

УДК 534.78+519.615

АНАЛІЗ ЧАСОВОГО РОЗПОДІЛУ ВИПРОМІНЮВАЛЬНИХ СПЕКТРАЛЬНИХ ЧАСТОТ І МЕТОДИКА ЇХ ОБЧИСЛЕННЯ

В. Ю. СЕМЕНОВ

Державне науково-виробниче підприємство “Дельта”, Київ

Одержано 15.09.2009

Розглянуто задачу обчислення випромінювальних спектральних частот (ВСЧ), які є альтернативною формою представлення параметрів голосового тракту. Досліджені властивості взаємного розташування ВСЧ на суміжних фреймах мовного сигналу. На їхній базі запропоновано новий ошадливий метод обчислення ВСЧ. Проведено порівняння обчислювальних характеристик отриманого методу з характеристиками стандартного голосового кодеку AMR-WB. Проаналізовано ефективність методу в умовах втрат голосових даних у каналі зв'язку.

Рассмотрена задача вычисления излучательных спектральных частот (ИСЧ), являющихся альтернативной формой представления параметров голосового тракта. Изучены свойства взаимного расположения ИСЧ на смежных фреймах речевого сигнала. На их основе предложен новый экономичный метод вычисления ИСЧ. Проведено сравнение вычислительных характеристик полученного метода с характеристиками стандартного речевого кодека AMR-WB. Проанализирована эффективность метода в условиях потери голосовых данных в канале связи.

The paper deals with considering of the problem of calculation of the immittance spectral frequencies (ISF) being the alternative representation of parameters of the vocal tract. Properties of the inter-frame ISF relative position have been studied. On this base, the novel efficient method for ISF computation has been proposed. The computational characteristics of this method have been compared with those of the standard AMR-WB speech codec. Performance of the method has been analyzed for conditions of voice data loss in the communication channel.

ВСТУП

Розвиток методів цифрового аналізу та обробки мовних сигналів потребує простих і водночас ефективних моделей мовних сигналів, заснованих на знанні фізичних механізмів їх генерування та сприйняття. Більшість сучасних методів обробки мовних сигналів базується на авторегресивній (АР) моделі мовного сигналу [1, 2], в якій форма голосового тракту визначається за допомогою АР коефіцієнтів a_k , $k=1, 2, \dots, p$.

Проте безпосередньо АР коефіцієнти зазвичай не застосовуються в системах кодування або розпізнавання мовних сигналів. Це пояснюється їхньою великою спектральною чутливістю й відсутністю чітких динамічних діапазонів. Тому їх зазвичай перетворюють у деякі альтернативні набори параметрів, наприклад, лінійні спектральні частоти (line spectral frequencies) – ЛСЧ [5] або випромінювальні спектральні частоти (immittance spectrum frequencies) – ВСЧ. Останні були введені дещо пізніше [6] і знайшли застосування, зокрема, при кодуванні широкосмугових мовних сигналів (з частотою дискретизації > 16 кГц) [3, 4].

Обчислення ЛСЧ та ВСЧ пов'язано з розв'язанням нелінійних рівнянь [5–9], що небажано з точки зору реалізації в системах реального часу внаслідок можливості непередбачених затримок та накопичення похибок. Особливо це стосу-

ється цифрових сигнальних процесорів з фіксованою комою. Найрозповсюдженіші методи розв'язання таких рівнянь – сіткові [10] та методи пониження ступеня [11]. У праці [12] для прискорення швидкодії сіткових методів було застосовано попереднє прогнозування можливих значень ЛСЧ.

Новий підхід до обчислення ЛСЧ, заснований на загальних принципах розв'язання нелінійних рівнянь, було запропоновано у праці [7]. У працях [8, 9] показана можливість покращити цей підхід на основі введеної властивості міжфреймової впорядкованості.

У цьому дослідженні розвинуто підхід до обчислення випромінювальних спектральних частот, який було намічено в публікації [9]. Основні завдання статті полягають у:

- 1) вивчення властивостей часового розподілу ВСЧ, корисних для покращення процедур їх обчислення;
- 2) розробка алгоритму обчислення ВСЧ, який би ефективно використовував ці властивості;
- 3) дослідження ефективності методу, в тому числі за умов втрати голосових даних у каналі зв'язку.

У першому розділі аналізується взаємне розташування ВСЧ на суміжних часових відрізках

(фреймах) мовних сигналів та пояснюється, як встановлені властивості їхнього часового розподілу можуть бути використані для обчислення ВСЧ. На базі властивості міжфреймової впорядкованості ВСЧ у другому розділі запропоновано алгоритм їх обчислення. Третій розділ присвячено аналізу ефективності методу, зокрема за умов втрати фреймів мовного сигналу в каналі зв'язку.

1. ДОСЛІДЖЕННЯ РОЗТАШУВАННЯ ВСЧ НА СУМІЖНИХ ЧАСОВИХ ФРЕЙМАХ

1.1. Означення та властивості ВСЧ

Як відомо, обчислення ВСЧ ω_k , $k=1, 2, \dots, p-1$ зводиться до пошуку коренів поліномів $P_1(x)$ та $P_2(x)$ [6]:

$$P_1(x) = \sum_{k=0}^{p/2} r_k^{(1)} x^{p/2-k} = 0,$$

$$P_2(x) = \sum_{k=0}^{p/2-1} r_k^{(2)} x^{p/2-1-k} = 0.$$

Корені цих рівнянь відповідають косинусам ВСЧ ($x = \cos(\omega)$) і задовольняють властивості внутрішньофреймової впорядкованості:

$$x_{i-1} < x_i < x_{i+1}, \quad i = 2, \dots, p-2, \quad (1)$$

де непарні косинуси відповідають поліному $P_1(x)$, а парні – $P_2(x)$.

Традиційно відшукування усіх коренів рівняння складається з процедур локалізації та уточнення їхніх значень. Методика обчислення ВСЧ, яка пропонується, включає такі етапи:

- 1) локалізацію коренів полінома $P_2(x)$;
- 2) уточнення коренів полінома $P_2(x)$;
- 3) уточнення коренів полінома $P_1(x)$ (їхня локалізація відбувається автоматично завдяки властивості (1)).

Розглянемо властивості ВСЧ, важливі для їхньої локалізації.

На рис. 1 зображено осцилограму та відповідну динаміку зміни ВСЧ для мовного фрагменту, що належить диктору-чоловіку. Запис було дискретизовано з частотою 16000 Гц і потім переквантовано до 12800 Гц. Обчислення ВСЧ відбувалось на часових фреймах тривалістю 20 мс при порядку АР моделі $p=16$.

З графіків видно, що у більшості випадків ВСЧ мають властивість міжфреймової впорядкованості [9] і, як наслідок, їхніх косинусів:

$$x_{i-1}^{(n-1)} < x_i^{(n)} < x_{i+1}^{(n-1)}, \quad i = 2, \dots, p-2. \quad (2)$$

Тут верхніми індексами позначені номери фреймів.

З точки зору локалізації ВСЧ важливими є відповіді на такі питання.

1. Чи може інтервал $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$, $i=2, 4, \dots, p-2$ містити більше двох коренів полінома $P_2(x)$?
2. Наскільки часто інтервал $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$, $i=2, 4, \dots, p-2$ містить єдиний корінь полінома $P_2(x)$, тобто виконується властивість міжфреймової впорядкованості ВСЧ?

Для відповіді на ці питання ми дослідили ВСЧ широкосмугових мовних сигналів, переквантованих з 16000 до 12800 Гц. При цьому використовувалась база мовних сигналів дев'яти дикторів загальною тривалістю 16 хвилин. До того ж застосовувались АР моделі з порядками від 14 до 20. У результаті були зроблені такі висновки.

1. Зафіксовано лише 0.002 % випадків, у яких більш ніж два кореня полінома $P_2(x)$ розташовувались між коренями полінома $P_1(x)$ попереднього часового фрейму.
2. З табл. 1 видно, що досліджувана властивість міжфреймової впорядкованості наявна у переважній більшості випадків (від 95.3 % для $p=20$ до 97.2 % для $p=14$).

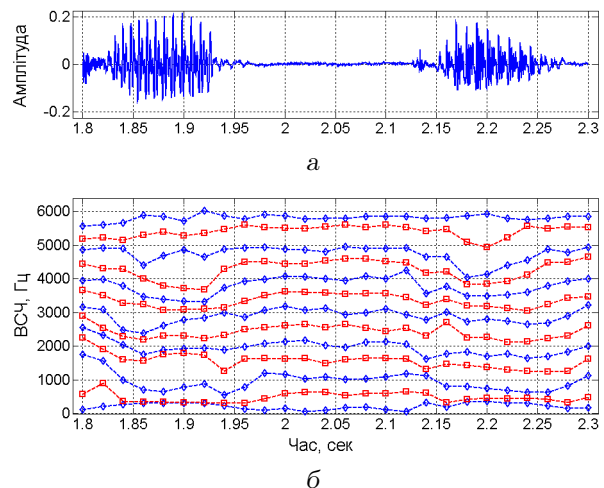


Рис. 1. Осцилограма мовного сигналу (а) і відповідний часовий розподіл ВСЧ (б)

1.2. Обговорення встановлених властивостей ВСЧ

Встановлені властивості ВСЧ мають ключове значення для процедури їхньої локалізації. Відповідно до другої властивості, локалізація коренів полінома $P_2(x)$ практично зводиться до порівняння знаків полінома $P_2(x)$ у точках, які відповідають непарним косинусам попереднього фрейму:

$$\{-1, x_3^{(n-1)}, x_5^{(n-1)}, \dots, x_{p-3}^{(n-1)}, 1\}.$$

Якщо знаки $P_2(x_{i-1}^{(n-1)})$ і $P_2(x_{i+1}^{(n-1)})$ не співпадають, то на цьому інтервалі знаходиться єдиний корінь (присутність будь-якої іншої непарної кількості коренів виключається завдяки першій властивості).

Щоб унаочнити сказане, звернімося до прикладу, зображеного на рис. 2. Інтервали $[-1, x_3^{(n-1)}]$, $[x_5^{(n-1)}, x_7^{(n-1)}]$, $[x_9^{(n-1)}, x_{11}^{(n-1)}]$, $[x_{13}^{(n-1)}, 1]$ задовольняють властивості міжфреймової впорядкованості, а значить, містять один корінь. Для встановлення присутності коренів на інтервалах $[x_3^{(n-1)}, x_5^{(n-1)}]$ та $[x_9^{(n-1)}, x_{11}^{(n-1)}]$ застосовується метод їхньої послідовної бісекції [4].

Табл. 1. Процентна кількість часових фреймів, для яких корені полінома $P_2(x)$ чергуються з коренями полінома $P_1(x)$ попереднього фрейму

p	14	16	18	20
%	97.2	96.6	96.0	95.3

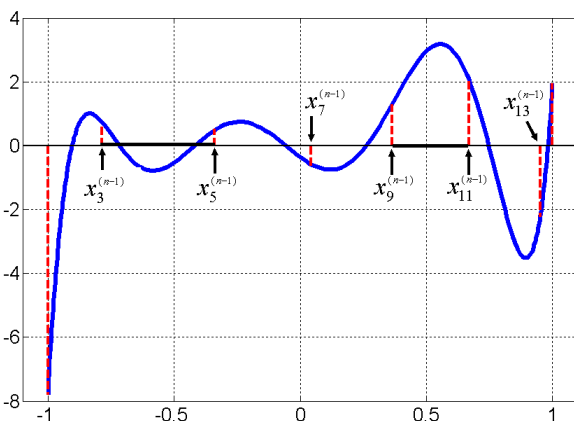


Рис. 2. Приклад взаємного розташування косинусів ВСЧ на суміжних часових фреймах

2. АЛГОРИТМ ОБЧИСЛЕННЯ ВСЧ

Відповідно до встановлених властивостей міжфреймового розташування ВСЧ, пропонується наступна структура алгоритму їх обчислення.

Вихідні дані

АР коефіцієнти поточного фрейму a_k , $k=1, 2, \dots, p$; значення косинусів ВСЧ $x_1^{(n-1)}$, $x_2^{(n-1)}, \dots, x_{p-1}^{(n-1)}$ обчислені на попередньому часовому фреймі.

Крок 1. Обчислення коефіцієнтів $P_2(x)$

АР коефіцієнти $\{a_k\}$ трансформуються у коефіцієнти полінома $P_2(x)$ за допомогою Чебишевських поліномів [10].

Крок 2. Локалізація парних косинусів ВСЧ

Перевіряються знаки полінома $P_2(x)$ у точках $\{-1, x_3^{(n-1)}, x_5^{(n-1)}, \dots, x_{p-3}^{(n-1)}, 1\}$.

Якщо $\text{sign}(P_2(x_{i-1}^{(n-1)})) \neq \text{sign}(P_2(x_{i+1}^{(n-1)}))$, то інтервал $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$ містить єдиний корінь.

Якщо $\text{sign}(P_2(x_{i-1}^{(n-1)})) = \text{sign}(P_2(x_{i+1}^{(n-1)}))$, то присутність кореня на інтервалі $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$ перевіряється за допомогою методу послідовної бісекції інтервалу [4].

Крок 3. Уточнення парних косинусів ВСЧ

Значення парних косинусів ВСЧ уточнюються аналогічно до алгоритму [4].

Крок 4. Уточнення непарних косинусів ВСЧ

Значення непарних косинусів ВСЧ уточнюються на інтервалах $[-1, x_2^{(n)}]$, $[x_2^{(n)}, x_4^{(n)}]$, \dots , $[x_{p-2}^{(n)}, 1]$ аналогічно до алгоритму [4].

Крок 5. Перетворення до частотної області

ВСЧ остаточно обчислюються за допомогою перетворення $\omega = \arccos(x)$.

Зауваження

У разі втрати попереднього фрейму в каналі зв'язку за опорні точки на кроці 2 обираються непарні косинуси останнього коректно прийнятого фрейму.

Табл. 2. Порівняння обчислювальних характеристик методів обчислення ВСЧ

МЕТОД	$N_{\text{сєр}}$	$N_{\text{макс}}$	WMOPS _{сєр}	WMOPS _{макс}
AMR-WB	136.6	142.0	0.80	0.83
Запропонований, без втрат	58.8	82.0	0.34	0.47
Запропонований (FER = 5 %)	58.8	82.0	0.34	0.47
Запропонований (FER = 20 %)	59.5	82.0	0.35	0.47

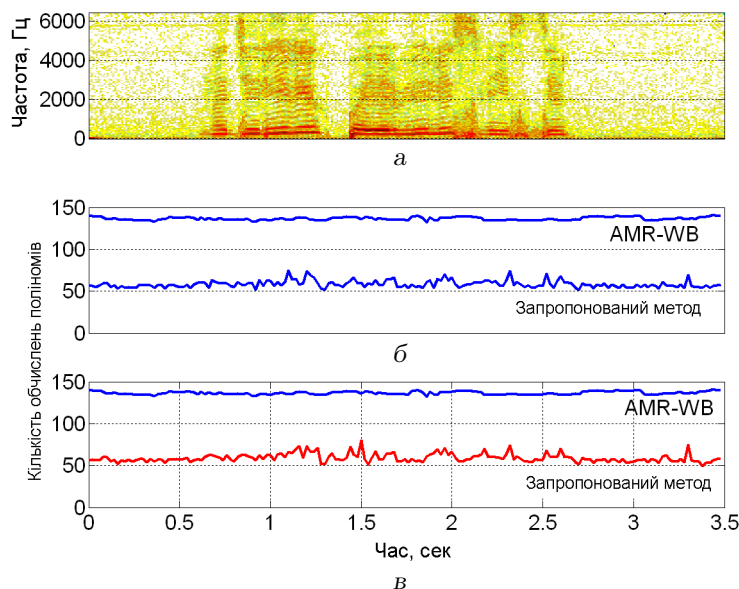


Рис. 3. Спектрограма тестового сигналу (а), кількості обчислень поліноміальних функцій для FER = 5 % і 20 % (б і в відповідно)

3. ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Ефективність запропонованого методу обчислення ВСЧ було перевірено для порядку AP моделі $p=16$. При цьому використовувалась база мовних сигналів, описана у попередніх розділах. У табл. 2 наведені середні й максимальні обчислювальні витрати у термінах WMOPS (weighted millions operations per second – мільйонів зважених операцій за секунду) та кількість обчислень поліноміальних функцій за один часовий фрейм (N), які відповідають запропонованому алгоритму й алгоритму голосового кодеку AMR-WB [4]. Доречною є також перевірка ефективності методу в умовах втрати голосових пакетів. Для цього були промодельовані канали зв'язку з імовірностями втрати фреймів (FER – frame error rates) 5 та 20 % відповідно.

З таблиці видно, що запропонований алгоритм забезпечує зниження середніх обчислювальних ви-

трат приблизно в 2.3 рази у порівнянні із сітковим підходом AMR-WB. Іншою важливою його перевагою є те, що відповідна максимальна кількість поліноміальних функцій $N_{\text{макс}}$ в 1.7 разів нижча за середнє значення для підходу AMR-WB. Ще один важливий висновок полягає в тому, що навіть втрати фреймів з ймовірністю 20 % майже не впливають на працю методу. Це пояснюється повільною зміною форми голосового тракту на протязі кількох десятків мілісекунд.

На рис. 3 наведено приклад розподілу кількості обчислень поліноміальних функцій N для підходу AMR-WB [4] й запропонованого методу для мовного сигналу диктора-жінки тривалістю 3.5 с. Очевидно, що мінімальні обчислювальні витрати мають місце в паузах, заповнених фоновим шумом, а невеликі піки в розподілі операцій пов'язані з раптовими переходами від одного звуку до іншого.

Зазначимо, що на відміну від методів послідовного пониження ступеня [11], запропонований ме-

тод не призводить до послідовного накопичення похибок обчислень. Тому він добре підходить для реалізації на цифрових сигнальних процесорах з фіксованою комою (fixed-point DSPs). Так, одна з версій методу була застосована при побудові голосового вокодеру (4800 біт/с), реалізованого на процесорі ADSP 2191.

ВИСНОВКИ

1. Розглянуто задачу обчислення випромінювальних спектральних частот (ВСЧ). Досліджені властивості взаємного розташування ВСЧ на суміжних фреймах мовного сигналу, зокрема властивість їх міжфреймової впорядкованості, на базі яких запропоновано новий економічний метод обчислення ВСЧ.
2. Показано, що середню кількість операцій зменшено у 2.3 рази у порівнянні з сітковим підходом. Окрім цього, максимальна кількість операцій виявилась у 1.7 рази меншою, ніж при стандартному підході, застосованому в голосовому кодеку AMR-WB. Також проаналізовано ефективність методу в умовах втрат голосових даних у каналі зв'язку. Показано, що втрати фреймів з імовірністю до 20 % майже не впливають на роботу методу, що пояснюється повільною зміною форми голосового тракту з часом. Стійкість роботи запропонованого методу продемонстровано при застосуванні в широкосмуговому вокодері з фіксованою комою.

1. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов.– М.: Радио и связь, 1981.– 496 с.
2. Калужный А. Я., Семенов В. Ю. Экономичный метод очистки речи от шума, основанный на блочном представлении сигнала в пространстве состояний и векторном квантовании // Акуст. вісн.– 2002.– 5, N 3.– С. 28–34.
3. Besette B., Salami R., Lefebvre R., Jelinek M., Rotola-Pukkila J., Vainio J., Mikkola H., Jarvinen J. The adaptive multirate wideband speech codec (AMR-WB) // IEEE Trans. Speech Audio Proces.– 2000.– 10.– P. 19–41.
4. 3GPP 26.190. Adaptive multi-rate-wideband (AMR-WB) speech codec; Transcoding functions.– 2005 [электронный ресурс]. Режим доступа к сайту: <http://www.3gpp.org>.
5. Itakura F. Line spectrum representation of linear predictive coefficients of speech signals // J. Acoust. Soc. Amer.– 1975.– 57, Suppl. 1.– P. S35.
6. Bistritz Y., Peller S. Immittance spectral pairs (ISP) for speech encoding // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– 1993.– 2.– P. 9–12.
7. Semenov V., Kalyuzhny A., Kovtonyuk A. Efficient calculation of line spectral frequencies based on new method for solution of transcendental equations // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– 2003.– 2.– P. 457–460.
8. Семенов В. Ю. Новый подход к вычислению линейных спектральных частот речевых сигналов, основанный на свойстве межфреймовой упорядоченности // Акуст. вісн.– 2004.– 7, N 3.– С. 55–64.
9. Semenov V. Computation of immittance and line spectral frequencies based on inter-frame ordering property // J. Comput.– 2007.– 2, N 7.– P. 75–80.
10. Kabal P., Ramachandran R. P. The computation of line spectral frequencies using Chebyshev polynomials // IEEE Trans. Acoust. Speech Signal Proces.– 1980.– 28.– P. 562–574.
11. Wu C.-H., Chen J.-H. A novel two-level method for the computation of LSP frequencies using a decimation-in-degree algorithm // IEEE Trans. Speech Audio Proces.– 1997.– 5.– P. 106–115.
12. Dumitrescu B., Tabus I. Predictive LSF computation // Signal Proces.– 2002.– 81.– P. 2019–2031.