Мир, 1982. – 790 с. 3. N. Papamarkos and B. A new approach for multilewel threeshold selection. Gatos Grafhhical Models and Image Processing. 56(5), 1994– p. 357–370. 4. Szala J. Zastosovwanie metod kompputerowej analizy obrazu do ilosciowej oceny stryktury materialow. // W. POLITECHNIKA SLASKA, Zeszyty naukowe. – 2000. – № 1518.–167с. 5. Иванюк В.Г., Лобур М.В., Лау Г. Распознавание компонентного состава по изображению. // Материалы IV Междунар. научно – методической конф. "Дистанционное обучение – образовательная среда XXI века" (10–12 ноября 2004г.). Минск, БГУИР, 2004.-С.444-447. 6. Іванюк В.Г., Капшій О.В., Косаревич Р.Я., Лау Г. Інформаційна оцінка і виділення фрагментів кольорових зображень // Радіоелектроніка і інформатика. – 2004. – № 3. С. 99–102. 7. Базылев И.Т., Дуничев К.И., Иваницкая В. П. Геометрия. – М.: Просвещение, 1974. – 351 с.

УДК 638.235.231

Б.П. Русин, Л.В. Чирун
Національний університет "Львівська політехніка",
кафедра ІСМ

АДАПТИВНІ АЛГОРИТМИ АПРОКСИМАЦІЇ КОЕФІЦІЄНТІВ МОДЕЛІ СИНТЕЗУ МОВНИХ СИГНАЛІВ

Описано адаптивну модель синтезу голосових сигналів, основану на косинусному перетворенні Фур'є. Запропоновано застосування неперервних дробів для апроксимації коефіцієнтів цієї моделі. Описано алгоритми: знаходження значення неперервних дробів, розбіжних у класичному розумінні. Ця процедура використовується для реалізації моделі голосового тракту людини.

In this article adaptive model for speech synthesis based on cosine Fourier transform is described. Application of continued fractions for better aproximation of model coefficients is proposed. Algorithms for continued fraction calculation, non-convergent in classical meaning, are proposed. This procedure is used for the design of the zero-pole model of human vocal tract.

Вступ. Всюди, де виникає потреба в доставці голосової інформації, є місце для якісного синтезу мови. З використанням систем синтезу мови *мова-текст* (*TTS - Text-to-Speech System*) задача доставки голосової інформації може бути розв'язана будь-де. Користувач має можливість вибирати голос залежно від потреб для оптимізації свого застосування, і, навіть більше того, існують голоси спеціальних типів залежно від виду доставки. Сучасні TTS системи можна оптимізувати до розмірів малого пакета. Сучасний розвиток технологій мобільного зв'язку дає змогу використовувати TTS системи для передавання синтезованої мови на ручні пристрої з використанням мереж стільникового зв'язку. Синтезована мова підтримує *SSML*, *Voice XML* теги, стандарти *Microsoft(R) SAPI* для розширення галузей застосування, зручності клієнтів і спрощення користування системами.

Сучасні галузі застосування TTS систем:

– *Допоміжні* – несуть силу мовлення, читання і навчання в життя.

– *Стільниковий зв'язок* – доставка голосових повідомлень у реальному часі на різноманітні пристрої.

– *Урядові проекти* – застосування TTS систем для побудови аплікацій, що дають змогу передавати голосову інформацію як в приміщеннях, так і на виїзді.

– *Медицина* – Синтезована мова дає можливість реалізації різноманітних аплікацій для галузі охрони здоров'я (довідники, екстренна медична допомога).

– *Навігація* – Використання якісної синтезованої мови в ручних чи бортових системах навігації.

– *Іграшки та комп'ютерні ігри* – створення ефекту живої іграшки та надання комп'ютерним іграм більшої реалістичності.

– *Індустрія* – побудова необхідних комунікацій у різноманітних виробничих чи індустріальних сферах.

– *Прогноз погоди* – прогнози погоди у режимі реального часу.

Сьогодні існує два основних підходи, що використовуються для генерування синтетичної мови людини: *конкатенативний синтез* і *формантний синтез*. *Конкатенативний синтез* оснований на конкатенації (об'єднанні) сегментів записаної мови. Загалом він продукує природнішу мову. Однак природна варіація в мові і технології сегментування записів деколи в результаті дають мову, зовсім відмінну від природної.

Основними підтипами конкатенативного синтезу є:

– *Синтез з вибором одиниць мови.* На жаль, для отримання хорошої якості мови необхідні великі бази одиниць мови. Але при такій реалізації з погляду TTS систем отримуємо найкращу якість мови, при вдалому виборі мовних одиниць.

– *Дифонний синтез.* Використовує мінімальну базу сегментів мови, що складається з дифонів (звук-в-звук транскрипції), що використовує дана мова. Кількість дифонів залежить від фонетики мови. Речення перекладається на набір дифонів з використанням технології цифрової обробки сигналів та лінійного редиктивного кодування. У цьому випадку ми втрачаємо в якості мови, але виграємо в тому, що такі системи є набагато менші, ніж у попередньому випадку і тут отримується виграш з погляду комерції.

– *Синтез з спеціалізованою предметною областю.* У цьому випадку використовується перший підхід, але до розгляду береться лише база слів і фраз з певної предметної області (оголошення графіка руху, прогноз погоди). Така система є дуже обмежена з погляду різноманіття речень і слів, які можемо використовувати.

– *Формантний синтез* – взагалі не використовує бази звукових сигналів. Натомість вихідний сигнал повністю генерується, використовуючи акустичну модель. Параметри моделі змінюються з часом для створення сигналу штучної мови. Деколи цей метод синтезу називають *синтез за правилами.* На виході такі синтезатори дають роботоподібну мову, ЯКУ ніяк не можна прийняти за мову людини. Однак максимальна натуральність мови не завжди є головною метою TTS, і в цьому формантні синтезатори виграють у конкатенативних систем.

Постановка задачі. Основними проблемами під час побудови сучасних синтезаторів мови є мінімізація:

– *жорсткої пам'яті* (проблема виникає через необхідність збереження великих кількостей мовних одиниць у базі даних)

– *алгоритмічна складність* (складні алгоритми згладжування для відтворення якісної синтезованої мови)

– *швидкість обчислень* (велика кількість коефіцієнтів моделей, які потрібно обчислювати).

Частину цих проблем можна вирішити, якщо в моделях синтезу мови використовувати алгоритми теорії ланцюгових дробів. Ці алгоритми є простими, що дасть змогу зменшити необхідні об'єми жорсткої пам'яті при побудові моделі. Вони є легкі в реалізації, що спростить алгоритмічну складність задачі синтезу мови.

Використовуючи авторегресивний підхід, отримаємо змогу відкинути прив'язку до фізичних моделей, тим самим зменшивши кількість коефіцієнтів, необхідних для реалізації моделі, тобто спростивши її.

А оскільки ланцюгові дроби є хорошим апаратом апроксимації, що відомо ще з XV–XVI стст., то використання алгоритмів теорії ланцюгових дробів дасть можливість покращити точність обчислень коефіцієнтів моделі синтезу мови та отримати кращу синтезовану мову.

Виклад основного матеріалу дослідження. Загальна система голосового синтезу зображена на рис.1.

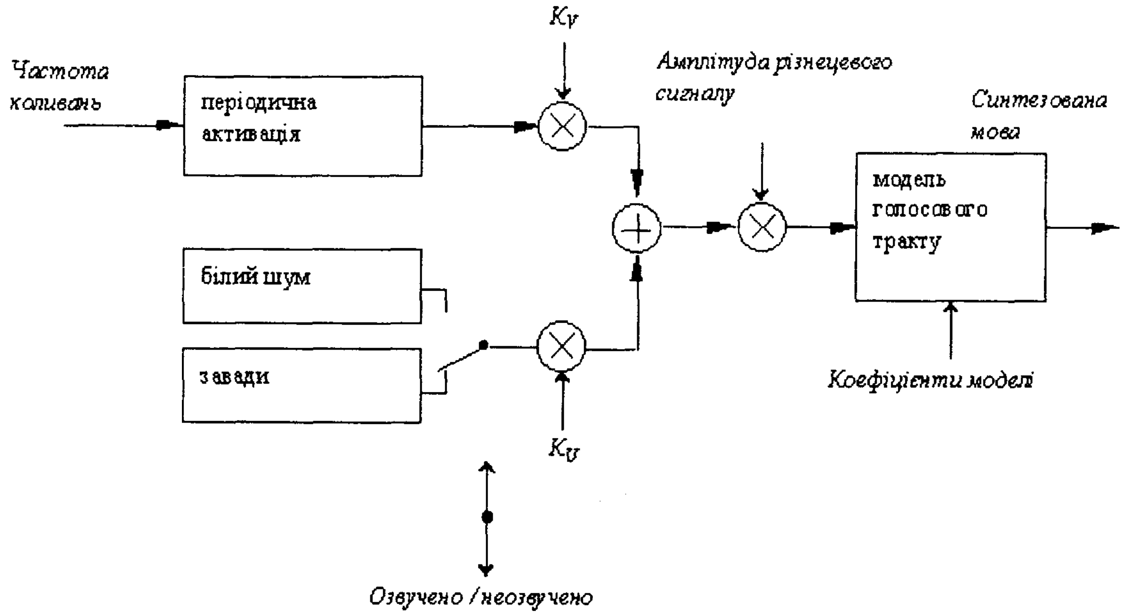

Рис. 1. Система синтезу мови

Голосова модель, основана на косинусному перетворені Фур'є для синтезу мови, має кращі властивості і застосування, меншу чутливість до ефектів квантування і в результаті дає природнішу синтезовану мову. Параметрами цієї моделі є коефіцієнти косинусного перетворення Фур'є. Ця модель базується на косинусному розкладі логарифмічного короткочасового голосового діапазону і синтез реалізований наближеним оберненим косинусним перетворенням Фур'є з використанням неперервних ланцюгових дробів. Цей підхід є параметричним і не оснований на ніяких спрощувальних припущеннях про голосову модель, тому що полюси як і нулі голосової моделі є обгрунтовані.

Нехай маємо логарифмічний діапазон $\ln\left|S\left(e^{j\omega T}\right)\right|$ сегмента голосових даних $\{s(n)\}$, де T – вибірковий інтервал, $f_s = \dfrac{1}{T}$ – вибіркова частота, і ω – кутова частота. Для подання даної функції використаємо дійсні коефіцієнти косинусного перетворення Фур'є $\{c_n\}$

$$\ln\left|S\left(e^{j\omega T}\right)\right| = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega T} \tag{1}$$

Комплексні коефіцієнти косинусного перетворення Фур'є $\{g_n\}$ дискретної системи з мінімальною стабільністю фази є випадкові і можуть бути пов'язані з $\{c_n\}$ такими співвідношеннями:

$$\begin{aligned} g_n &= c_n, \quad n = 0, N_F/2, \\ g_n &= 2c_n, \quad 0 < n < N_F/2, \\ g_n &= 0, \quad n < 0 \end{aligned} \tag{2}$$

Цифровий фільтр, чия логарифмічна величина відповідності апроксимує функцію $\ln\left|S\left(e^{j\omega T}\right)\right|$, визначається системою трансферних функцій

$$\tilde{S}(z) = e^{c_0} \exp \sum_{n=1}^{N_0-1} 2c_n z^{-n} = e^{c_0} \prod_{n=0}^{N_0-1} \exp\left[2c_n z^{-n}\right] \tag{3}$$

З (3) випливає, що система трансферних функцій $\tilde{S}(z)$ є добутком трансцендентних трансферних функцій

$$H_n(z) = e^{2c_n z^{-n}}, \quad 0 < n \le N_0 - 1. \tag{4a}$$

Відповідна імпульсна характеристика визначається так:

$$h_n(m)=\begin{cases}\dfrac{(2c_n)^i}{i!}, & m=ni, \quad i=0,1,2,\dots\\ 0, & m\neq ni\end{cases} \tag{4b}$$

Це означає, що система трансферних функцій $\tilde{S}(z)$ набуде вигляду

$$\tilde{S}(z)=e^{c_0}\prod_{n=0}^{N_0-1}H_n(z). \tag{5}$$

Для реалізації трансферної функції $H_n(z)$ за допомогою цифрового фільтра, необхідно знайти апроксимацію $H_n(z)$, яку можливо практично реалізувати. Одним з варіантів апроксимації експоненціальної функції в (4, а) є неперервні ланцюгові дроби [4]. Тоді система трансферних функцій в практичній голосовій моделі на основі косинусного перетворення Фур'є буде мати вигляд

$$\tilde{\tilde{S}}(z)=e^{c_0}\prod_{n=0}^{N_0-1}\tilde{H}_n(z). \tag{6}$$

Експоненціальну функцію виражену розкладом в неперервний дріб можна зобразити у вигляді такого розкладу [5], [6]:

$$e^x=\frac{1}{1-}\,\frac{x}{1+}\,\frac{x}{2-}\,\frac{x}{3+}\dots+\frac{x}{2-}\,\frac{x}{2s-1},\dots \tag{7}$$

де параметр $x=2c_nz^{-n}$. Точність наближення голосової моделі залежить не лише від кількості коефіцієнтів косинусного перетворення Фур'є в (3), але і від кількості членів неперервного дробу в (7), тобто від довжини неперервного дробу, що визначається s. Рекомендовано використовувати непарну кількість елементів неперервного дробу в (7). Це призводить до апроксимації експоненціальної функції раціональною функцією з однаковими степенями поліномів в чисельнику та знаменнику (при використанні апроксимацій Паде).

Граф потоку сигналу (*Signal Flow Graph* (SFG)) раціональної трансферної функції, що апроксимує часткову нераціональну трансферну функцію в (4, а), показаний на рис. 2.

Як було зазначено вище, похибка апроксимації для e^x визначається кількістю елементів неперервного дробу для розкладу експоненціальної функції в неперервний дріб. Ця похибка надалі також залежить від величин модулів дійсних коефіцієнтів косинусного перетворення Фур'є $|c_n|$.

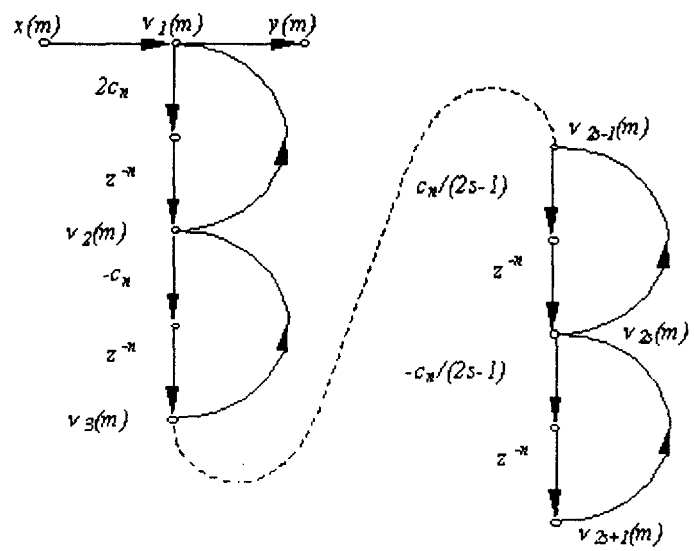

Рис 2. Граф потоку сигналу

Кількість відповідних комірок (рис.2) можна вибирати відповідно до величин коефіцієнтів косинусного перетворення Фур'є.

Для побудови адаптивного алгоритму апроксимації коефіцієнтів моделі синтезу мови необхідно реалізувати такі дії: обчислити значення ланцюгового дробу, який є коефіцієнтом моделі, визначити значення модуля ланцюгового дробу, визначити значення модуля аргументу ланцюгового дробу, встановити знак аргументу.

Відзначимо, що для встановлення з достатньо високою точністю величини модуля та аргументу комплексного числа, який є значенням розбіжного в класичному розумінні ланцюгового дробу, необхідно використати досить довгі послідовності підхідних дробів. Тому доцільно розглянути алгоритми обчислення значень ланцюгових дробів, виділяючи насамперед ті, що забезпечували б мінімальні витрати часу для отримання довгих послідовностей значень підхідних дробів. Алгоритмами, які б забезпечили мінімальні витрати часу, необхідного для отримання значення неперервного ланцюгового дробу та давали б достатньо високу точність апроксимації є Δ-алгоритм та ψ/φ-алгоритм. Для кожного з цих алгоритмів необхідно виконати 6n операцій для обчислення значення ланцюгового дробу та 7n операцій у разі обчислення серій підхідних дробів. Зауважимо, що ці два алгоритми є нисхідними, що дозволяє отримати значення наступного підхідного дробу без перерахунку значень всіх попередніх, що своєю чергою дає змогу отримати суттєвий виграш у часі, необхідному для обчислення значення ланцюгового дробу.

Δ-алгоритм виглядає так. Відома формула

$$\Delta f_n = f_n - f_{n-1} = \frac{P_n}{Q_n} - \frac{P_{n-1}}{Q_{n-1}} = \frac{(-1)^{n+1} a_1 \cdot a_2 \cdots a_n}{Q_{n-1} \cdot Q_n}.$$

Отже,

$$\frac{\Delta f_n}{\Delta f_{n-1}} = -\frac{a_n \cdot Q_{n-2}}{Q_n} = \frac{b_n}{z_n} - 1,$$

де

$$z_n = Q_n / Q_{n-1}, \quad z_n = b_n + \frac{a_n}{z_{n-1}}, \quad z_1 = b_1, \quad n = 1,2,...$$

З вищенаведених формул маємо

$$\Delta f_n = \left(\frac{b_n}{z_n} - 1\right) \cdot \Delta f_{n-1}, \quad \Delta f_1 = \frac{a_1}{b_1}.$$

Паралельно, обчислюючи значення підхідних дробів розкладу неперервного ланцюгового дробу, що зображає коефіцієнт моделі, будемо визначати значення модуля та значення модуля аргументу неперервного дробу.

Значення модуля неперервного дробу та модуль аргументу встановлюється за допомогою формул

$$\lim_{s \to \infty} \sqrt[s]{\prod_{i=1}^{s} \left| P_i / Q_i \right|} = r_o,$$

$$\pi \cdot \lim_{s \to \infty} \frac{k_s}{s} = |\varphi_0| \tag{8}$$

де P_i / Q_i – значення i-го підхідного дробу з множини, що містить s підхідних дробів, k_s – число від'ємних підхідних дробів з s підхідних дробів вихідного розкладу.

Оскільки попередньо описаний алгоритм дозволяє лише встановити модуль аргументу, але не його знак, тому виникає потреба в розробці додаткових процедур, які б дали можливість визначення знака аргументу комплексного числа. Коротко нагадаємо алгоритми визначення знака комплексного числа.

Алгоритм Z_1. *Якщо модуль аргументу комплексного числа, що є значенням "розбіжного" ланцюгового дробу, встановлений з аналізу знаків підхідних дробів за формулою* (8), *лежить в інтервалі* $0 < |\varphi| < \frac{\pi}{2}$, *то знак аргументу комплексного числа буде* ***додатним***, *якщо на "періоді"* *знайдуться додатні підхідні дроби, що утворюють* ***спадні*** *послідовності. Якщо на періоді знайдуться додатні підхідні дроби, що утворюють* ***зростаючі*** *послідовності, то розклад визначає значення комплексного числа з* ***від'ємним*** *аргументом.*

Алгоритм Z_2. *Якщо модуль аргументу комплесного числа, що є значенням "розбіжного" ланцюгового дробу, визначеного за формулою* (8) *знаходиться в інтервалі* $\frac{\pi}{2} < |\varphi| < \pi$, *то знак аргументу цього комплексного числа буде* ***додатним****, якщо на "періоді" модулі підхідних дробів, що мають від'ємні значення, утворюють* ***спадну*** *послідовність. Якщо на "періоді" від'ємні підхідні дроби утворюють по модулю* ***зростаючу*** *послідовність, то ланцюговий дріб визначає значення комплексного числа, що має* ***від'ємний*** *аргумент.*

Висновки. Голосове моделювання на основі косинусного перетворення Фур'є насправді належить до спектрального синтезу голосових сигналів і не основане на будь-яких спрощуючих апріорних міркуваннях про систему відтворення мови. Воно також містить інформацію про діапазон активуючого голосового тракту. Процедура є параметричною. Якщо використовувати синхронну висоту тону звуку, згенерована синтезована мова не відрізняється від природної мови. Процедура голосового моделювання на основі косинусного перетворення Фур'є вимагає виконання більше арифметичних операцій, ніж підходи, основані на лінійному предиктивному кодуванні, але структуру цифрового фільтра можна оптимізувати. Неперервні дроби пропонують цікавий інструмент не лише в синтезі мови. Апроксимація високого порядку алгебраїчних трансцендентних функцій може бути використана в системах біологічного і виробничого моделювання. Пряма реалізація неперервних дробів надалі надає можливість реалізації багатокоміркових структур.

1. J.D. Markel and A.H. Gray, linear Prediction of Speech (Berlin; Springer Verlag, 1976). 2. J. Makhoul, Stable and Efficient Lattice Methods for Linear Prediction, IEEE Trans. Acoustics, Speech and signal Processing, Vol., ASSP-25,No.5,October 1977, 423-428. 3. R. Vich and Z. Smekal, Continued Fractions in Digital Filter Synthesis, Proc. of Inter. Scient. Colloquium, September 18-21, 1995, Ilmenau, Germany, 353-356. 4. R. Vich and Z. Smekal, Digital Filter Realization of Nonrational Transfer Functions, Proc. of the First European Conference on Signal Analysis and Prediction, ECSAP-97, June 24-27, 1997, Prague, Czech Republic, pp. 179-182. 5. Шмойлов В.И. Периодические цепные дроби. – Львов: Академический Экспресс, 1998.- 219с. 6. Шмойлов В.И., Чирун Л.В. Комплексные числа и непрерывные дроби. – Львов, Меркатор, 2001. – 564с.