

Отзыв научного руководителя на выпускную квалификационную работу  
Михайловского Дмитрия Владимировича  
на тему:  
«End2End моделирование голоса»

Направление: 01.03.01 «Математика»

Основная образовательная программа: СВ.5000.2018 «Математика»

В своей работе Дмитрий Владимирович обращается к теме синтеза голоса. В связи с развитием голосовых интерфейсов, используемых сегодня во многих областях (от голосовых ассистентов и роботов в телефонии, до ведения медицинских записей), данная область представляется актуальной. Несмотря на долгую историю развития данных технологий, задача синтеза до сих пор не решена. Одной из важных проблем сегодняшнего дня является создание голосовых моделей, способных обеспечить качественный синтез голоса при существенных ограничениях на вычислительные ресурсы, например, в мобильных, или встроенных устройствах. Одной из наиболее сложных задач с этой точки зрения является задача «сшивания» фрагментов синтезированного голоса в последовательную связную речь, т.н. вокодер. На данный момент задачу «сшивания» принято решать с помощью машинного обучения, и полученные таким способом модели являются узким местом в системах синтеза голоса в условиях ограничений на вычислительные ресурсы. В работе проведен анализ альтернативных путей решения данной проблемы и предложен способ комплексного подхода к решению этой задачи.

Основная идея работы заключается в том, чтобы воспользоваться обратимостью преобразования Фурье, и по предсказанным спектрограммам – задача, которая решается множеством существующих моделей – восстановить сигнал. Для этого необходимо кроме восстановления амплитуд восстановить также и фазовые сдвиги. Именно вокруг этого аспекта и сосредоточена большая часть работы. В работе исследованы инварианты в пространстве фазовых сдвигов с точки зрения человеческого восприятия. Важным новым знанием стало то, что случайные фазовые сдвиги по разным частотам не влияют на восприятие. На основе этого факта были выдвинуты и проверены несколько гипотез о способах восстановления фазовых сдвигов, один из которых показал высокое качество восстановления сигнала на оригинальных спектрограммах голоса. К сожалению, на предсказанных спектрограммах результат хотя и превосходит существующие аналоги алгоритмического восстановления фазовых сдвигов, не достаточно хорош для практического применения. Тем не менее я ожидаю, что результаты будут существенно улучшены через адаптацию процедуры генерации спектрограммы для алгоритмического восстановления сигнала.

К недостаткам данной работы приходится отнести ее оформление. В связи со сжатыми

сроками подготовки текста жертвой пали как чистота изложения, так и оформление экспериментов. Тем не менее, работу можно прочитать и вынести из нее тезисы главных результатов.

Несмотря на перечисленные недостатки, считаю, что работа на тему «End2End моделирование голоса» является законченным научным исследованием и соответствует требованиям университета, предъявляемым к бакалаврским диссертациям, а её автор, Михайловский Дмитрий Владимирович, заслуживает оценки «отлично» и рекомендации к поступлению в магистратуру.



А.Ю. Авдюшенко,  
доцент, к.ф.-м.н.

Дата: 7 июня 2022 г.