**УДК 004.934**

# СРАВНЕНИЕ ОБЩЕЙ УСТОЙЧИВОСТИ К ШУМАМ И УСТОЙЧИВОСТИ К МАСКИРОВКЕ ШУМОВЫХ ДИПФЕЙКОВ ДЛЯ СИСТЕМ ОБНАРУЖЕНИЯ ДИПФЕЙКОВ

**Галимов А. А.**[1]
**Научный руководитель – Чирковский А. Д.**[1]
[1]Университет ИТМО
mr.artyomgalimov@gmail.com

## Введение

Развитие технологий синтеза речи и клонирования голосов на основе ИИ создаёт угрозы для систем голосовой аутентификации, требуя разработки надёжных методов антиспуфинга. Однако эффективность таких систем снижается в реальных условиях из-за шумовой маскировки. В данной работе исследуется устойчивость моделей к шумовым искажениям, а также влияние аугментации шумами на качество обнаружения голосовых дипфейков. Результаты показывают, что шумовая маскировка усиливает асимметричную деградацию моделей, где ошибки второго рода (FNR) растут сильнее, чем ложные срабатывания (FPR), особенно у моделей, обученных с шумовыми аугментациями.

## Основная часть

В данном исследовании было проведено тестирование четырёх моделей (xls-r-2b-antideepfake, xls-r-2b-antideepfake-nda, df_arena_1b, df_arena_500M) для оценки устойчивости систем голосового антиспуфинга к шумовой маскировке голосовых дипфейков и общей устойчивости к шумам [1, 2, 3]. Тесты проводились на нескольких наборах данных. Первый, InTheWild (31,779 записей) [4]. Второй, InTheWild аугментированный шумами из DEMAND (381,348 семплов) [5]. Третий, InTheWild аугментированный шумами из MS-SNSD (381,308 семплов) [6]. Аугментированные корпуса данных смешивались с соотношением сигнала к шуму 6, 12, 18, 24 дБ (на каждые 3 случайных шума 4 уровня SNR) при 16 кГц.

Для исследования различий в общей устойчивости к шумам и в устойчивости к шумовой маскировке, были взяты несколько метрик: Equal Error Rate (EER), а также специально разработанная для аугментированной шумами базы метрика Noise Influence, высчитываемая как среднее арифметическое метрик Noise Tolerance (отношение роста FNR к росту FPR на зашумленных данных в точке порога EER) и Noise Vulnerability (общий рост среднего числа ошибок на зашумленных данных в точке порога EER). Замеры на моделях одного размера с использованием и неиспользованием шумовой аугментации, а также на одной модели, но с разным числом параметров показали двоякость эффекта: С одной стороны, на моделях с применением аугментации наблюдается понижение EER, что часто воспринимается как хороший признак увеличения общей устойчивости к шумам, но с той же стороны понижается устойчивость к шумовой маскировке.

## Выводы

Результаты показали, что общая устойчивость модели к шумам и устойчивость к шумовой маскировке это разные величины, которые могут меняться независимо, что подтверждает необходимость в специальных метриках, бенчмарках и методиках оценки. Предлагаемая метрика позволяет выявлять подобные несовпадения.

## Литература

1. Ge W. et al. Post-training for Deepfake Speech Detection // arXiv. – 2025. – ApXiv:2506.21090. – Режим доступа: https://arxiv.org/abs/2506.21090 (дата обращения: 27.02.2026).

2. Kulkarni A. et al. DF Arena 1B V 1 - Universal Audio Deepfake Detection [Электронный ресурс] // Hugging Face. – 2025. – Режим доступа: https://huggingface.co/Speech-Arena-2025/DF_Arena_1B_V_1 (дата обращения: 27.02.2026).
3. Kulkarni A. et al. DF Arena 500M V 1 - Universal Audio Deepfake Detection [Электронный ресурс] // Hugging Face. – 2025. – Режим доступа: https://huggingface.co/Speech-Arena-2025/DF_Arena_500M_V_1 (дата обращения: 27.02.2026).
4. Muller N. M. et al. Does audio deepfake detection generalize? // arXiv. – 2022. – ApXiv:2203.16263. – Режим доступа: https://arxiv.org/abs/2203.16263 (дата обращения: 27.02.2026).
5. Thiemann J., Ito N., Vincent E. The diverse environments multi-channel acoustic noise database (DEMAND): a database of multichannel environmental noise recordings // Proceedings of Meetings on Acoustics. – 2013. – Vol. 19, № 1. – DOI: 10.1121/1.4800452 (дата обращения: 27.02.2026).
6. Reddy Ch. K. A. et al. A scalable noisy speech dataset and online subjective test framework // arXiv. – 2019. – ApXiv:1909.08050. – Режим доступа: https://arxiv.org/abs/1909.08050 (дата обращения: 27.02.2026).