

## РЕАЛИЗАЦИЯ МНОГОЗАДАЧНОЙ СИСТЕМЫ СИНТЕЗА РЕЧИ РЕАЛЬНОГО ВРЕМЕНИ

Кондратьев А.Ю. (ООО «ЦРТ-Инновации»)

Научный руководитель – к.ф.-м.н., Рыбин С.В. (Университет ИТМО)

Представленная работа посвящена разработке информационной системе многоканального синтеза речи из текста на основе нейронных сетей, удовлетворяющая всем нефункциональным требованиям прикладных систем.

Автоматический синтез речи является одной из основных задач человеко-компьютерного взаимодействия. Синтез может применяться во многих прикладных задачах и автоматизированных системах, например в качестве части голосового ассистента. В настоящее время, задача синтеза речи успешно решается при помощи глубоких нейронных сетей, которые предоставляют высокое качество выходного звука. В тоже время, к недостатку данного метода относится высокая вычислительная сложность и долгое время это было одной из причин отсутствия массового промышленного внедрения систем на основе данного метода. Сейчас подобные системы активно внедряются и идут активные работы по снижению алгоритмической сложности синтеза речи, а также построению эффективных систем реального времени.

В задаче синтеза речи выделяется два основных этапа – преобразование текста или звуковых фонем в промежуточное представление в виде спектрограммы сигнала, а также восстановление по спектрограмме речевого сигнала. Обе подзадачи решают нейронные сети, которые могут быть как авторегрессионные, при которых следующий отсчет зависит от предыдущего, так и параллельные, в которых синтез всех отсчетов спектра или сигнала осуществляется одновременно. В настоящее время, преобладают авторегрессионные алгоритмы. К их преимуществу, помимо качества звучания, относится то, что первоначальный результат в них получается с минимальной задержкой и может быть использован на следующем этапе. Это позволяет организовать конвейер с минимальным временем отклика при длительном общем времени работы.

В данной работе предлагается асинхронная система вычислений длительных абстрактных задач, направленная на максимальную утилизацию вычислительных ресурсов. Таким образом, задачу можно рассмотреть, как несколько этапов взаимосвязанных производителей и потребителей. Для отсутствия больших задержек и разрывов в синтезе речи используется общий балансировщик нагрузки, который точно определяет, у какой подзадачи наибольший приоритет на выполнение в данный момент времени.

В работе продемонстрировано, что общая производительность подобной системы почти в два раза превышает производительность классической реализации, в которой вычисления производятся конкурентно. Это позволяет подавать на вычислительный узел вдвое больше запросов на одновременный синтез речи, что значительно влияет на конечную стоимость внедрения и эксплуатации системы синтеза речи.