

УДК 004.934.5

М. Кліщ; В. Яцишин, к. т. н., доц.

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

СТРУКТУРНІ ОБМЕЖЕННЯ ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ГЕНЕРАТИВНОГО МОДЕЛЮВАННЯ АУДІОСИГНАЛІВ

UDC 004.934.5

M. Klishch; V. Yatsyshyn, PhD., Assoc. Prof.

STRUCTURAL CONSTRAINTS FOR IMPROVING THE EFFICIENCY OF GENERATIVE AUDIO SIGNAL MODELING

У багатьох задачах обробки мовлення та аудіо ключовим обмеженням є нестача якісно розмічених даних та ліцензованих даних. Це стосується як малоресурсних мов, так і спеціалізованих доменів, включно зі співочим мовленням. У таких умовах виникає потреба в методах, які забезпечують ефективне представлення сигналу та стійке навчання моделей за обмежених обсягів даних. Структурні обмеження в генеративних аудіомоделях є ключовим чинником підвищення даноефективності, оскільки вони зменшують розмірність простору функцій, який має бути вивчений мережею, та узгоджують оптимізацію з фізичною природою сигналу.

У рамках диференційовної цифрової обробки сигналів (DDSP) такі обмеження реалізуються через апіорно фіксовані гармонічні й шумові генератори, що дозволяє моделі навчатися лише керуванню низьковимірними тембральними параметрами замість відтворення високовимірного аудіосигналу [1, 2]. Зокрема у дослідженнях [3, 4, 5] показано ефективність даного фреймворку для моделювання мовлення та співочого мовлення.

Додаткове структурне обмеження у вигляді багаторівневого залишкового векторного квантування (RVQ) дискретизує латентний простір і формує компактне, стійке до шумів представлення, яке зменшує надлишкову варіативність і сприяє регуляризації моделі [6].

Література

1. Engel, J., Hantrakul, L., Gu, C., & Roberts, A. (2020). DDSP: Differentiable digital signal processing. arXiv preprint arXiv:2001.04643.
2. Hayes, B., Shier, J., Fazekas, G., McPherson, A., & Saitis, C. (2024). A review of differentiable digital signal processing for music and speech synthesis. *Frontiers in Signal Processing*, 3, 1284100.
3. Liu, Y., Yu, B., Lin, D., Wu, P., Cho, C. J., & Anumanchipalli, G. K. (2024, December). Fast, high-quality and parameter-efficient articulatory synthesis using differentiable DSP. In *2024 IEEE Spoken Language Technology Workshop (SLT)* (pp. 711-718). IEEE.
4. Nercessian, S. (2022). Differentiable WORLD synthesizer-based neural vocoder with application to end-to-end audio style transfer. arXiv preprint arXiv:2208.07282.
6. Agrawal, P., Koehler, T., Xiu, Z., Serai, P., & He, Q. (2024, April). Ultra-lightweight neural differential dsp vocoder for high quality speech synthesis. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 10066-10070). IEEE.
2. 6. Défossez, A., Copet, J., Synnaeve, G., & Adi, Y. (2022). High fidelity neural audio compression. arXiv preprint arXiv:2210.13438.