Разработка системы синтеза эмоциональной речи

Ян Якубчик

Дмитрий Климов

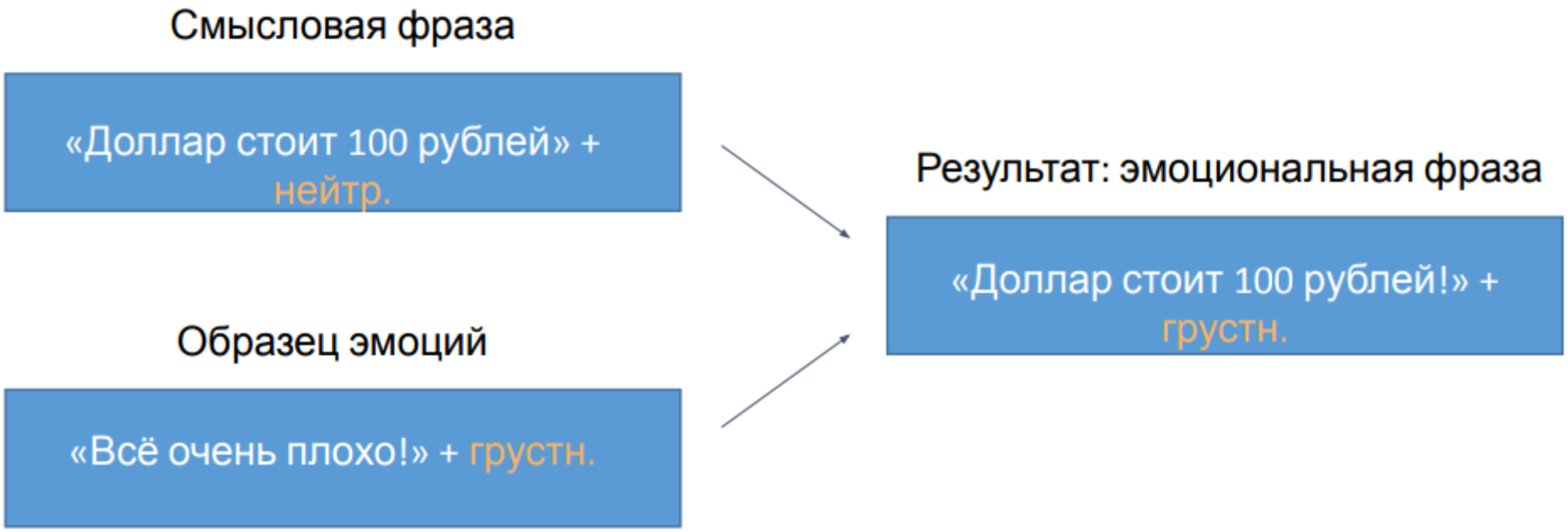
Команда ЭМОТЕХ

Проблема

- Роботы могут разговаривать с помощью синтезаторов речи
  - Фразы произносятся каждый раз одинаково
  - Нет эмоций
- Люди передают значительную часть информации в эмоциях
- В распознавании эмоций учёные достигли больших успехов
  - Искусственные нейронные сети
  - Гильбертовы контуры
- Синтез эмоциональной речи – непростая задача
- Необходимо:
  - Разработка системы изменения эмоций в звуковом сигнале
  - Дополнение к синтезатору речи

Предлагаемый подход

Переносим эмоции с фразы-источника на фразу-назначение (прямое обучение по аналогии)
Для переноса эмоций мы переносим интонацию и ритм, сохраняя форманты.



Модель речевого звука «источник-фильтр» [1][2]

[1] Chiba, T.; Kajiyama, M. (1942). The Vowel: Its Nature and Structure. Tokyo: Tokyo-Kaiseikan Pub. Co., Ltd.
[2] Fant, G. (1960) Acoustic theory of speech production, The Hague, The Netherlands, Mouton. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

- Источник *h(t)* (голосовые связки+шум)*фильтр *x(t)* (речевой тракт)

$$y(t) = h(t) * x(t) \quad (1)$$

(* означает свёртку)

- Спектр Фурье: $Y(\omega) = H(\omega) * X(\omega)$ (2) (* означает умножение)

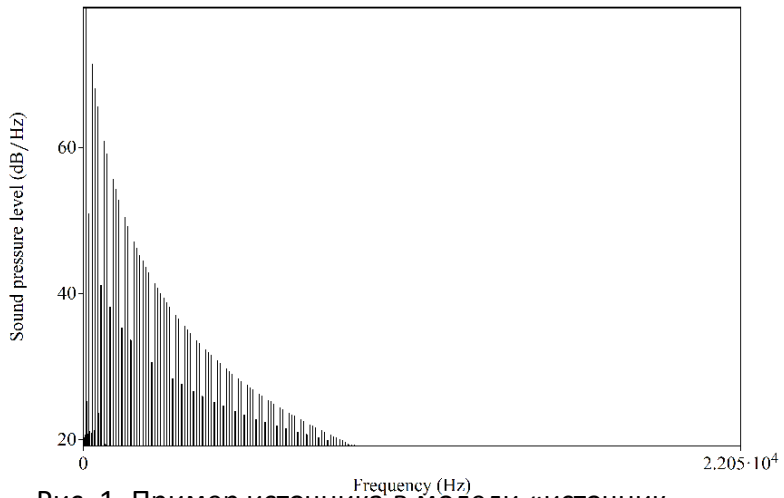


Рис. 1. Пример источника в модели «источник-фильтр». Основной тон равен 100 Гц. Звук не содержит шума, что нормально для гласных звуков.

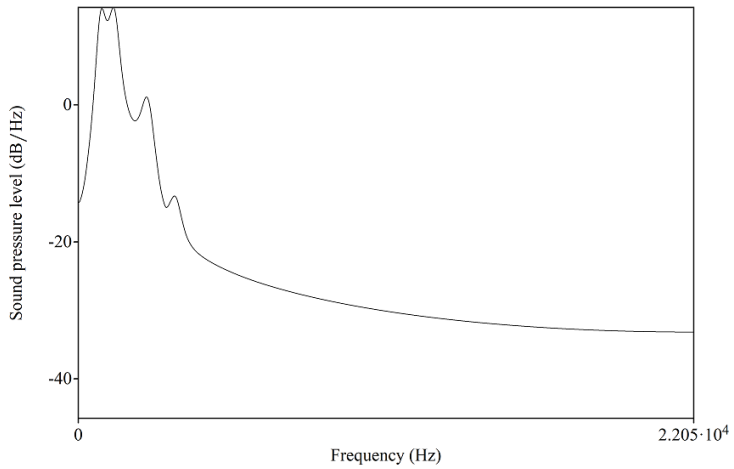


Рис. 2. Пример фильтра. Первые две форманты равны 800 Гц и 1300 Гц, что соответствует гласной [a].

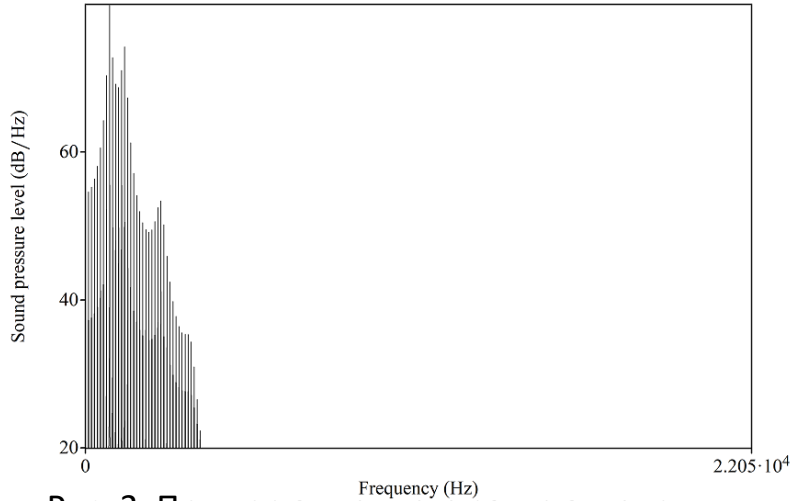


Рис. 3. Пример выходного звука в модели «источник-фильтр». Спектр содержит гармоники источника, усиленные вокруг формант.

Спектрограмма и форманты

- Каждый спектральный срез подчиняется модели «источник-фильтр»

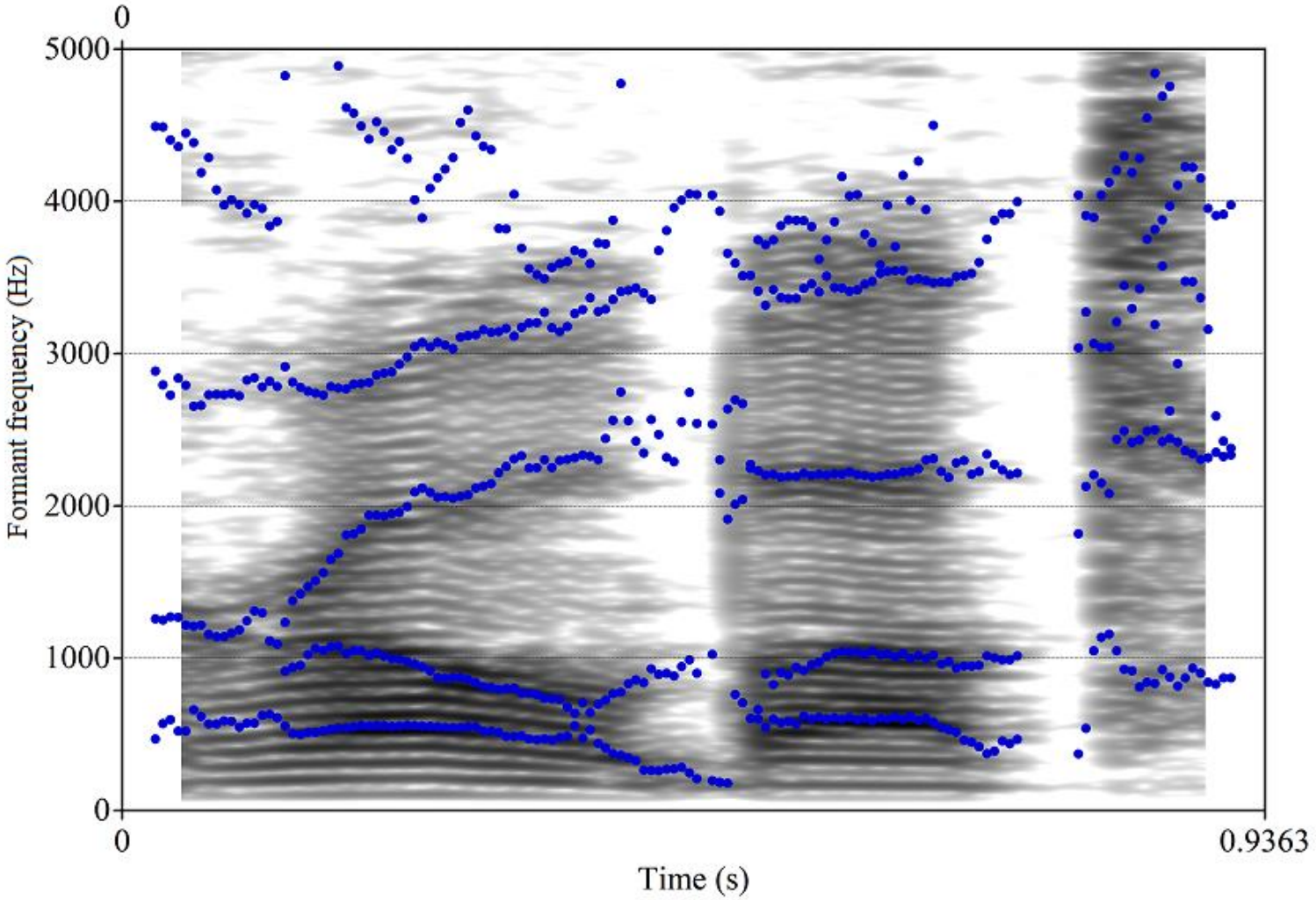


Рис. 4. Спектрограмма и форманты записи слова «робот».

Методы анализа и синтеза речи

- Метод автокорелляции для определения высоты звука (Boersma 2000) [3]
- Метод Бург для определения формант (Anderson 1978) [4]
- Метод Overlap-add для изменения интонации и ритма (Moulines&Charpentier 1990) [5]

[3] Paul Boersma “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.” *Proceedings of the Institute of Phonetic Sciences* 17:97-110. University of Amsterdam.
[4] N. Anderson (1978): “On the calculation of filter coefficients for maximum entropy spectral analysis.” In Childers: *Modern Spectrum Analysis*, IEEE Press:252-255.

DTW (dynamic time warping)
Пример соответствия между двумя звуками

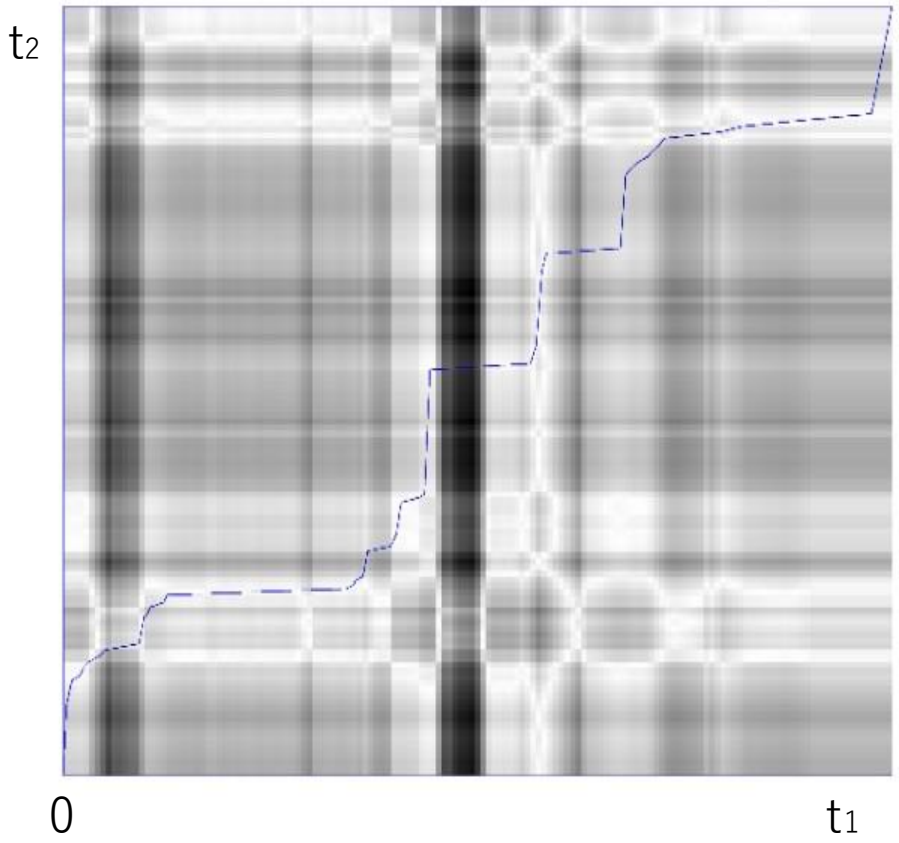


Рис. 5. Пример ритмического соответствия между двумя звуками

Результаты

- С помощью автоматического поиска соответствий методом DTW мы успешно перенесли эмоции между одинаковыми фразами, сказанной разными голосами
- На данном этапе разрабатывается перенос эмоций между разными фразами
- Разрабатывается модель языка
    - Некоторые части интонации показывают эмоции, некоторые части указывают содержание (ударение, вопросительная и восклицательная интонация, смыслоразличительная функция в тоновых языках), сильно зависит от языка

Закономерности интонации русских слов

Интонация зависит от положения относительно ударения

Может быть понижение на ударном слоге или после него

Может быть плавное движение интонации перед ударением и после ударения

Рис. 19. Модель интонации русского слова

Темп перед, на и после ударения может быть разным

Предлагается модель, основанная на 6 моментах времени, 6 частотах и 3 темпах

Высота перепадов может зависеть от эмоций

Перспективы коммерциализации

- Динамичное развитие сервисных роботов требует от человекоподобных роботов способности коммуникации с людьми. Мы собираемся выходить на мировой рынок с данной системой синтеза эмоциональной речи.
- На данный момент мы ведём переговоры с компанией Промобот в Перми по поводу покупки данной системы.

Бизнес-модель

ЭМОТЕХ

Разрабатывает проект системы вывода эмоциональной речи, состоящей из данного программного обеспечения (дополнения к существующему синтезатору речи) и плана интеграции с синтезатором речи и роботом в целом, оформляет патент и продает проектную документацию компаниям производителям

→

Промобот

Устанавливает систему на своих роботов и платит ЭМОТЕХу

100-200 $ с каждого робота

При продаже 1000 роботов в год выручка 100 000$ в год

→

→

Другие компании

Устанавливают систему на своих роботов и платят ЭМОТЕХу

100-200 $ с каждого робота

→

Конкурентный подход

Показатель	EVS	ATR CHATR	DAVID	Tinkoff
Языки	Японский	Японский	любой	Русский
Количество эмоций	6	3	3	4
Платформа	Sony PlayStation	ПК клиента	ПК клиента	Сервер
Качество	высокое	Среднее	Низкое	Высокое
Вычислительная сложность	Низкая	Средняя	Низкая	Высокая
Объём базы данных	Большой	Большой	Маленький	Большой
Метод	Компиляционный	Компиляционный	Параметрическая обработка	ИИ

Таблица Существующие решения в области синтеза эмоциональной речи

Применение

АДМИНИСТРАЦИЯ ГУБЕРНАТОРА
ПЕРМСКОГО КРАЯ

ул. Куйбышева, д. 14, г. Пермь, 614006,
тел. (342) 253 70 69, 253 79 69,
факс (342) 253 66 49, www.admin.permkrai.ru
e-mail: obladm@permkrai.ru
ОКПО 78884744, ОГРН 1055900364156,
ИНН/КПП 5902293202/590201001

30.09.2021 № 02-10-12

На № б/и от 04.10.2021

Руководителю проекта «Робот-социолог»

Д.А. Тиунову

Уважаемый Дмитрий Александрович!

Департамент общественных проектов Администрации губернатора Пермского края заинтересован в развитии проекта «Робот-социолог» с использованием системы голосования Ярейтинг и робота «Промобот». Научно-технические разработки и программные решения позволят поднять социологические опросы на новый уровень. Также интересны психологические и социологические исследования в этой области, подтверждающие положительную реакцию респондентов на использование роботов для проведения опросов. Считаем, что работа над данным проектом позволит довести его, в том числе, до уровня коммерческого использования. Со своей стороны, мы также готовы использовать его для получения корректных данных при проведении социологических опросов.

Директор департамента
общественных проектов

Е.И. Хузин

14.06.2022 20:44 Примерное время чтения: меньше минуты

Эмоциональный голос для роботов создали пермские учёные

правительство Пермского края

Пермь, 14 июня - АиФ-Прикамье.

Рынки

Рынок сервисных роботов стремительно растёт

		2023 г	2026 г
объем рынка	млн. $	3952.50	9630.00
количество	шт.	200 000	500 000

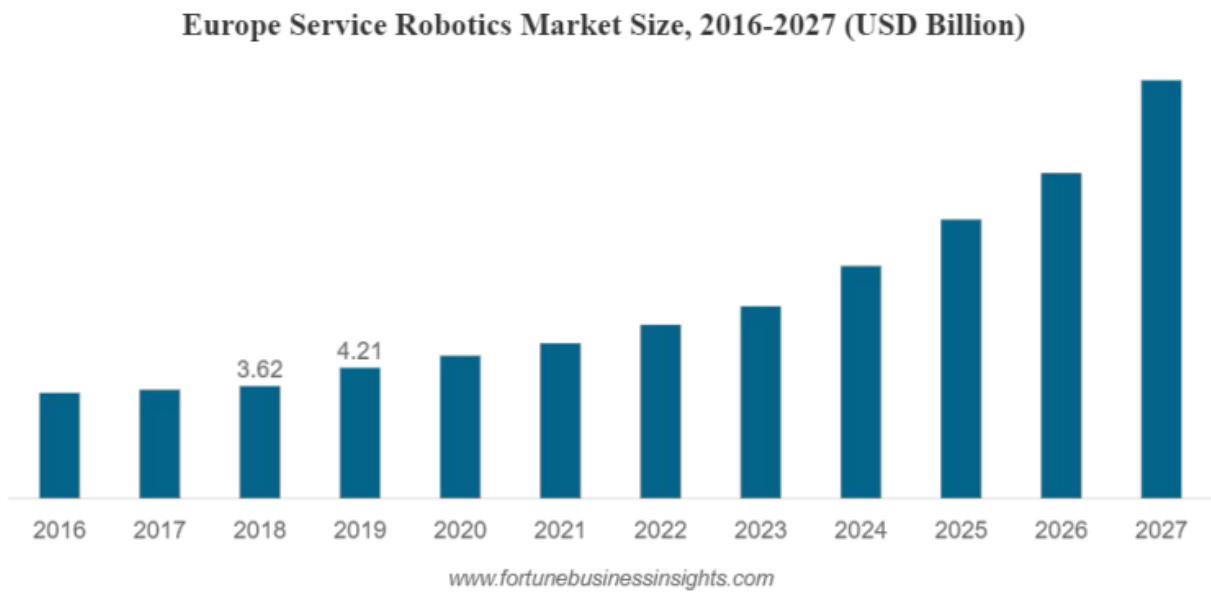


Рис.19. Объем рынка человекоподобных роботов

Рынки

- Данный продукт наиболее актуален для человекоподобных роботов, работающих с людьми. Однако он может быть интересен и для нечеловекоподобных роботов или для программ и онлайн-систем, не имеющих физического тела, но общающихся с людьми голосом, для создания эмоциональных реплик в фильмах и других целей. На диаграмме показаны различные сценарии продаж, при которых ЭМОТЕХ захватывает от 1 до 5% доли рынка (в количестве роботов и проданных экземплярах нашей системы)

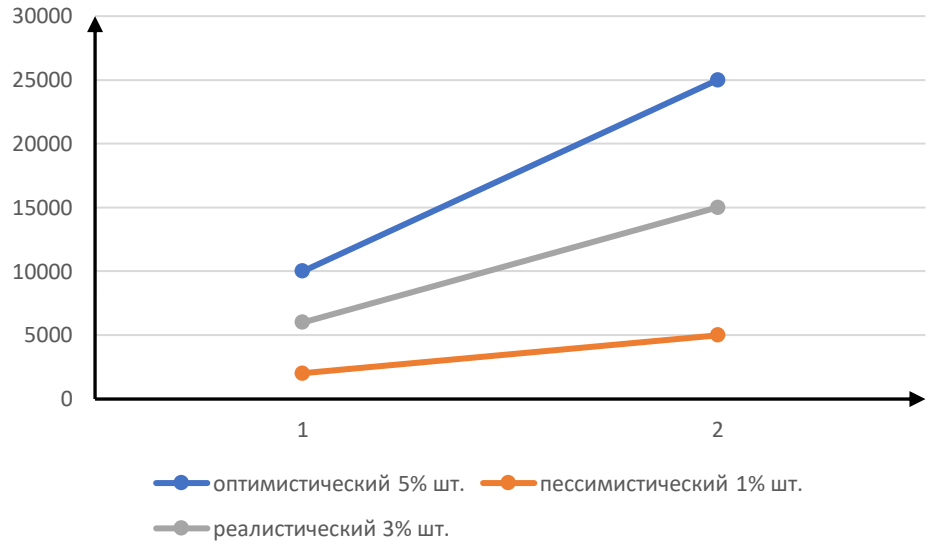


Рис.20. Объем рынка продаж системы эмоциональной речи, шт.

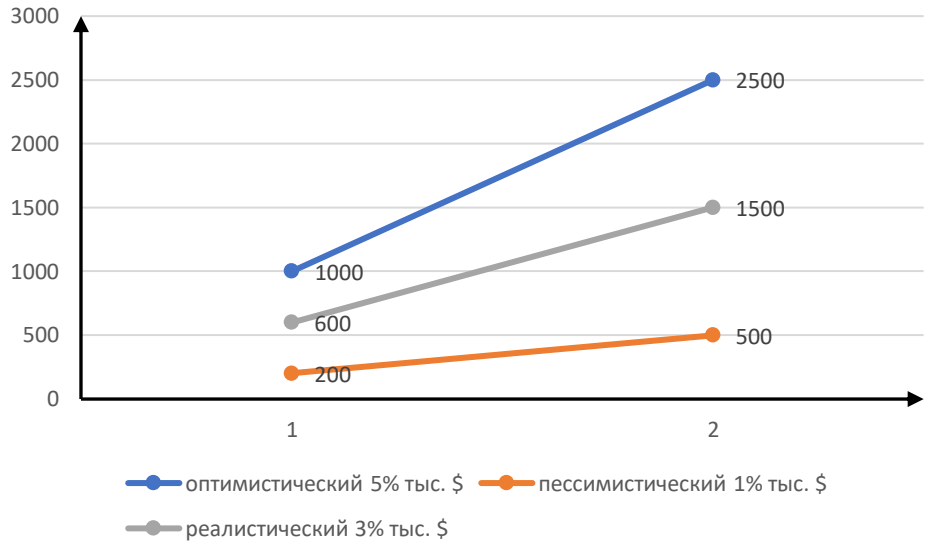


Рис.21. Объем рынка продаж системы эмоциональной речи, тыс.$

НОВОСТИ

- Исследование начато Яном Якубчиком во время обучения в магистратуре ПНИПУ (Пермского нациоального исследовательского политехнического университета) по специальности Автономные сервисные роботы в лаборатории кафедры Автоматики и телемеханики на роботах Промобот, предоставленных компанией Промобот для исследований
- Главный результат: научная статья
  Ian Iakubchik, Anna Iakubchik, Yuri Lipin “Emotional Voice Reconstruction Based on Intonation Alteration”, ElConRus, IEEE, 2021.
  https://ieeexplore.ieee.org/abstract/document/9396450
- В 2021-м году получен получен грант УМНИК, выполнен первый этап НИР по гранту (исследованы признаки аудиозаписи, отвечающие за эмоции и за содержание, разработан алгоритм переноса эмоций между различными фразами), выполняется второй этап (разработка образца ПО и испытания; работы близятся к завершению). Планируется преакселерация.

Команда

Ян Якубчик – лидер команды и научный руководитель. Образование: 2015-2019 Йокогамский Государственный Университет, информационный инженер, бакалавр. 2019-2021, Токийский университет, механо-информатика, магистр информатики и технологии, 2020-2022, ПНИПУ (Пермский Национальный Исследовательский Политехнический Университет), специальность Автономные сервисные роботы, магистр. Является аспирантом Токийского университета по теме робототехники. Публикации: "Implant Email Attack That Does Not Require URL Access by Target User", Proceedings of Computer Security Symposium, 2019. "Continuum Model Simulation and Congestion Reduction Control of Four-directional Pedestrian Flows assuming Scramble Crossing", DARS-SWARM, 2021. ElConRus, IEEE, 2021, "Emotional Voice Reconstruction Based on Intonation Alteration ", ElConRus, IEEE, 2021 "Acoustic Determination of Contact on the Exterior Surface of the Robot". https://ieeexplore.ieee.org/abstract/document/9396486
https://ieeexplore.ieee.org/abstract/document/9396450

Помогают Дмитрий Климов, Борис Погорелый