

Інформаційні технології

УДК 004.424

Лєсна Наталя Совєтївна

*кандидат технічних наук, професор,
професор кафедри програмної інженерії
Харківський національний університет радіоелектроніки*

Стьопін Олександр Сергійович

*студент
Харківський національний університет радіоелектроніки*

МЕТОДИ ТА МОДЕЛІ СИНТЕЗУ ІНТОНАЦІЙНОЇ ЗВУКОВОЇ МОВИ

***Анотація.** У статті розглянуті питання які зв'язані з синтезом інтонаційної звукової мови. Зокрема, розглядається роль фонетично-акустичного процесору а саме про методи модифікації довжини алофонів. При цьому етапі відбувається виправлення темпо-ритмічних і частотних артефактів. Такі виправлення здійснюються за допомогою модифікації основних компонентів алгоритму LP-PSOLA.*

***Ключові слова:** синтез мови, алофон, синтезування інтонації.*

До теперішнього часу системи синтезу мови досягли певної досконалості і вже використовуються в ряді практичних додатків. Однак комфортність сприйняття синтезованої мови в реальних умовах, особливо при довгих повідомленнях, залишається ще далекою від задовільної. Світова тенденція розвитку мовленнєвих технологій вказує на актуальність створення систем синтезу виразної інтонації в синтезованій мові. На даний момент кількість досліджень в області створення систем синтезу

виразної мови постійно збільшується. Дослідження в області створення систем синтезу інтонаційної мови проводяться для різних мов.

Поняття «інтонація» в лінгвістиці сформувалося як міждисциплінарне поняття однієї з функцій усного мовлення людини. Одним з головних компонентів звукової реалізації інтонаційного мовлення є просодія, тобто ритміка, динаміка мови, та мелодика. Ці елементи взаємодіють, підтримують один одного і всі разом обумовлюють інтонації у мові. Сама інтонація у синтезованому тексті формується на етапі проходження фонетико-акустичного процесору, з таких ключових елементів як мелодика, гучність, темп мови та її окремих відрізків, ритміки, та особливостей фонації повідомлення [1, с. 46].

Одним з аспектів формування інтонації у синтезованій мові є регулювання темпу мови, тобто модифікація тривалості. Зміни основного тону які у нашому випадку здійснюються за допомогою дискретного перетворення Фур'є, призводять до зміни довжини алофона, звук якого піддається модифікації. Це обумовлює втрату природності, диктор починає говорити то швидше, то повільніше. У таких випадках також необхідно виправляти тривалість алофонів. Підвищення основного тону періодичного сигналу за допомогою дискретного перетворення Фур'є зменшує його тривалість. Для його відновлення зазвичай використовується повтор періодів сигналу. При цьому необхідно уникнути виникнення двох основних дефектів, що знижують якість синтезованого сигналу. Перший пов'язаний з тим, що при кожному повторі збивається фаза сигналу, що призводить до характерного потріскування при відтворенні сигналу. Другий пов'язаний з тим, що багаторазове повторення одного періоду людське вухо сприймає як гудіння або дзвін.

Одним з стандартних рішень цієї проблеми є наступна схема повтору періодів. На сам перед нам необхідно розділити алофон на періоди. Далі для збільшення алофона необхідно з двох суміжних періодів

які мають довжину l_1 та l_k отримати $k \leq 2$ періодів з довжинами l_1, \dots, l_k [2, с. 20-45]. При цьому перший та останній періоди є основними. Періоди $2, \dots, k - 1$ синтезуються з першого та останнього k періоду ця довжина вираховується за формулою.

$$l_j = l_1 + \left(\frac{l_k - l_1}{k - 1} \right), j = 2, \dots, k - 1$$

Перед модифікацією кожного відрізка проводиться розрахунок кількості доповнених періодів. Якщо, $L_m, n = 0, \dots, N$ - масив значень першого періоду та коректної стиковки, а $R_p, p = 0, \dots, P$ - масив значень другого періоду та некоректної стиковки [3 с. 2]. Якщо масив значень нового періоду, який необхідно вставити між першим та другим, визначається суперпозицією:

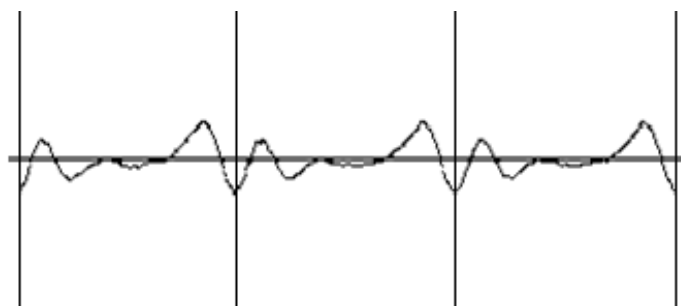


Рис. 1. Періоди алофона

На рисунку [1, с. 3] ми бачимо періоди після подовження алофона. Доданий період ми можемо побачити в центрі. Отриманий в результаті період ідеально стикується по фазі як з першим так і з другим своїм сусідом, при цьому не є повтором ні того, ні іншого, що і є необхідним нам рішенням дефектів.

$$M_k = TR_k + TL_k, k = 0 \dots K, \begin{cases} TL_k = 0, k = 0..G \\ TL_k = L_{k-G} W_{k-G}^1, k = G..K' \end{cases} \begin{cases} TR_k = 0, k = G..K \\ TR_k = R_k W_k^2, k = 0..F' \end{cases}$$

$$W_n^1 = \frac{1}{2} \left(1 - \cos \left(\frac{\pi n}{S} \right) \right), n = 0 \dots S, W_n^2 = \frac{1}{2} \left(1 - \cos \left(\frac{\pi n}{F} \right) \right), n = 0 \dots F,$$

$$G = \max(0, K - N), S = \min(N, K), F = \min(P, K).$$

Алгоритм придатний і для багаторазового повторення, якщо його застосувати послідовно для вставки кожного нового періоду. Відповідно, зниження висоти основного тону періодичного сигналу вищенаведеним алгоритмом збільшує його тривалість. У цьому випадку, для компенсації (зменшення тривалості) також важливо уникнути збою фази [4, с. 27]. Тому використовується аналогічний підхід, а саме масив значень нового періоду, який необхідно вставити замість наявних двох, визначається також суперпозицією:

$$M_k = TR_k + TL_k, k = 0 \dots K, \begin{cases} TL_k = 0, k = 0..G \\ TL_k = L_{k+P-k} W_{k-G}^1, k = G..K' \end{cases} \begin{cases} TR_k = 0, k = G..K \\ TR_k = R_k W_k^2, k = 0..F' \end{cases}$$

$$W_n^1 = \frac{1}{2} \left(1 - \cos \left(\frac{\pi n}{S} \right) \right), n = 0 \dots S, W_n^2 = \frac{1}{2} \left(1 - \cos \left(\frac{\pi n}{F} \right) \right), n = 0 \dots F,$$

$$G = \max(0, K - P), S = \min(P, K), F = \min(N, K).$$

Вдосконалення систем синтезу мови можливо лише за допомогою використання всебічної та всеохоплюючої інформації про мову, яка може бути органічно впроваджена в систему на основі запропонованого вище підходу. Шляхом удосконалення і ускладнення інтонаційних фарб різного рівня можливо покращити якість синтезованої мови, не змінюючи при цьому загальну структуру і основні модулі синтезатора мови, що дозволяє залучати до вдосконалення голосових інтерфейсів лінгвістів, не знайомих з обробкою сигналів і основними алгоритмами синтезу [5, с. 7]. Запропонований підхід дозволяє зробити технологію синтезу мови більш наближеною до людської мови.

Таким чином можна зробити висновок що за допомогою запропонованого в роботі алгоритму ми можемо скорегувати тривалість мовних алофонів не втрачаючи природність мови. Зміна тривалості є неодмінною частиною яка дозволяє нам регулювати темп мови. Подальше

дослідження у цьому напрямку дозволить покращити синтез інтонаційної мови, а саме одіну з основних складових темп мови.

Література

1. Рыбчинский С.В. СИНТЕЗ РЕЧИ / Учебное пособие по дисциплине «Синтез речи» // СПб: Университет ИТМО, 2014 – 92 с.
2. Сегментация и дифонное распознавание речевых сигналов / А.К. Бурибаева, Г.В. Дорохина, А.В. Ниценко, В.Ю. Шелепов // Тр. СПИИРАН. – 31 (2013). – С. 20-42.
3. Чистиков П.Г., Рыбчинский С.В. Проблемы естественности речевого сигнала в системах синтеза / Компьютерные инструменты в образовании, 2011. №1.
4. Malah D. Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals / IEEE Transactions on Acoustics, Speech and Signal Processing. - 1979. - Vol. 27, 2. - P. 121-133.
5. Pitrelli, J.F. et al. The IBM expressive text-to-speech synthesis system for American English / Audio, Speech, and Language Processing, IEEE Transactions on V. 14, Issue 4, July 2006 Page(s):1099-1108.