

УДК 004.94:336.7:519.237

**Кокорев Борис Сергеевич**, к.т.н., преподаватель, Московский  
Финансовый Юридический Университет, г. Москва

**Корнеев Антон Артурович**, аспирант, Московский Финансовый  
Юридический Университет, г. Москва

## **СРАВНИТЕЛЬНЫЙ АНАЛИЗ АППАРАТНЫХ РЕСУРСОВ И МЕТОДОВ ХРАНЕНИЯ ГОЛОСОВЫХ ДАННЫХ В СИСТЕМАХ ГОЛОСОВОЙ ОБРАБОТКИ**

**Аннотация.** В статье представлен сравнительный анализ аппаратных архитектур и методов хранения голосовых данных в системах голосовой обработки, включая автоматическое распознавание речи, диалоговые системы и голосовую биометрию. Рассмотрены CPU-, GPU- и специализированные ускорители (TPU, FPGA, ASIC), а также файловые, распределённые, объектные и комбинированные модели хранения данных. Оценка проводилась по трём ключевым критериям: производительность, масштабируемость и экономическая эффективность. Показано, что оптимальный выбор архитектуры определяется требованиями конкретного сценария применения и согласованным сочетанием вычислительных и хранилищных решений. Работа позволяет формализовать подход к проектированию систем голосовой обработки и повышает практическую ценность их внедрения.

**Annotation.** This paper presents a comparative analysis of hardware architectures and voice data storage methods in voice processing systems, including automatic speech recognition, dialogue systems, and voice biometrics. CPU-, GPU-, and specialized accelerators (TPU, FPGA, ASIC) are considered,

along with file-based, distributed, object-based, and hybrid storage models. Evaluation is conducted according to three key criteria: performance, scalability, and cost-efficiency. It is shown that the optimal choice of architecture is determined by the specific application requirements and a coordinated combination of computational and storage solutions. The study provides a formalized approach to designing voice processing systems and enhances their practical value.

**Ключевые слова:** голосовая обработка, автоматическое распознавание речи, голосовая биометрия, аппаратные ускорители, CPU, GPU, TPU, FPGA, ASIC, хранение данных, объектные хранилища, файловые системы, масштабируемость, производительность, экономическая эффективность.

**Keywords:** voice processing, automatic speech recognition, voice biometrics, hardware accelerators, CPU, GPU, TPU, FPGA, ASIC, data storage, object storage, file systems, scalability, performance, cost-efficiency.

Современные технологии голосовой обработки — включая автоматическое распознавание речи, голосовые интерфейсы и голосовую биометрию — прочно входят в практику информационных систем. Глобальный рынок технологий распознавания речи и голоса в 2024 году оценивался почти в 19 млрд долларов США с ожидаемым многократным ростом к 2032 году, что отражает устойчивый интерес к этим технологиям в разнообразных приложениях, от корпоративных сервисов до пользовательских устройств [1, 2].

Голосовая биометрия применяется в задачах аутентификации и идентификации пользователей, повышая удобство и безопасность в банковской, телекоммуникационной и клиентской сферах [3].

Эффективность работы таких систем напрямую зависит от аппаратных ресурсов и методов хранения голосовых данных, которые обеспечивают скорость обработки, масштабируемость и надёжность, делая выбор архитектурных решений критическим этапом проектирования.

Существуют различные подходы к организации хранения голосовой информации: локальные и распределённые файловые системы, объектные хранилища и комбинированные решения, что требует их сравнительного анализа с учётом технических и экономических критериев. В научной литературе отсутствует комплексный обзор, систематически сопоставляющий аппаратные ресурсы и методы хранения голосовых данных в единой методологической рамке [4–10].

Цель настоящей работы — сравнительный анализ аппаратных ресурсов и методов хранения голосовых данных в системах голосовой обработки по критериям производительности, масштабируемости и экономической эффективности. Это позволяет выделить характерные особенности архитектурных решений и соотнести их с практическими требованиями современных информационных систем.

Системы голосовой обработки включают разнообразные приложения с различными требованиями к функциональности, объёму данных и условиям эксплуатации.

Автоматическое распознавание речи (ASR) преобразует устную речь в текст с использованием алгоритмов обработки сигналов и моделей машинного обучения [4]. Для таких систем критическими являются высокая производительность обработки и минимальные задержки доступа к данным, особенно в интерактивных интерфейсах и службах документооборота.

Диалоговые системы интегрируют ASR, обработку естественного языка и генерацию ответов. Они предъявляют повышенные требования к масштабируемости вычислительных и хранилищных подсистем, обеспечивая обслуживание большого числа параллельных сессий без снижения качества взаимодействия [4, 5].

Голосовая биометрия решает задачи идентификации и верификации пользователей на основе параметров голосового сигнала [3, 5]. Для этих систем важны высокая скорость доступа к данным и структурирование хранимых моделей, поскольку точность распознавания напрямую зависит от организации хранения и обработки аудиопризнаков [5].

Архивные и аналитические системы предназначены для долговременного хранения больших массивов голосовых данных и последующего анализа. Здесь ключевыми становятся масштабируемость хранения и экономическая эффективность, так как объёмы данных могут достигать терабайт и петабайт [4–10].

В совокупности, рассматриваемые области применения определяют различные приоритеты: ASR и интерактивные интерфейсы — производительность; диалоговые и биометрические системы — масштабируемость; архивные решения — экономическая эффективность, что обосновывает необходимость системного сравнительного анализа архитектурных решений.

Аппаратные ресурсы определяют производительность и масштабируемость систем голосовой обработки, включая задачи глубокого обучения и ASR. Различные архитектуры обеспечивают параллелизм вычислений, эффективность использования энергии и влияние на стоимость эксплуатации, что критично при выборе решений для разных сценариев [6–10].

Центральные процессоры (CPU) обеспечивают универсальность и гибкость интеграции, что делает их подходящими для небольших задач и прототипирования [6, 7]. CPU эффективно обрабатывают предварительные этапы аудиосигналов и управление вычислительными задачами, однако ограничены по масштабируемости и скорости инференса современных моделей глубокого обучения. Их преимущество — относительно низкая стоимость и простота сопровождения.

Графические процессоры (GPU) обеспечивают высокую производительность благодаря тысячам параллельных ядер, ускоряя обучение и инференс нейросетей [8, 9]. Гибридные решения, сочетающие CPU и GPU, позволяют распределять нагрузку между универсальными и специализированными вычислениями, повышая масштабируемость при работе с большим числом одновременных запросов. Они оптимальны для задач ASR, диалоговых систем и голосовой аналитики, где важна скорость обработки и пропускная способность [8, 9].

К специализированным ускорителям относятся TPU, FPGA и ASIC-решения [7–10]. TPU оптимизированы для конкретных моделей, повышая производительность инференса и энергоэффективность. FPGA позволяют настраивать аппаратные конфигурации под специфические алгоритмы, обеспечивая баланс между вычислительной мощностью и энергопотреблением. Специализированные ускорители особенно эффективны в масштабных облачных и промышленных системах, где важна совокупная экономическая эффективность при высокой нагрузке [7–10].

Таблица 1. Сравнительная оценка аппаратных архитектур систем голосовой обработки

Тип архитектуры	Производительность обработки аудиоданных	Масштабируемость вычислений	Экономическая эффективность	Итог
CPU	2	2	4	Подходит для прототипирования, предварительной обработки аудиоданных, небольших и средненагруженных систем голосовой обработки
GPU	5	4	3	Оптimalен для обучения и инференса нейросетевых моделей ASR, диалоговых систем и голосовой аналитики
TPU (ASIC)	5	4	4	Эффективен для масштабных облачных систем и промышленного инференса нейросетей с высокой нагрузкой

FPGA	4	3	4	Подходит для специализированных и энергоэффективных решений, ориентированных на конкретные алгоритмы обработки голоса
------	---	---	---	---

Краткое пояснение по шкале оценок

- 1 — крайне низкий уровень соответствия критерию
- 3 — средний, приемлемый уровень
- 5 — высокий уровень соответствия критерию

Голосовые данные характеризуются большим объёмом, потоковой природой и необходимостью хранения метаданных (временные метки, параметры записи, результаты обработки). Выбор модели хранения напрямую влияет на производительность системы, масштабируемость и экономическую эффективность [4–10].

Файловые системы и их распределённые аналоги обеспечивают простоту реализации и прямой локальный доступ к аудиофайлам [11, 12]. Горизонтальное расширение возможно через добавление узлов, но масштабируемость ограничена архитектурой и сложностью управления консистентностью данных [11, 12]. Экономическая эффективность высока для небольших и средних объёмов хранения.

Объектные хранилища предназначены для больших объёмов неструктурированных данных и сопровождаются расширенными

метаданными [10, 11]. Они обеспечивают масштабируемость за счёт распределённой архитектуры и позволяют эффективно работать с аналитикой и обучением моделей на больших наборах аудиоданных [10, 11]. Экономическая эффективность проявляется при долгосрочном хранении и облачных сценариях.

Комбинированные подходы разделяют хранение аудиофайлов и метаданных между различными системами — объектными или файловыми для данных и реляционными/NoSQL для метаданных [10–16]. Это обеспечивает гибкость архитектуры, быстрый поиск и фильтрацию по метаданным, сохраняя масштабируемость и высокую производительность обработки данных. Несмотря на более сложное сопровождение, такой подход оптимален для промышленных систем с высокой нагрузкой.

Таблица 2. Сравнительная оценка методов хранения голосовых данных

Метод хранения	Скорость доступа и работы с данным и	Масштабируемость хранения	Экономическая эффективность эксплуатации	Итог
Файловые и распределённые файловые системы	4	3	3	Подходит для архивных систем, офлайн-анализа и локальных решений с ограниченным объёмом

				ГОЛОСОВЫХ данных
Объектные хранилища	3	5	4	Оптимальны для масштабируем ых облачных систем, долгосрочного хранения и аналитики больших массивов аудиоданных
Комбинированн ые подходы (данные + метаданные)	5	4	4	Наиболее эффективны для промышленных систем голосовой обработки с высокой нагрузкой и требованиями к быстрому поиску

В системах голосовой обработки эффективность вычислительных и хранилищных компонентов определяется их согласованной работой. Целесообразно рассматривать архитектурные решения как типовые связки «метод хранения + аппаратная архитектура», оценивая компромиссы между

производительностью, масштабируемостью и экономической эффективностью [6–16].

Файловые и распределённые файловые системы наиболее эффективно работают с CPU-ориентированными архитектурами [11, 12]. Такая связка обеспечивает прямой и предсказуемый доступ к аудиофайлам, достаточную производительность для офлайн-анализа и архивных сценариев, а также экономичность за счёт низкой стоимости оборудования и простоты сопровождения. Использование GPU с файловыми системами возможно, но обычно не позволяет полностью раскрыть потенциал ускорителей [8, 9].

Объектные хранилища оптимальны в сочетании с GPU-ускоренными архитектурами [10, 11]. Распределённая модель хранения и поддержка параллельного доступа обеспечивают стабильную загрузку графических ускорителей, высокую масштабируемость по объёму данных и вычислительным ресурсам. Такая связка подходит для систем автоматического распознавания речи, диалоговых платформ и аналитических сервисов, где важны скорость обработки и возможность параллельного анализа больших массивов данных [10, 11, 16].

Комбинированные подходы к хранению данных и метаданных демонстрируют максимальную гибкость и применимы в архитектурах с GPU и специализированными ускорителями (TPU, FPGA) [10–16]. Раздельное хранение аудиофайлов и метаданных позволяет оптимизировать доступ и распределять вычислительную нагрузку между слоями системы. Эти связки обеспечивают высокую производительность и масштабируемость и наиболее подходят для промышленных систем голосовой биометрии и интеллектуальной аналитики, требующих быстрой обработки больших объёмов данных [10–16].

Таким образом, выбор архитектурного решения определяется не универсальными технологиями, а согласованным подбором аппаратных ресурсов и методов хранения в зависимости от требований конкретного сценария [6–16]. Применение единых критериев анализа позволяет формализовать выбор и обосновать архитектурные решения с научной точки зрения [6–16].

В статье проведён сравнительный анализ аппаратных архитектур и методов хранения голосовых данных в системах голосовой обработки с использованием единых критериев: производительность, масштабируемость и экономическая эффективность [6–16]. Такой подход позволил обеспечить целостность и сопоставимость результатов.

Показано, что универсального решения для всех классов систем не существует. Выбор архитектуры всегда является компромиссом между вычислительной мощностью, масштабируемостью и затратами, а оптимальность определяется соответствием требованиям конкретного сценария [6–16].

Наибольшую производительность обеспечивают GPU и специализированные ускорители в сочетании с масштабируемыми методами хранения, что критично для ASR, голосовой аналитики и биометрии [8–11]. CPU-ориентированные решения подходят для офлайн-обработки и архивных систем с низкой нагрузкой [6, 7].

По масштабируемости лидируют объектные и комбинированные хранилища, позволяющие независимо расширять вычислительные и дисковые ресурсы, тогда как файловые системы ограничены и требуют дополнительного управления при росте нагрузки [10–16].

С точки зрения экономической эффективности простые решения уступают по производительности и масштабируемости более сложным архитектурам, но высокопроизводительные системы оправдывают свою стоимость при высокой интенсивности обработки [10–16]. Комбинированные подходы обеспечивают оптимальное соотношение затрат и эффективности при корректной формализации требований [10–16].

Таким образом, ключевым фактором проектирования является точное определение целевых требований и использование единой методики анализа, что позволяет обоснованно выбирать архитектурные решения и повышает практическую ценность систем голосовой обработки [6–16].

### **Список использованной литературы**

1. Kings Research. Speech and Voice Recognition Market Size, Share, Growth & Industry Analysis, By Technology, Deployment, Vertical, and Regional Forecast, 2025-2032. — Jul. 2025. — Режим доступа: <https://www.kingsresearch.com/report/speech-and-voice-recognition-market-2521> (дата обращения: 19.12.2025).
2. Fortune Business Insights™. Speech and Voice Recognition Market Size, Share & Industry Analysis, By Technology, By Deployment, By End-user, and Regional Forecast, 2025-2032. — 2025. — Режим доступа: <https://www.fortunebusinessinsights.com/speech-and-voice-recognition-market-101382> (дата обращения: 19.12.2025).
3. Поваложская А. А., Карпов А. А. Аналитический обзор методов автоматического анализа спонтанной речи в системах ASR // Информатика и автоматизация. 2024. № 1. С. 5–38. DOI:10.15622/ia.2024.23.1.

4. Стефаниди А. Ф. Исследование мультимодальных алгоритмов биометрической идентификации на основе методов цифровой обработки речевых сигналов и изображений: автореф. дис. ... наук. — ВлГУ, 2022.
5. Alam S., et al. Review on FPGA-Based Accelerators in Deep Learning // Journal of Computer Engineering and Applications. 2024. Vol. 13, No. 15. P. 2988. DOI:10.3390/electronics13152988.
6. Silvano C., Ielmini D., Ferrandi F., et al. A Survey on Deep Learning Hardware Accelerators for Heterogeneous HPC Platforms // ACM Computing Surveys. 2025. — Режим доступа: <https://arxiv.org/abs/2306.15552> (дата обращения: 20.12.2025).
7. Mittal S., Umesh S. A survey on hardware accelerators and optimization techniques for RNNs // Journal of Systems Architecture. 2021. Vol. 112:101839. DOI:10.1016/j.sysarc.2020.101839.
8. Shipitsin S. P. Развитие аппаратно-ориентированных нейронных сетей на FPGA и ASIC // Computer Science Review, 2019. — Режим доступа: <https://cyberleninka.ru/article/n/razvitie-apparatno-orientirovannyh-neyronnyh-setey-na-fpga-i-asic> (дата обращения: 20.12.2025).
9. TPU (Tensor Processing Unit) // Google Cloud TPU Documentation. — Режим доступа: <https://cloud.google.com/tpu/docs> (дата обращения: 20.12.2025).
10. Zhang Q., Chen M., Li L. Big Data Storage Technologies: A Survey // IEEE Access. 2018. Vol. 6. P. 42440–42457. DOI:10.1109/ACCESS.2018.2855859.
11. Ji C., Li Y., Qiu W., Awada U., Li K. Big Data Processing in Cloud Computing Environments // IEEE Transactions on Big Data. 2019. Vol. 5, No. 1. P. 17–33. DOI:10.1109/TBDDATA.2016.2599878.
12. Shvachko K., Kuang H., Radia S., Chansler R. The Hadoop Distributed File System // Proceedings of the 2010 IEEE 26th Symposium

- on Mass Storage Systems and Technologies (MSST). IEEE, 2010. P. 1–10. DOI:10.1109/MSST.2010.5496972.
13. Ghemawat S., Gobiuff H., Leung S.-T. The Google File System // ACM SIGOPS Operating Systems Review. 2003. Vol. 37, No. 5. P. 29–43. DOI:10.1145/945445.945450.
  14. Gibson G., Van Meter R. Network Attached Storage Architecture // Communications of the ACM. 2000. Vol. 43, No. 11. P. 37–45. DOI:10.1145/353360.353363.
  15. Stonebraker M., Abadi D., DeWitt D., et al. MapReduce and Parallel DBMSs: Friends or Foes? // Communications of the ACM. 2010. Vol. 53, No. 1. P. 64–71. DOI:10.1145/1629175.1629197.
  16. SiddiqA et al. Big data storage technologies: a survey // Frontiers of Information Technology & Electronic Engineering. — FITEE.

## References

1. Kings Research. *Speech and Voice Recognition Market Size, Share, Growth & Industry Analysis, By Technology, Deployment, Vertical, and Regional Forecast, 2025-2032*. — Jul. 2025. — Accessed at: <https://www.kingsresearch.com/report/speech-and-voice-recognition-market-2521> (accessed: 19.12.2025).
2. Fortune Business Insights™. *Speech and Voice Recognition Market Size, Share & Industry Analysis, By Technology, By Deployment, By End-user, and Regional Forecast, 2025-2032*. — 2025. — Accessed at: <https://www.fortunebusinessinsights.com/speech-and-voice-recognition-market-101382> (accessed: 19.12.2025).
3. Povolozkaya A. A., Karpov A. A. Analytical Review of Methods for Automatic Analysis of Spontaneous Speech in ASR Systems // *Informatics and Automation*. 2024. No. 1. P. 5–38. DOI:10.15622/ia.2024.23.1.

4. Stefanidi A. F. *Study of Multimodal Biometric Identification Algorithms Based on Digital Processing Methods of Speech Signals and Images: Abstract of Dissertation ... Sciences.* — VLGU, 2022.
5. Alam S., et al. Review on FPGA-Based Accelerators in Deep Learning // *Journal of Computer Engineering and Applications.* 2024. Vol. 13, No. 15. P. 2988. DOI:10.3390/electronics13152988.
6. Silvano C., Ielmini D., Ferrandi F., et al. A Survey on Deep Learning Hardware Accelerators for Heterogeneous HPC Platforms // *ACM Computing Surveys.* 2025. — Accessed at: <https://arxiv.org/abs/2306.15552> (accessed: 20.12.2025).
7. Mittal S., Umesh S. A Survey on Hardware Accelerators and Optimization Techniques for RNNs // *Journal of Systems Architecture.* 2021. Vol. 112:101839. DOI:10.1016/j.sysarc.2020.101839.
8. Shipitsin S. P. Development of Hardware-Oriented Neural Networks on FPGA and ASIC // *Computer Science Review,*2019. — Accessed at: <https://cyberleninka.ru/article/n/razvitie-apparatno-orientirovannyh-neyronnyh-setey-na-fpga-i-asic> (accessed: 20.12.2025).
9. TPU (Tensor Processing Unit) // *Google Cloud TPU Documentation.* — Accessed at: <https://cloud.google.com/tpu/docs> (accessed: 20.12.2025).
10. Zhang Q., Chen M., Li L. Big Data Storage Technologies: A Survey // *IEEE Access.* 2018. Vol. 6. P. 42440–42457. DOI:10.1109/ACCESS.2018.2855859.
11. Ji C., Li Y., Qiu W., Awada U., Li K. Big Data Processing in Cloud Computing Environments // *IEEE Transactions on Big Data.* 2019. Vol. 5, No. 1. P. 17–33. DOI:10.1109/TBDDATA.2016.2599878.
12. Shvachko K., Kuang H., Radia S., Chansler R. The Hadoop Distributed File System // *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST).* IEEE, 2010. P. 1–10. DOI:10.1109/MSST.2010.5496972.

13. Ghemawat S., Gobioff H., Leung S.-T. The Google File System // *ACM SIGOPS Operating Systems Review*. 2003. Vol. 37, No. 5. P. 29–43. DOI:10.1145/945445.945450.
14. Gibson G., Van Meter R. Network Attached Storage Architecture // *Communications of the ACM*. 2000. Vol. 43, No. 11. P. 37–45. DOI:10.1145/353360.353363.
15. Stonebraker M., Abadi D., DeWitt D., et al. MapReduce and Parallel DBMSs: Friends or Foes? // *Communications of the ACM*. 2010. Vol. 53, No. 1. P. 64–71. DOI:10.1145/1629175.1629197.
16. Siddiq A et al. Big Data Storage Technologies: A Survey // *Frontiers of Information Technology & Electronic Engineering*.— FITEE.