УДК 004.04

А. В. Грінін

АЛГОРИТМІЧНІ МЕТОДИ СИНТЕЗУ ЗВУКУ ТА ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ У ГЕНЕРАЦІЇ АУДІОСИГНАЛІВ

Вінницький національний технічний університет

Анотація:
Систематизовано сучасні алгоритмічні методи синтезу звуку та проаналізовано застосування методів машинного навчання у задачах генерації й трансферу аудіосигналів. Розглянуто ключові архітектурні рішення та їхній вплив на якість синтезованого звуку.
Ключові слова: синтез звуку, генерація аудіосигналів, машинне навчання.
Abstract:
Modern algorithmic methods of sound synthesis have been systematized, and the application of machine learning techniques in tasks of audio signal generation and transfer has been analyzed. Key architectural solutions and their impact on the quality of synthesized sound are considered.
Keywords: sound synthesis, audio signal generation, machine learning.

Вступ

Розробка методів синтезу звуку є однією з центральних задач у галузі цифрової обробки сигналів та аудіотехнологій. Починаючи з 1960-х років, еволюція синтезу звуку пройшла шлях від аналогових схем на дискретних компонентах до програмних алгоритмів і, зрештою, до систем на основі глибокого навчання. Сьогодні синтезовані аудіосигнали застосовуються у музичній індустрії, кінематографі, відеоіграх, системах мовленнєвого синтезу та медичних дослідженнях [1].

Ключовою проблемою традиційних цифрових методів синтезу є обмежена виразність при генерації тембрально складних або динамічно змінних звуків. Класичні алгоритми потребують ретельного ручного налаштування параметрів і не завжди здатні відтворити всю складність акустичної поведінки реальних джерел звуку [2]. Це обумовлює актуальність дослідження та впровадження методів машинного навчання, здатних навчатися необхідних аудіопредставлень безпосередньо з даних.

Метою роботи є аналіз основних алгоритмічних методів синтезу звуку та огляд сучасних нейромережевих підходів до генерації аудіосигналів, визначення їхніх переваг, обмежень і перспектив практичного застосування.

Методи синтезу звуку

Сучасна теорія цифрового синтезу виділяє кілька принципово різних підходів до генерації звукового сигналу, кожен із яких характеризується власним математичним апаратом та областю застосування.

Адитивний синтез базується на теоремі Фур'є: будь-який звуковий сигнал s(t) може бути представлений як скінченна або нескінченна сума синусоїдальних компонент [2]:

$$s(t) = \sum A_n \cdot \sin(2\pi \cdot f_n \cdot t + \varphi), \quad (1)$$

де A_n – амплітуда n-ї гармоніки, f_n – її частота, φ_n – початкова фаза. Теоретично метод дозволяє синтезувати довільний тембр, однак для реалістичного відтворення складних акустичних сигналів може знадобитися кілька сотень незалежних осциляторів, що накладає значні вимоги на обчислювальні ресурси [3].

Субтрактивний синтез реалізує принцип, протилежний адитивному: початковий спектрально насичений сигнал (пилоподібна або прямокутна хвиля) пропускається через фільтри, що відсікають небажані гармоніки. Архітектура класичного субтрактивного синтезатора включає генератор (VCO), фільтр із керованою частотою зрізу (VCF) та підсилювач (VCA) з огинаючою ADSR. Інтуїтивність

налаштування та характерне «тепле» звучання пояснюють широке застосування цього методу в аналоговому синтезі [4].

FM-синтез (частотна модуляція), розроблений Дж. Чоунінгом у 1967 р. та запатентований Yamaha у 1973 р., забезпечує генерацію складних спектрів шляхом модуляції несучого осцилятора (оператора-носія) модулятором [5]. Миттєва частота несучого сигналу описується виразом:

$$f(t) = f_c + I \cdot f_m \cdot \cos(2\pi \cdot f_m \cdot t), \quad (2)$$

де f_c – частота носія, $f_м$ – частота модулятора, I – індекс модуляції. Зміна індексу I в часі дозволяє динамічно перебудовувати спектральний склад сигналу, що особливо ефективно при синтезі металевих, дзвінкових та ударних тембрів. Реалізації на 4-6 операторах (DX7, FM8) досі широко застосовуються у звуковому дизайні.

Вейвтейблінг (wavetable synthesis) здійснює відтворення та плавну інтерполяцію між заздалегідь збереженими одноциклічними формами хвиль. Ефект динамічно змінюваного тембру при обчислювальній ефективності, близькій до простого осцилятора, зробив цей метод основою для широкого класу синтезаторів – від PPG Wave (1981) до сучасних Serum та Vital. Семплінг оперує реально записаними аудіофрагментами, транспонуючи їх за висотою методами тайм-стретчингу та пітч-шифтингу [3].

Нейронні мережі у синтезі аудіосигналів

Застосування методів глибокого навчання у задачах синтезу та обробки звуку набуло стрімкого розвитку протягом 2016-2024 рр. Ключовою відмінністю нейромережевих методів від класичних є здатність автоматично навчатися ефективних аудіопредставлень із великих масивів даних без ручного проектування синтетичних алгоритмів.

Авторегресивна модель WaveNet (Google DeepMind, 2016) стала першою архітектурою, що досягла якості синтезу мовлення, суб'єктивно невідмінної від натуральної. WaveNet генерує відліки аудіосигналу послідовно, умовно на контекст попередніх відліків, використовуючи дилатовані причинні згорткові мережі. Основним недоліком є значна обчислювальна вартість генерації в реальному часі (O(N) операцій на відлік).

Генеративно-змагальні мережі (GAN) виявилися ефективним інструментом для генерації коротких звукових семплів. Архітектура GANSynth (Google Magenta) та WaveGAN навчаються відображенню латентного простору у простір часових аудіосигналів у змагальному тренувальному процесі. GANSynth демонструє здатність до синтезу семплів музичних інструментів з контрольованим тембром та висотою тону, проте синтез тривалих музичних структур залишається складним завданням для цієї архітектури.

Принципово важливим підходом є DDSP (Differentiable Digital Signal Processing, Google Research, 2020), що поєднує диференційовані блоки класичного DSP (осцилятори, фільтри, ревербератори) з нейронними мережами, що навчаються керуючих параметрів для цих блоків. Такий гібридний підхід дозволяє використати фізичні обмеження задачі як структурний пріоритет моделі, що забезпечує більш ефективне навчання та інтерпретовані параметри синтезу. DDSP успішно застосовується для трансферу тембру (timbre transfer) та реалістичного синтезу одноголосних інструментів.

Дифузійні моделі (diffusion models) стали найпотужнішим напрямком генерації аудіо у 2022–2024 рр. Системи AudioLM та AudioCraft (Meta), засновані на архітектурі MusicGen, здатні генерувати музичні фрагменти тривалістю кілька хвилин за текстовим описом або мелодичним прикладом. На відміну від GAN, дифузійні моделі відрізняються стабільністю тренування та здатністю моделювати складні умовні розподіли аудіоданих, що відкриває перспективи їх застосування у системах звукового дизайну.

Паралельно з архітектурними інноваціями розвиваються методи оцінки якості синтезу, що поєднують об'єктивні метрики (PESQ, STOI, FAD) та суб'єктивні слухові тести. Для практичних застосувань важливими залишаються питання обчислювальної ефективності та латентного контролю, що стимулює дослідження оптимізованих моделей і апаратного прискорення для реального часу. Також зростає увага до етичних і правових аспектів генерації аудіо, зокрема до проблеми авторства, фальсифікації голосу та необхідності механізмів маркування синтетичного контенту.

Висновок

Кожен із розглянутих чотирьох основних алгоритмічних методів відповідає специфічному класу задач звукового дизайну. Класичні методи забезпечують детерміновану та ресурсоефективну генерацію, але потребують ручного налаштування параметрів і не адаптуються до нових тембральних завдань без перепроектування алгоритму.

Аналіз нейромережевих підходів свідчить про принципово вищий рівень тембральної виразності та гнучкості порівняно з класичними методами. Гібридна архітектура DDSP, що інтегрує диференційовані DSP-блоки у нейромережевий конвеєр, видається найбільш перспективною для задач з обмеженими даними та вимогами до інтерпретованості параметрів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Roads C. The Computer Music Tutorial. ─ Cambridge, Massachusetts: MIT Press, 2023. ─ 1288 p.
2. Zölzer U. Digital Audio Signal Processing – 2nd ed. ─ Chichester: John Wiley & Sons, 2008. ─ 618 p.
3. Vail M. The Synthesizer: A Comprehensive Guide to Understanding, Programming, Playing, and Recording. ─ New York: Oxford University Press, 2014. ─ 448 p.
4. Réveillac J.-M. Synthesizers and Subtractive Synthesis. Vol 2. – Chichester: John Wiley & Sons, 2024. ─ 304 p.
5. Chowning J. M. The Synthesis of Complex Audio Spectra by Means of Frequency Modulation // Journal of the Audio Engineering Society. – 1973. – Vol. 21, №7. – P. 526-534.

Грінін Андрій Вікторович – студент групи 1ПІ-25м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, email: andergrinin@gmail.com;

Andrii Hrinin – student of group 1PI-25m, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine, email: andergrinin@gmail.com;