три різні ІК, які

окрім переліку “обслуговують” також розповідальні та питальні висловлювання.

В експериментальній версії синтезатора, що дозволяє інтонувати тексти вручну, використовуються 17 типів ІК, які для зручності позначено як різновиди інтонаційних конструкцій О.А. Бризгунової [5]:

ІК-1а, ІК-1б, ІК-1в, ... ІК-6б. Надто умовно можна казати, що контури ІК-1 задають розповідну інтонацію, ІК-2 — пояснювальну, ІК-3 та ІК-4 — питальну, ІК-5 — окличну, ІК-6 — інтонацію незавершеності (“Сьогодні ж Софія...”).

У версії синтезатора, що аналізує та озвучує тексти цілком автоматично, використовується лише 8 ІК, 7 з яких однозначно пов’язані із знаками пунктуації (. , : ; - ? !), а останній (“|”) — із незавершеністю, що не позначена пунктуаційно. Цей ІК з’являється при автоматичному членуванні довгих (довших ніж 6 фонетичних слів) ІФ. Наприклад,  
І все-таки вже тоді | на мене повіяло холодним  
мороком невідомості.

Інтонаційний контур ІФ складається з контурів акцентних груп. Контур акцентної групи, в свою чергу, складається з трьох частин [3]: перед’ядро, ядро (тональний центр) та позаядро. Контури перед’ядра та позаядра задаються двома точками: початок та кінець, а контур ядра, яке співпадає із наголошеним (виділеним тонально-темпорально) голосним, моделюється детальніше — за допомогою 6 цільових точок.

За наявності двох дикторів, 8 типів ІФ та 4 типів АГ моделюються 64 інтонаційні контури АГ, а конкретна цільова тональна точка задається, таким чином, як елемент чотиримірної масиви:

ІК[диктор][тип ІФ][тип АГ][номер цільової тональної точки]

#### 4 Просодичні дані

В ідеалі, просодичні характеристики фрази, що синтезується, повинні максимально відповідати тому, що спостерігається за природнього вимовляння цієї ж фрази людиною. Незважаючи на певні спільні закономірності у просодичному оформленні фрази, мовлення різних людей різняться за ритмікою, темпом та паузацією, що відображається у тривалостях фонем. Слід враховувати також, що різні люди надають перевагу різним інтонаційним засобам при вимовлянні однієї й тієї ж фрази, а в тих випадках, коли різними людьми використовується один і той самий інтонаційний контур, значення ЧОТ все одно різняться за рахунок різних діапазонів голосу. Тому аналізується просодична інформація, що міститься у існуючих усномовних базах даних. Це робиться частково автоматично за допомогою розроблених програмних засобів.

Отримані дані використовуються при обчисленні просодичних характеристик в рамках моделей ритміки та інтонації.

Наведемо деякі дані дикторів С. та Д.

Тривалість наголошених голосних диктора С. в середньому коливається від 40 до 205 мс, ненаголошених — від 25 до 170 мс. У диктора Д. дані такі: наголошені — від 42 мс до 136 мс, ненаголошені — від 28 мс до 108 мс.

Найтривалішим як серед наголошених, так і серед ненаголошених голосних у обох дикторів є [а] (відповідно, середня тривалість 117 та 85 мс у диктора С. та 93 і 54 у диктора Д.). Найкоротшим у диктора С. є голосний [и] (відповідно, середня тривалість 87 та 68 мс). У диктора Д. найкоротший наголошений голосний — [И] (68 мс), найкоротші ненаголошені — [е] та [и] (49 мс).

Середні тривалості алофонів вважаються базовими для подальших обчислень.



## 5 Формування просодичних характеристик у синтезаторі мовлення

Головним завданням лінгвістичного процесора є перетворення вхідного орфографічного тексту на фонетико-просодичну транскрипцію, яка містить як сегментну (послідовність алофонів), так і просодичну (тривалості алофонів, значення частоти основного тону) інформацію, необхідну для подальшої генерації акустичного сигналу.

Виділяються такі етапи формування просодичних характеристик:

- визначення наголосів у словах;
- інтонаційно-фразове членування;
- обчислення тривалостей алофонів;
- обчислення значень частоти основного тону.

Розглянемо, як це відбувається в існуючій версії синтезатора мовлення.

Визначення наголосів та членування довгих ІФ робиться акцентно-інтонаційним транскриптором ще в орфографічному тексті на підставі словника наголосів [6] та інших службових словників.

Далі орфографічний текст перетворюється у фонетичну транскрипцію, в якій зберігаються знаки пунктуації та наголосу, а також межі між фонетичними словами. Саме ця інформація суттєва для визначення ритміки та інтонації на наступному етапі, у блоці обчислення просодичних характеристик (БОПХ). Знаки наголосу можуть бути трьох типів: словесний (звичайний), фразовий та допоміжний [7].

Синтезатор орієнтується на стандартний для диктора темп вимовляння. Однак є можливість змінювати темп.

БОПХ працює останнім у лінгвістичному процесорі. В задачу БОПХ входить визначення темпоральних та тональних характеристик, що мають бути надані алофонам, з послідовності яких формується фонетична

транскрипція тексту, що синтезується. На виході блоку фонетична транскрипція має бути перетворена на фонетико-просодичну.

Просодичні характеристики алофонів обчислюються у відповідності до фонетичних характеристик цих алофонів, їх контексту, типу та місця тонального центру інтонаційної фрази, темпу вимовляння.

Просодичні характеристики задають значення фізичних параметрів алофонів: тривалість (в мс) та рух основного тону (послідовність значень ЧОТ) (у Гц).

Блок оперує універсальними та індивідуальними даними.

Алфавіт фонем української мови є спільним для усіх носіїв. Наголошені/ненаголошені голосні вважаються різними фонемами, таким чином, вже в алфавіті фонем міститься просодична інформація. Наголошені голосні позначаються великими літерами, ненаголошені — маленькими: А,а,О,о,Е,е,І,і,И,и,У,у. Усі приголосні крім [й] становлять пари твердий/м'який: б-б', в-в' і так далі. До алфавіту включено також фонему-паузу #. Усього алфавіт фонем синтезатора налічує 58 одиниць.

Оскільки в алфавіті фонем, і відповідно, у базах даних не розрізняються короткі та довгі приголосні, вважається, що короткий та довгий приголосні — це один алофон, але різної тривалості.

## **6 Алгоритм членування інтонаційних фраз на акцентні групи та визначення параметрів акцентних груп**

- Визначається кількість АГ та місце ядерної АГ. Кількість АГ визначається за кількістю знаків наголосу (враховуються лише словесні та фразовий наголоси). Якщо в ІФ знайдено знак фразового наголосу, то ядерною вважається АГ, що містить фонетичне слово, позначене цим наголосом, в іншому випадку — остання АГ в ІФ.

- Визначаються межі акцентних груп у випадку, коли їх більше одної. Вважається, що АГ починається з фонетичного слова, яке містить словесний або фразовий наголос. Усі ФС, що мають знак допоміжного наголосу, відносяться до попередньої АГ. Якщо такі ФС стоять на початку ІФ, то вони відносяться до першої АГ.
- Визначається тип АГ: перша АГ — початкова, усі, якщо вони є, між початковою та ядерною, — перед'ядерні, ті, що після ядерної — позаядерні.
- Для кожної АГ знаходиться опис ІК (послідовність 10 цільових тональних точок) в залежності від диктора та від типу АГ.
- Для кожної АГ визначається місце тонального центру (ТЦ) (голосний, позначений словесним або фразовим наголосом) та перша фонема наступної АГ (у випадку останньої АГ ІФ це буде пауза).

## **7 Алгоритм обчислення темпоральних характеристик алофонів фонем**

Обчислення тривалостей алофонів фонем за правилами ритміки виходить з того, що кожна фонема має базову тривалість та набір коефіцієнтів тривалості, що змінюють цю базову тривалість. Набори коефіцієнтів тривалості в кожного з дикторів свої. Вони можуть також бути різними й в одного диктора, якщо йдеться про різні стилі вимовляння (наприклад, читання художньої прози або новин у телепередачах).

Окремо для голосних та приголосних обчислюється вектор контексту, кожна з 4 компонент якого може приймати значення 1 або 0.

Для голосних вектор контексту обчислюється таким чином:

- компонента 1: 1, якщо голосний є останнім у складі (відкритий склад);  
0 в іншому випадку;
- компонента 2: 1, якщо рух основного тону є спадним;  
0 в іншому випадку;

компоненти 3 та 4 обчислюються окремо для наголошених та ненаголошених голосних.

Для наголошених голосних:

- компонента 3: 1, якщо голосний є останнім голосним ІФ;  
0 в іншому випадку;
- компонента 4: 1, якщо голосний має фразовий наголос;  
0 в іншому випадку;

Для ненаголошених голосних:

- компонента 3: 1, якщо голосний є першим голосним ІФ;  
0 в іншому випадку;
- компонента 4: 1, якщо голосний є останнім алофоном ІФ;  
0 в іншому випадку;

Для приголосних вектор контексту обчислюється таким чином:

- компонента 1: 1, якщо наступний алофон приголосний;  
0 в іншому випадку;
- компонента 2: 1, якщо обидва оточуючі алофони голосні;  
0 в іншому випадку;
- компонента 3: 1, якщо приголосний стоїть або після, або перед паузою;  
0 в іншому випадку;
- компонента 4: 1, якщо наступний алофон — наголошений голосний;  
0 в іншому випадку;

16 комбінацій, що їх задає вектор контексту, становлять набір коефіцієнтів тривалості. Значення цих 16 коефіцієнтів для кожної фонемі і для кожного диктора вираховано заздалегідь за допомогою автоматичного аналізу фонетико-акустичних баз даних. Під час синтезу ІФ за контекстом визначається, який з 16 коефіцієнтів треба вибрати.

Далі базова тривалість помножується на коефіцієнт тривалості та коефіцієнт темпу. Існують обмеження на тривалість алофонів: вона повинна знаходитися в діапазоні 20–250 мс.

Тривалість довгих приголосних помножується ще на коефіцієнт подовженості, що залежить як від окремих приголосних, так і від диктора.

## **8 Алгоритм обчислення тональних характеристик висловлювань**

Діапазон ЧОТ, нижня та верхня межі ЧОТ залежать від диктора. Так, для чоловічих голосів діапазон ЧОТ в синтезаторі становить від 110–120 до 185–195 Гц, жіночих — від 140 до 205 Гц.

Для обчислення значень ЧОТ використовуються відносні значення, що потім співвідносяться із конкретними даними дикторів. Діапазон основного тону має градацію від 1 до 16 значень. Значення 1 відповідає мінімальному значенню ЧОТ, 16 — максимальному.

За типом ІФ, визначеним акцентно-інтонаційним транскриптором, та типом АГ з бази даних диктора вибирається відповідний інтонаційний контур АГ — вектор з 10 цільових тональних точок. З них перші дві задають значення ЧОТ на перед'ядрі, середні 6 — на ядрі, а останні дві — на позаядрі АГ.

Для кожної АГ визначається кількість неглухих (вокальних) алофонів перед ТЦ та кількість неглухих фонем після ТЦ. До цієї групи відносяться дзвінки приголосні та голосні із допоміжним наголосом. Для глухих приголосних та паузи ЧОТ не обчислюється.

Для кожного неглухого алофона значення ЧОТ обчислюються із урахуванням його місця в АГ. Для перед'ядерних алофонів:

- якщо перед'ядерний алофон один, то початкове значення ЧОТ на ньому буде дорівнювати першому з пари, що описує перед'ядро;

- якщо перед'ядерних алофонів два, то початкове значення першого буде дорівнювати першому з пари, а початкове значення другого — другому;
- якщо перед'ядерних алофонів більше, то початкове значення ЧОТ поточного алофона обчислюється за формулою

$$\text{ЧОТ} = \text{ІК}[0] + ((\text{ІК}[1] - \text{ІК}[0]) \times I) / K,$$

де  $I$  — це порядковий номер поточного перед'ядерного неглухого, а  $K$  — кількість перед'ядерних не глухих.

Аналогічно обчислюються значення ЧОТ для позаядерних алофонів.

Значення ЧОТ для наголошеного голосного, що є тональним центром АГ, в залежності від тривалості дорівнюють компонентам 3–8 інтонаційного контура АГ.

Таким чином отримуються значення ЧОТ на початку усіх алофонів та додатково в середині алофонів, що є тональними центрами АГ.

На рис. 1 показано, як різняться контури ЧОТ різних дикторів при синтезі однієї й тієї ж ІФ.

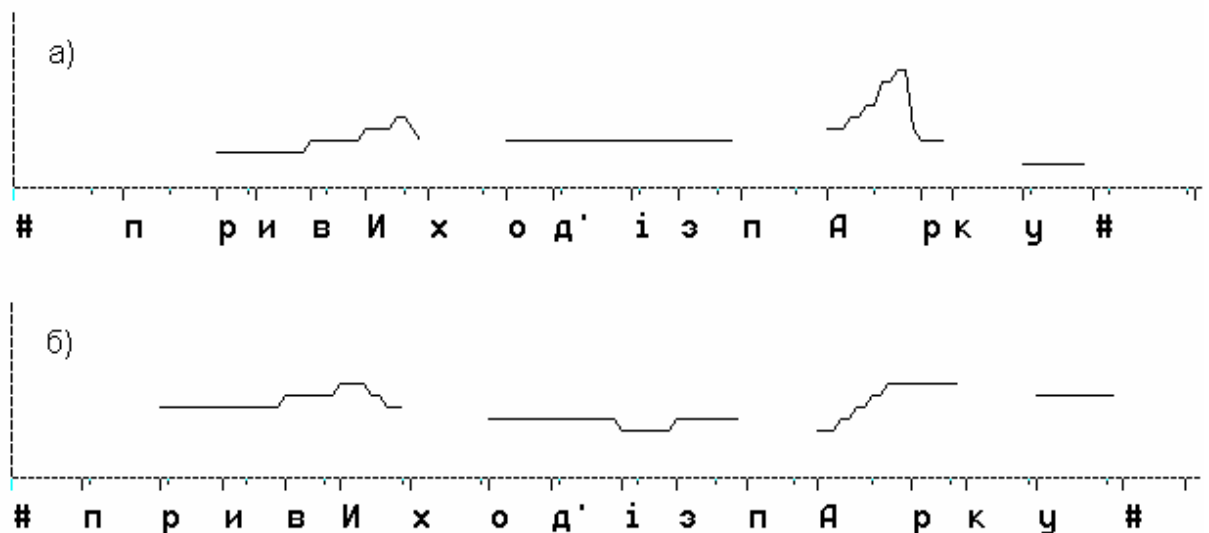


Рис. 1. Контури ЧОТ інтонаційної фрази “При виході з парку...” (синтезоване мовлення, а) диктор С., б) диктор Д.)

На Рис.2 наведено контур ЧОТ реального вимовляння цієї ІФ диктором С.

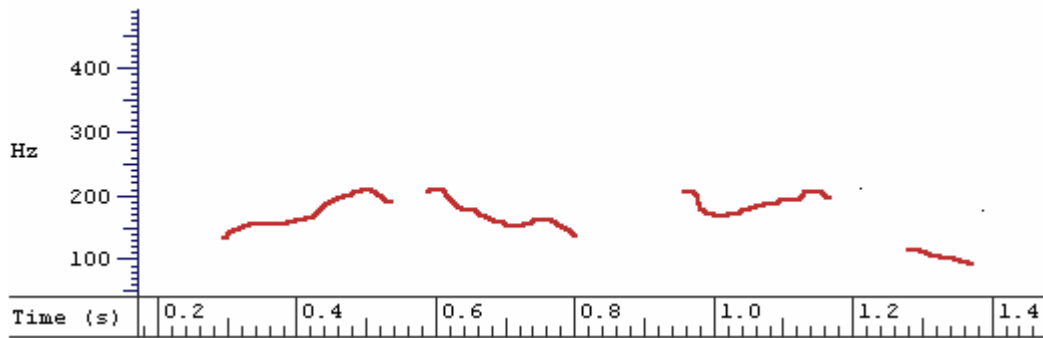


Рис.2. Контур ЧОТ фрази “При виході з парку...” (реальне мовлення, диктор С.)

## 9 Генерація фонетико-просодичної транскрипції

На виході лінгвістичного блоку кожна ІФ представлена у вигляді фонетико-просодичної транскрипції. В ній фіксуються одночасно як звуковий (алофонний) склад ІФ, так і її просодичні характеристики — тривалість та частота основного тону.

Нижче наведено фонетико-просодичну транскрипцію, що моделює вимовляння диктора С. Вона містить інформацію про тривалість звучання алофонів або їх частин (в мс), а також (для неглухих алофонів) значення ЧОТ на початку тривалості алофонів або їх частин.

```
# 140 мс
п 120 мс
р (50 мс 125 Гц)
и (70 мс 125 Гц)
в (70 мс 130 Гц)
И (40 мс 135 Гц) (20 мс 140 Гц) (10 мс 135 Гц) (10 мс 130 Гц)
х 100 мс
о (60 мс 130 Гц)
д' (100 мс 130 Гц)
і (60 мс 130 Гц)
з (80 мс 130 Гц)
п 110 мс
А (30 мс 135 Гц) (20 мс 140 Гц) (20 мс 145 Гц) (20 мс 155 Гц) (20 мс 160 Гц) (10 мс 135 Гц)
р (40 мс 130 Гц)
к 90 мс
у (90 мс 120 Гц)
# 130 мс
```

Фонетико-просодичні транскрипції використовуються надалі для обчислення довжин квазіперіодів ЧОТ на протязі тривалості алофонів, а потім для вибору оптимального алофону (фонемі-трифону) з фонетико-акустичної бази даних.

## 10 Прикінцеві положення

Синтезоване мовлення, що генерується з використанням наведених моделей та алгоритмів, значно відрізняється натуральністю від монотонного вимовляння. Це підтверджує загальну думку [8], що просодичні характеристики в значній мірі несуть відповідальність за натуральність звучання синтезованого мовлення.

Станом на сьогодні синтезатор українського мовлення здатен озвучувати орфографічні тексти трьома чоловічими голосами. Експеримент із синтезуванням жіночого мовлення засвідчив, що наведені моделі та алгоритми не потребують змін і дають гарні результати. Необхідність полягає лише в створенні нової фонетико-акустичної бази даних та отриманні на її підставі індивідуальних просодичних даних.

З метою покращення просодики синтезованого мовлення найближчим часом планується до моделей ритміки та інтонації (основних) додати модель гучності.

### Література

1. М. Сажок. *Автоматизовані засоби дослідження синтезу українського мовлення на основі фонемно-трифонної моделі* // Автоматизовані системи управління і нові інформаційні технології, Київ, 2003, вип.1, с. 101–113.
2. O.F. Krivnova. *Automatic synthesis of Russian speech* // *Proceedings of the XIV International Congress of Phonetic Sciences, Vol.1*, pp. 507–510, San Francisco, 1999.
3. B.M. Lobanov. *The Phonemophone text-to-speech system* // *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn, 1987, Vol.1*, pp.120–124.
4. Т. К. Vintsiuk, Т. V. Liudovyk, М. М. Sazhok. *Phonetic Knowledge Base for Ukrainian*// *Proceedings of the 3rd International Workshop “Speech and Computer” — SPECOM'98, St.-Petersburg, 1998*, с. 179–182.
5. Е.А. Брызгунова. *Звуки и интонация русской речи*, М., 1983.
6. В.А. Широков, В.В. Манако. *Організація ресурсів національної словникової бази* // *Мовознавство*, 2001, № 5, с. 3–13.
7. Т. Liudovyk. *Pitch Contour Modelling in a Speech Synthesis System* // *Proceedings of the First All-Ukrainian Conference “Signal/Image Processing and Pattern Recognition”*, Kyiv, 1992, с. 135–136.
8. В. Möbius. *Corpus-based speech synthesis: methods and challenges* // *Arbeitspapiere des Institut fur Maschinelle Sprachverarbeitung, AIMS 6(4)*, pp. 87–116, Univ. Stuttgart, 2000.