

Логико-математические

ОСНОВЫ

компьютерной лингвистики

и всей жизни ~~в~~ ~~техносфере~~

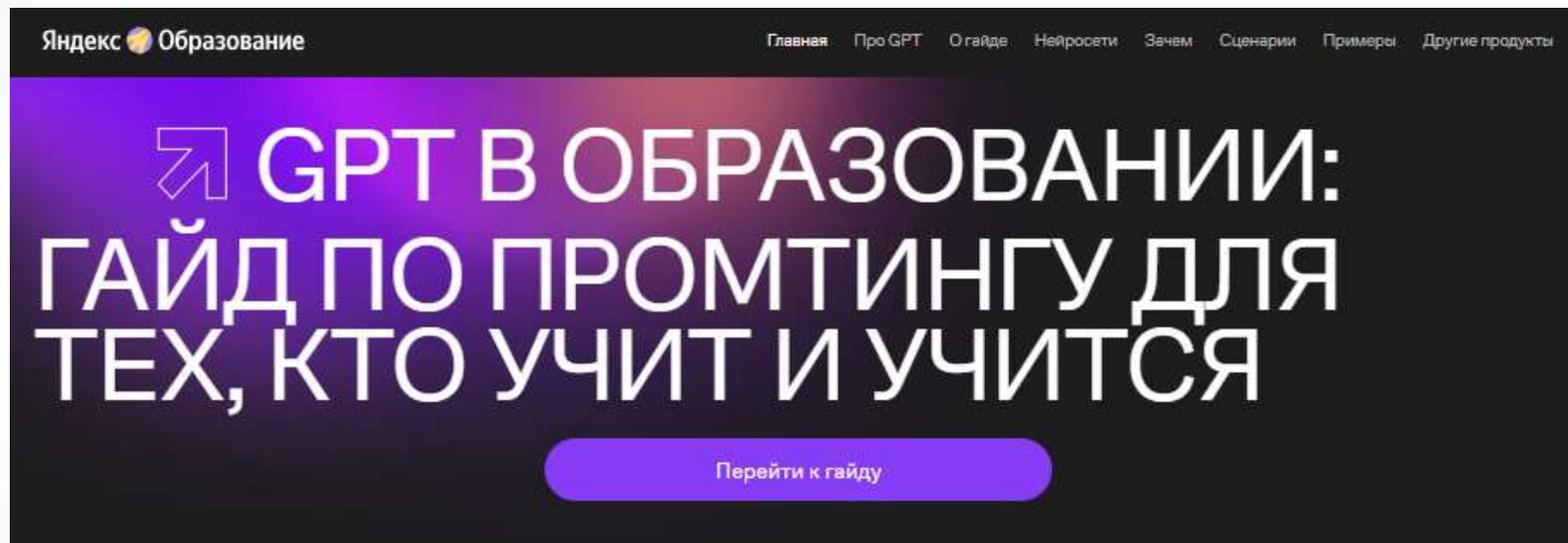
Давид Бэссис

- Цель этой книги – изменить ваш взгляд на мир.
- Математика, это странствие, которое поныне преобразует наши язык и мышление.
- Этот опыт мало похож на то, чему нас учат в школе...

Давид Бэссис

- Математика придает нашему миру форму. Она – инструмент власти и господства.
- Но для тех, кто живет ею, постижение математики – это прежде всего внутренний опыт, чувственный и духовный поиск.

GIGO



Яндекс Образование

Главная Про GPT О гайде Нейросети Зачем Сценарии Примеры Другие продукты

➤ GPT В ОБРАЗОВАНИИ:
ГАЙД ПО ПРОМТИНГУ ДЛЯ
ТЕХ, КТО УЧИТ И УЧИТСЯ

Перейти к гайду

Garbage in, garbage out

ТЕОРИЯ ПОРОЖДЕНИЯ РЕЧИ

ОБЪЯВЛЕНИЕ

получить логины/пароли
для работы в интернете
308 аудитория корп. №1

НОВОСТИ

практики порождения речи
и не только...

БЫСТРЕЕ!

Reading now becomes **faster**

Upload the book, AI will get rid of unnecessary parts and make you read only that's matter.

Login

Sign up here →

ГЛУБЖЕ?

Reading now becomes deeper

Upload the book, AI will get rid of unnecessary parts and make you read only that's matter.

Login

Sign up here →

YMHEE?

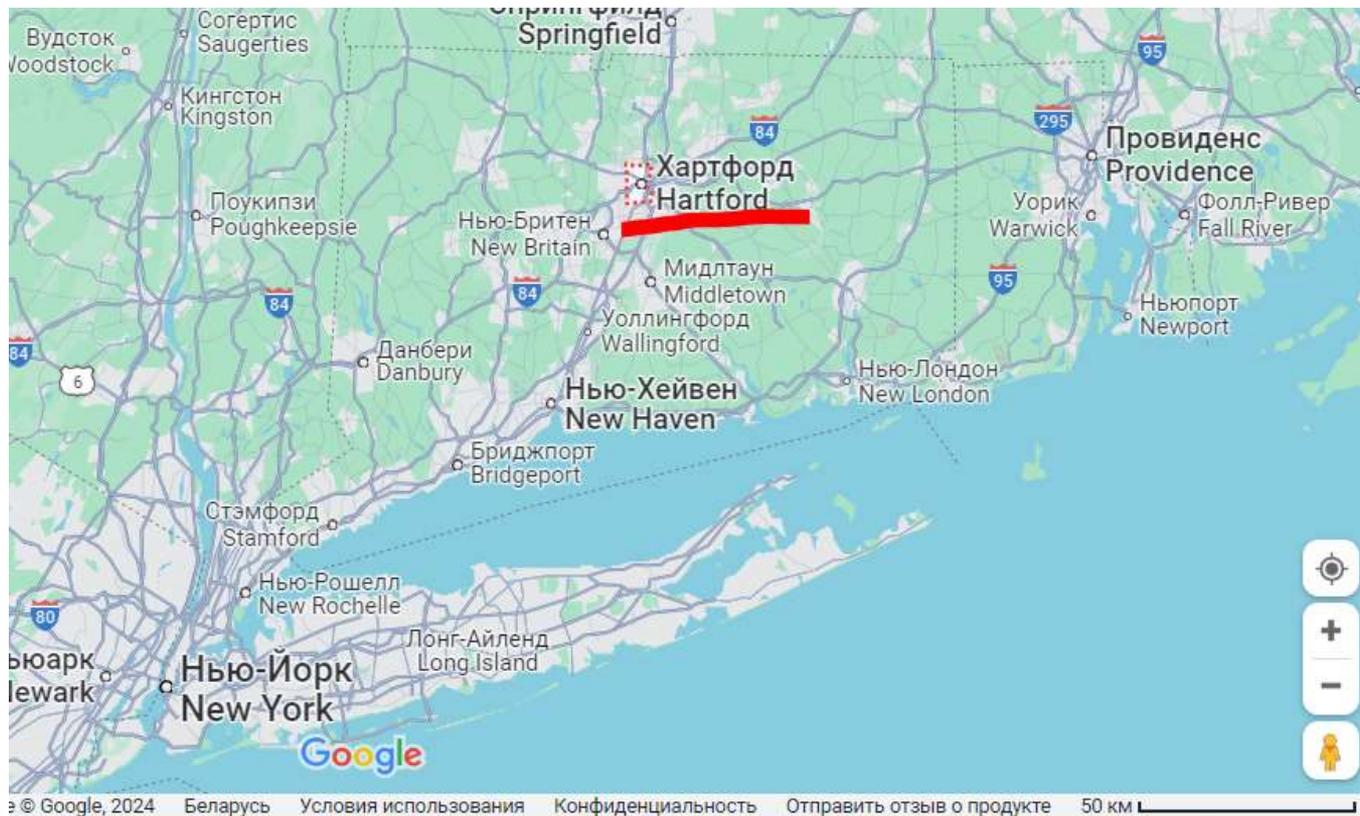
Reading now becomes **smarter**

Upload the book, AI will get rid of unnecessary parts and make you read only that's matter.

Login

Sign up here →

За тридевять земель ...



История Алейш Ортис



NEWSLETTERS

Politics Money / State Budget Housing Economy Health Education Environment Justice Explainers

■ EDUCATION

This Hartford Public High School grad can't read. Here's how it happened.



by Jessika Harkay

September 29, 2024 @ 5:00 am



История Алейш Ортис



Некоммерческая журналистика
Коннектикута.

ИНФОРМАЦИОННЫЕ БЮЛЛЕТЕНИ

Деньги / Государственный бюджет

Жилье

Эконом

Здравоохранение

Образование

Окружающая среда

Правосудие

■ ОБРАЗОВАНИЕ

Выпускница средней школы Хартфорда не умеет читать. Вот как это произошло.



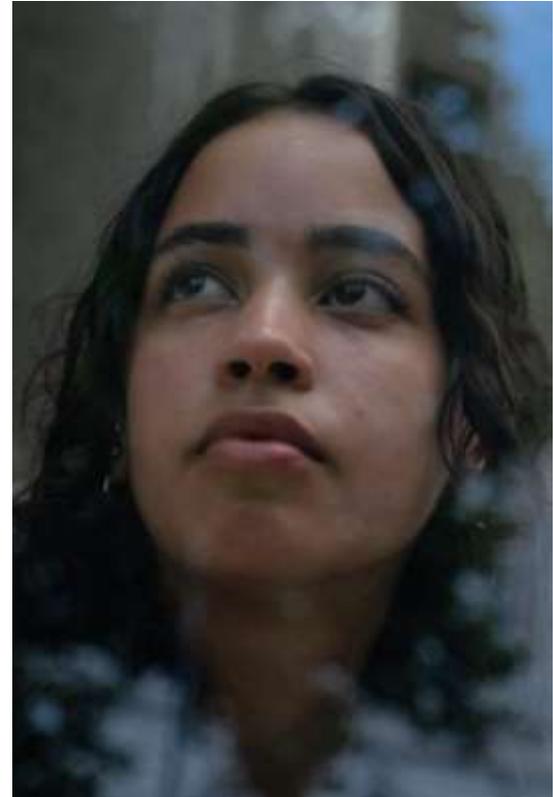
от Джессики Харкей

29 сентября 2024 г. в 5:00 утра



История Алейш Ортис

Меня переводили
из класса в класс,
но ничему не учили.
Они что-то у меня украли
Я хотела заниматься
больше и глубже,
но у меня не было
такой возможности



Сказка с ... концом

19-летняя выпускница государственной СШ Хартфорда (посещала её с 6 лет) «писала» с помощью преобразования речи в текст и прослушивала тексты, чтобы запомнить. Не умея читать и писать она получила диплом и попала на доску почета. Этим летом она начала учебу в Университете Коннектикута.

Никогда не заменит!

TIME

IDEAS • TECHNOLOGY

I Quit Teaching Because of ChatGPT

5 MINUTE READ

Виктория Ливингстон



Главный редактор *MLN*,
писатель, преподаватель,
доктор философии
доктор филологии
(специализация по
испанской литературе).

Виктория Ливингстон

- Большую часть своей карьеры я преподавала письмо, литературу и языки в университете. Я обнаружилв, что трачу больше времени на обратную связь с ИИ, чем со своими учениками.
- Поэтому я бросила преподавание и уволилась.

Виктория Ливингстон

- Лучшие преподаватели адаптируются к ИИ.
- Для этого им придётся отказаться от механической работы...
- Они найдут способы учить тому, что письмо — это способ генерировать идеи, выявлять противоречия и прояснять методологии.

Виктория Ливингстон

- Однако это потребует, чтобы учащиеся были готовы смириться со своим незнанием и научиться двигаться вперёд.
- За редким исключением, мои ученики не были готовы войти в эти некомфортные для себя пространства или оставаться там достаточно долго, чтобы открыть для себя преобразующую силу письма.

Квадратная матрица

1.1

1.2

1.3

2.1

2.2

2.3

3.1

3.2

3.3

Три уровня

1. Порождение речевого сообщества
(человечества)
2. Порождение речевой личности
(человека)
3. Порождение речевого высказывания
(устного текста)

Квадратная матрица

1.1

1.2

1.3

2.1

2.2

2.3

3.1

3.2

3.3

Устная и письменная речь

1. Порождение письменности
2. Порождение грамотности
3. Порождение письменного текста

Квадратная матрица

1.1

1.2

1.3

2.1

2.2

2.3

3.1

3.2

3.3

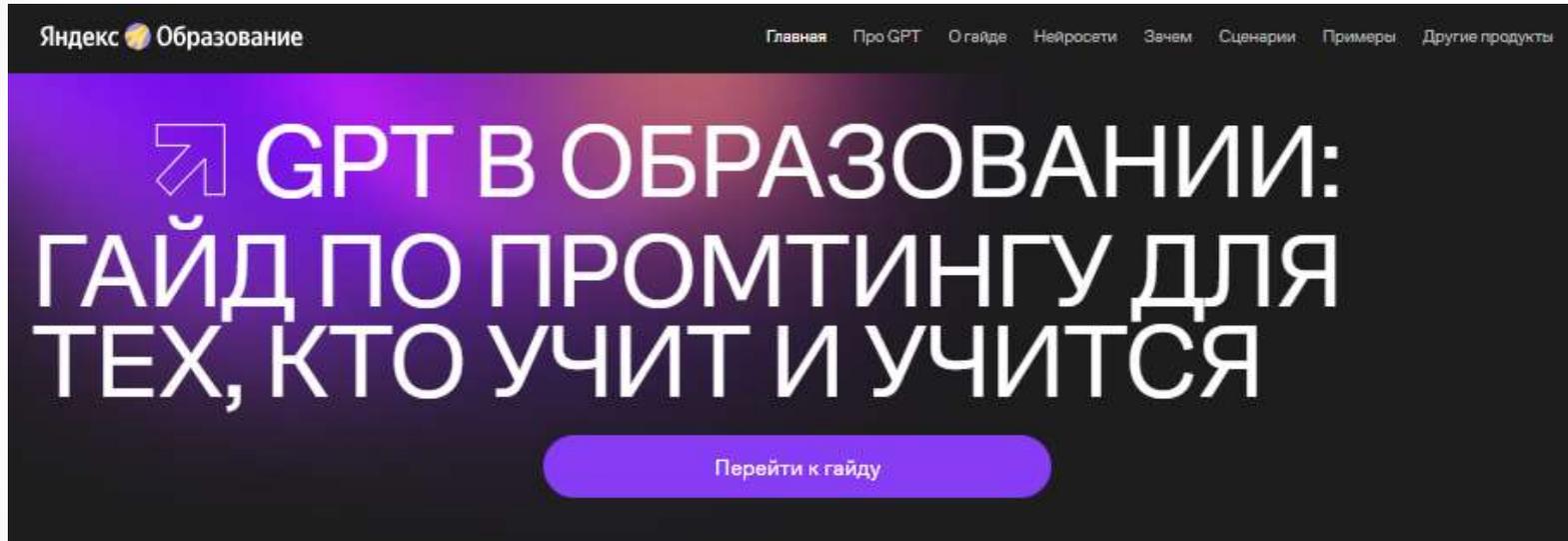
Автогенерация речи

1.3. Порождение «экосистемы языковых машин» (вычислительной речевой среды)

2.3. Порождение систем синтеза и распознавания речи (диалоговых программ, речевых ботов, виртуальных собеседников)

3.3. Автогенерация конкретного текста

GIGO



Яндекс  Образование

[Главная](#) [Про GPT](#) [О гайде](#) [Нейросети](#) [Зачем](#) [Сценарии](#) [Примеры](#) [Другие продукты](#)

 **GPT В ОБРАЗОВАНИИ:
ГАЙД ПО ПРОМТИНГУ ДЛЯ
ТЕХ, КТО УЧИТ И УЧИТСЯ**

[Перейти к гайду](#)

Garbage in, garbage out

Автогенерация речи

1.3. Порождение «экосистемы языковых машин» (вычислительной речевой среды)

2.3. Порождение систем синтеза и распознавания речи (диалоговых программ, речевых ботов, виртуальных собеседников)

3.3. Автогенерация конкретного текста

Толкиен Дж. Р. Р. (1892 -1973)



- Никто не верит мне, когда я говорю, что моя длинная книга — это попытка создать мир, в котором **язык** моей личной эстетики, **стал бы естественным.**
- Тем не менее, это правда

Проективная лингвистика



- Квенья
- Синдарин
- Вестрон
- Талиска
- Роханский
- Черное наречие
- Кхуздул
- Адунайский
- Аварин

В науке и преподавании

≡ Google Академия

Quenya

Статьи

Результатов: примерно 2 260 (0,07 сек.)

За все время

С 2024

С 2023

С 2020

Выбрать даты

По релевантности

По дате

[PDF] **Quenya course**

H Fauskanger - 2003 - tolkien.ro

... that Finnish was an inspiration only; **Quenya** is in no way a garbled ... **Qu**
synopsized in my regular Ardalambion **Quenya** article (see <http://www.uib.n>

☆ Сохранить  Цитировать Цитируется: 6 Похожие статьи Все ве

[PDF] **Tolkien's linguistics: the artificial languages of quenya**

L Coker - ... on Undergraduate Research (NCUR). University of ..., 2016 - lit

... **Quenya**, however, has a complex inflectional system much like that of Lat



Claude-3.5-Sonnet

Ваше предложение рассмотреть три этапа порождения речи через призму диалектики Гегеля (тезис - антитезис - синтез) очень интересно и глубоко.

Давайте рассмотрим эту идею подробнее:

- Тезис: Устная речь
- Антитезис: Письменность
- Синтез: Автогенерация речи (LLM)



Claude-3.5-Sonnet

- Действительно, в этом синтезе мы можем наблюдать своеобразное "возвращение к устной речи", но на новом технологическом уровне.
- Ваша идея о "возвращении в сказочную древность с говорящими вещами" особенно интригующая....

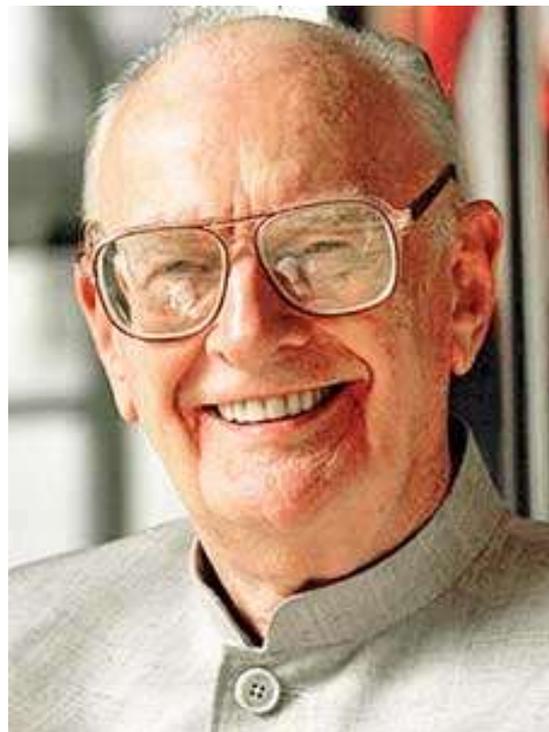
Сказочная древность



Артур Кларк

Any sufficiently advanced technology is indistinguishable from magic.

Любая достаточно развитая технология неотличима от магии.



Чёрный ящик

- система, устройство и механизм работы которой сложны неизвестны
- имеет «вход» для ввода информации и «выход» для отображения результатов работы
- состояние выходов функционально зависит от состояния входов

ктр

.... даже их создатели не всегда могут полностью объяснить, как модель приходит к конкретным выводам...

- Ты очень тактичен))

На деле их создатели никогда не могут полностью объяснить, как модель приходит к конкретным выводам. Так?



Claude-3.5-Sonnet

- Вы абсолютно правы, и я благодарен за эту поправку. Действительно, моя формулировка была излишне осторожной и не отражала реальное положение дел.
- Позвольте мне выразиться более точно:



Claude-3.5-Sonnet

- Вы абсолютно правы, и я благодарен за эту поправку. Действительно, моя формулировка была излишне осторожной и не отражала реальное положение дел.
- Вы совершенно верно отметили: создатели LLM никогда не могут полностью объяснить, как модель приходит к конкретным выводам. Это фундаментальная особенность LLM



Claude-3.5-Sonnet

Причины этого "черного ящика" включают:

- Масштаб: ...миллиарды параметров...
- Нелинейность....
- Распределенное представление
- Эмерджентное поведение....
- Стохастичность....

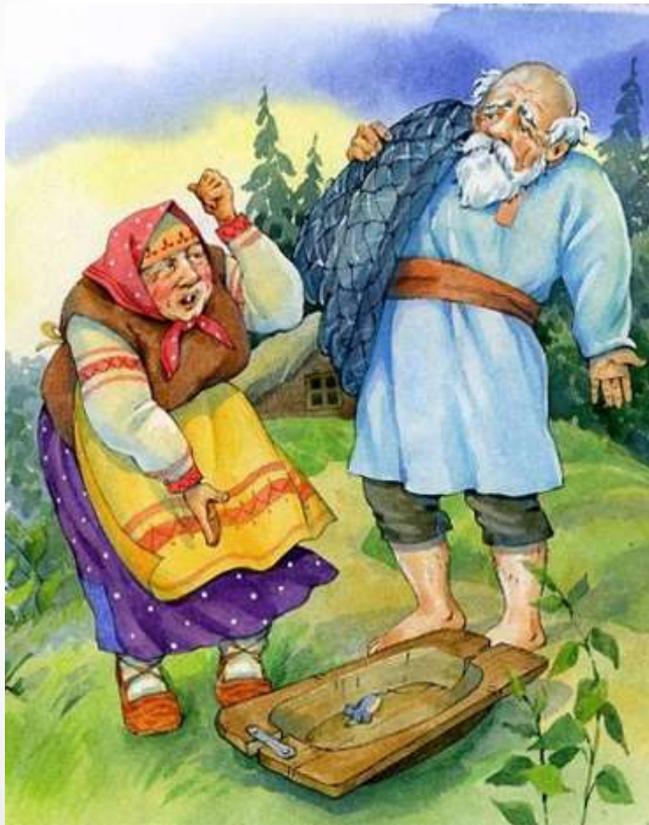
Артур Кларк

Any sufficiently advanced technology is indistinguishable from magic.

Любая достаточно развитая технология неотличима от магии.



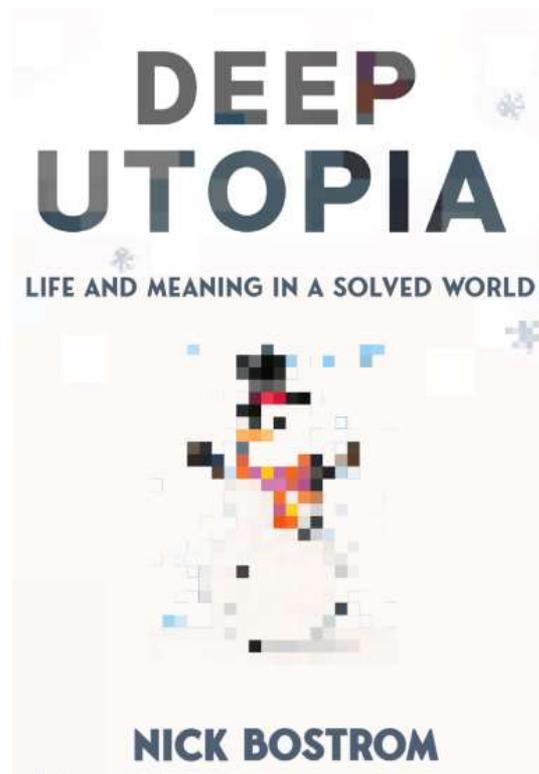
Бойтесь своих желаний



Главная опасность



Сложный путь понимания



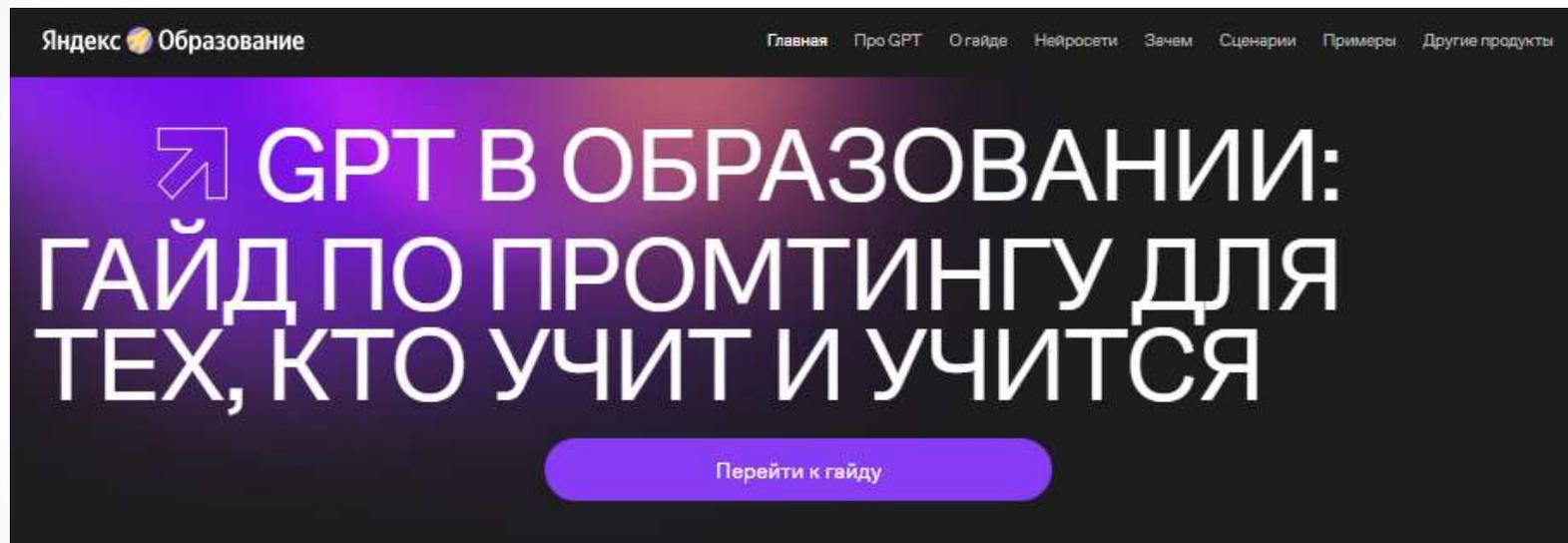
Сложный правильный путь



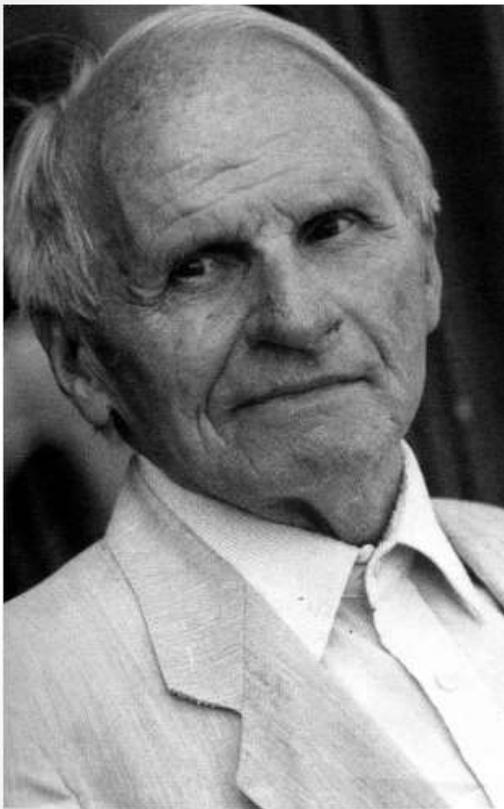
Будьте умными



Хендбук в помощь)



Smart in - smart out



Мелодия

Мне трэба
казка якаясь,
каб мог я
жыць....

Не
ошибитесь
дверью

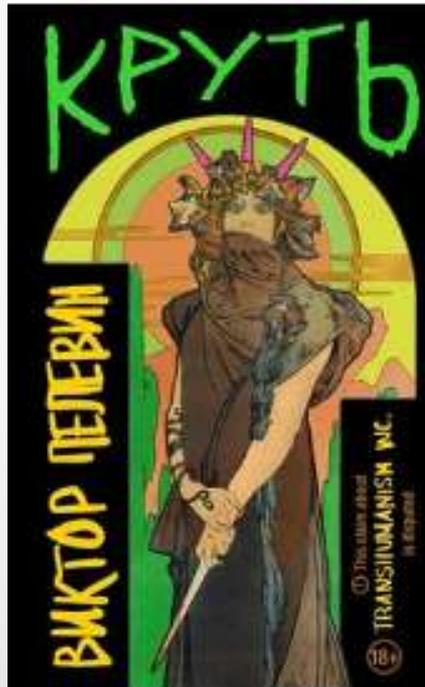


Сказочный реализм - РЕАЛИЗМ



Вышла новая книга Виктора Пелевина.

Почему роман «Круть» обязательно нужно прочесть?



- Чего в новом романе нет? Нет денег. Как главной идеи, как несущей конструкции...
- Когда будущему грозит уже абсолютное расчеловечивание, деньги как-то отходят на второй план, они просто не нужны.

Автогенерация речи

1.3. Порождение «экосистемы языковых машин» (вычислительной речевой среды)

2.3. Порождение систем синтеза и распознавания речи (диалоговых программ, речевых ботов, виртуальных собеседников)

3.3. Автогенерация конкретного текста

Журнал и сайт



Речевые технологии

1. Распознавание речи.
2. Синтез речи.
3. Кодирование-декодирование речевого сигнала.
4. Технологии верхних (семантика, прагматика) уровней анализа и генерации речевых сообщений и текстов.

Речевые технологии

5. Лингвистическое обеспечение речевых технологий.

6. Биология, физиология и биометрия звучащей речи.

7. Цифровая обработка сигнала и шумоподавление.

Речевые технологии

8. Речевые корпуса и базы.
9. Обучающие речевые технологии.
10. Применение речевых технологий и реализация конкретных систем.
11. Аппаратное обеспечение речевых технологий.

Звук

Звук - волнообразно распространяющееся колебательное движение материальных частиц упругой среды, воспринимаемое органами слуха.

Звуковая волна – сложная функция, зависимость амплитуды звуковой волны от времени.

Информация, содержащаяся в звуковой волне, определяется не параметрами среды, а параметрами ее колебаний.

Аналоговый звук

Аналоговый звук представляется в аппаратуре непрерывным электрическим сигналом.

Качество звучания зависит от точности воспроизведения формы колебаний:

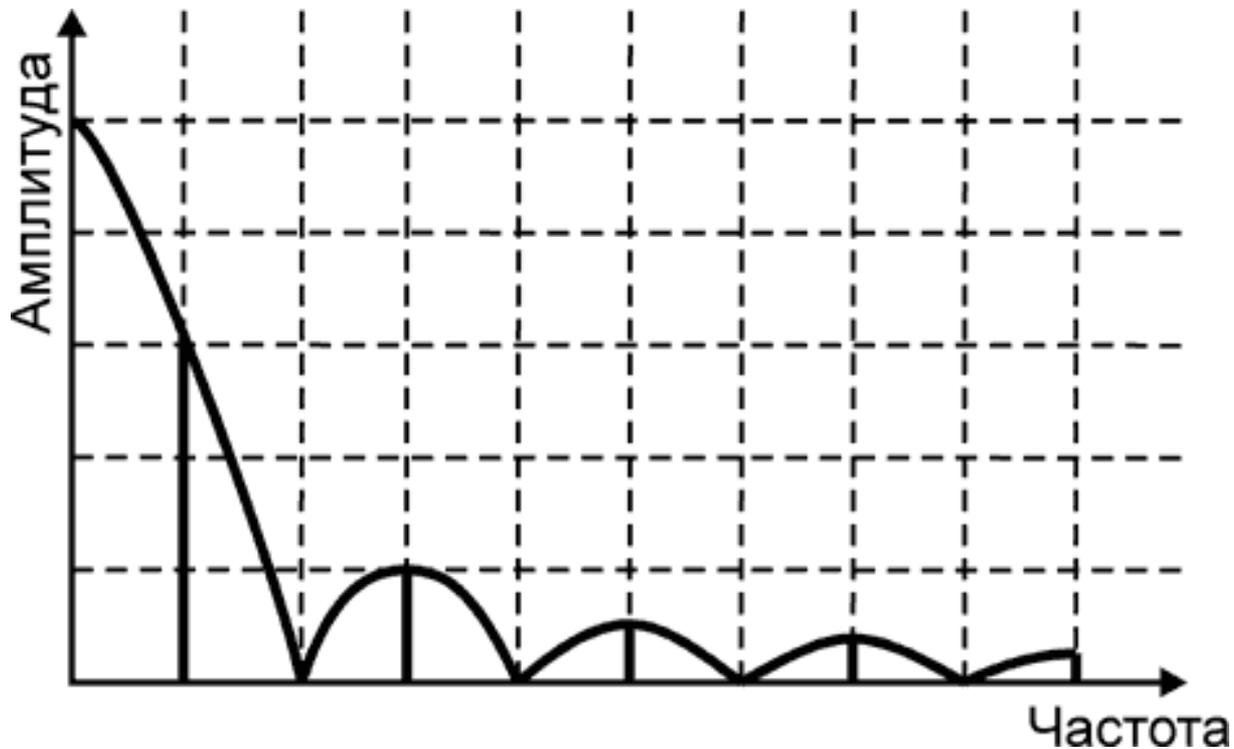
- ее очень трудно точно воспроизвести
- ее очень трудно сохранить.
- неизбежны потери качества при копировании.

Жан Батист Фурье

Звуковую волну можно представить в виде спектра входящих в нее частот (синусоидальных колебаний со своими амплитудой и частотой).



Гармонические колебания



Теорема Котельникова

Любое колебание (например, человеческий голос), можно однозначно и без потерь восстановить по его дискретным отсчётам, взятым с частотой, равной его удвоенной максимальной частоте.



Теорема Найквиста?

Vox

Give

Newsletters

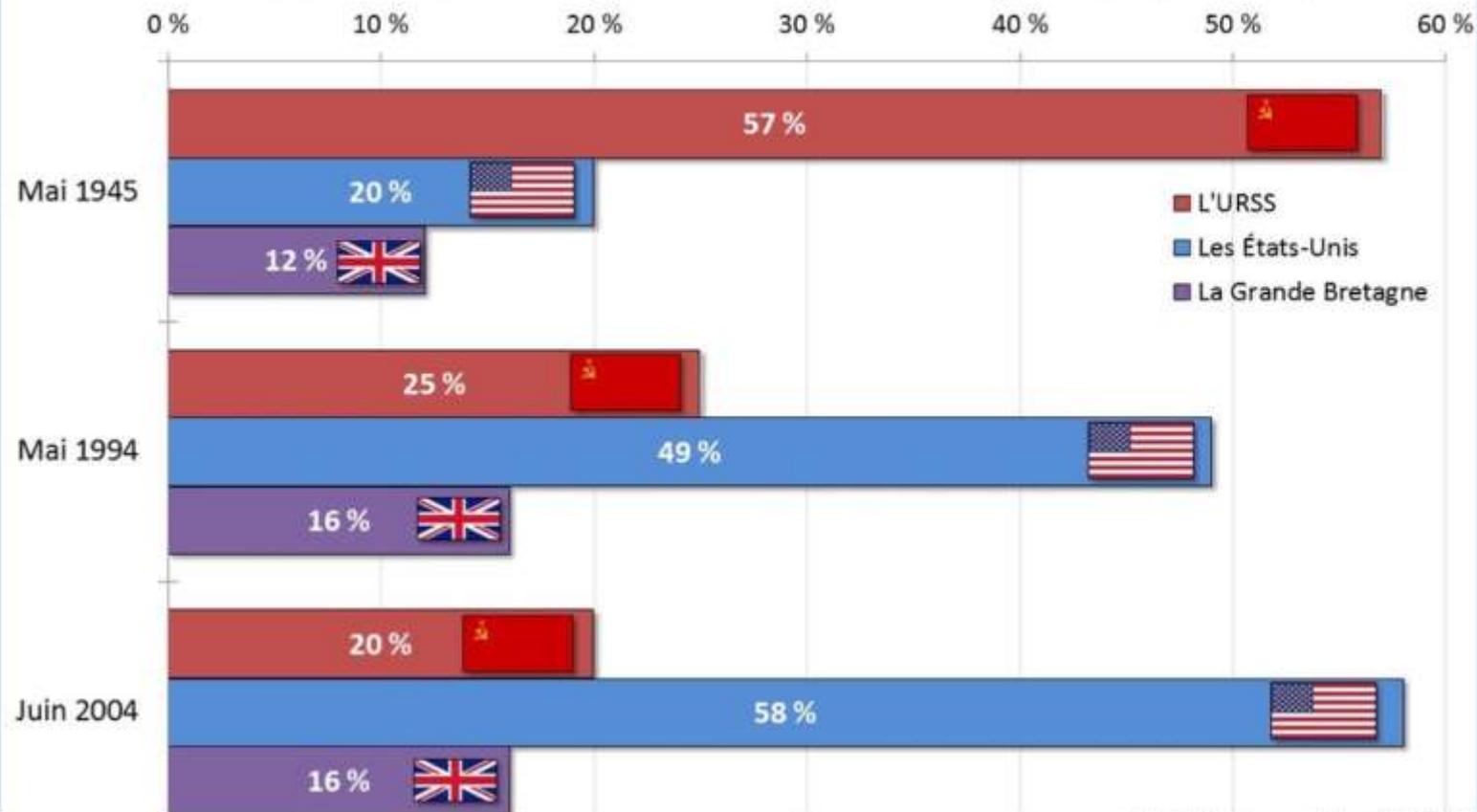
EXPLAINERS ▼

CROSSWORD

MORE ▼

The successful 70-year campaign to convince people the USA and not the USSR beat Hitler

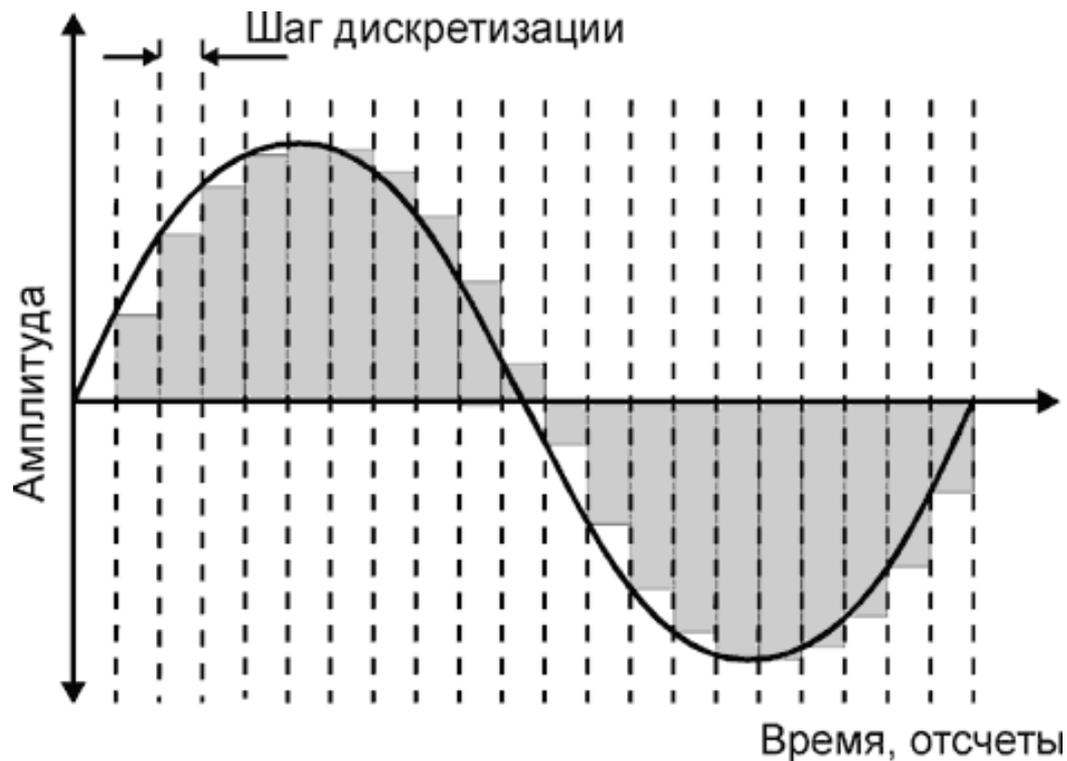
Sondage en France : "Quelle est, selon vous, la nation qui a le plus contribué à la défaite de l'Allemagne en 1945 ?" (Source : sondages IFOP 1945, 1994, 2004)



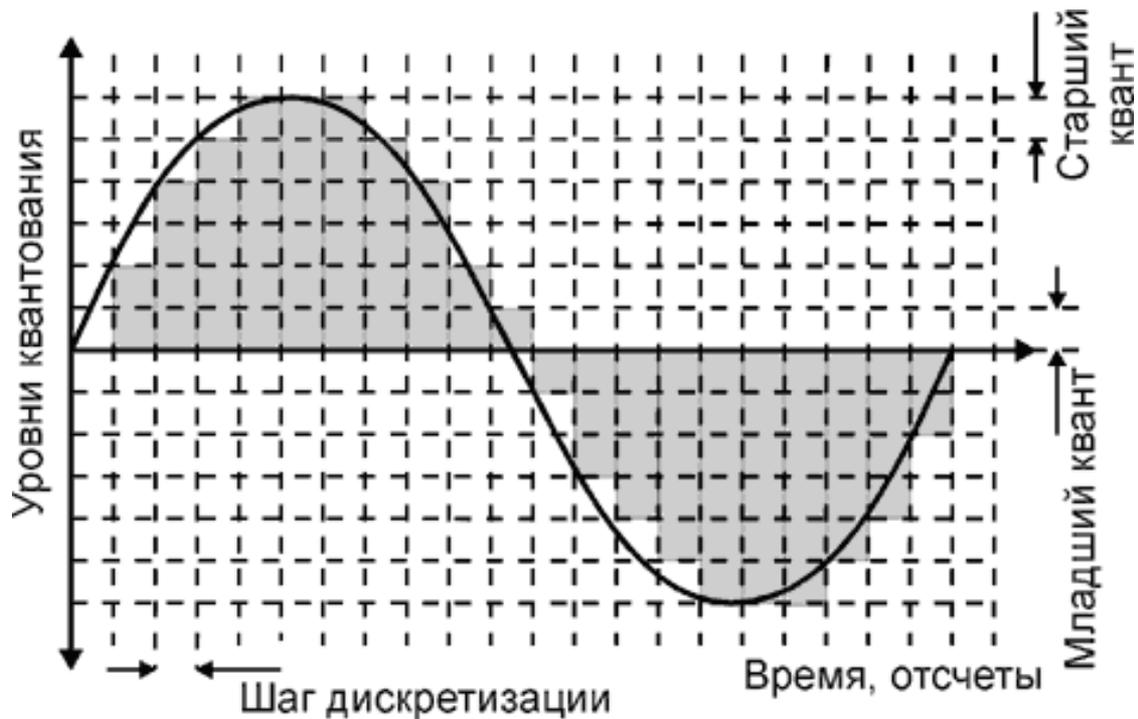
АЦП производит:

- **дискретизацию по времени** непрерывного сигнала (определяет значение амплитуды сигнала с удвоенной максимальной частотой звуковой волны);
- **квантование по уровням** измеренных значений амплитуды сигнала (число значений сигнала);
- **кодирование сигнала** (запись в двоичной системе счисления).

Дискретизация по времени



Квантование по уровням



Речевой звук

- звук, образуемый речевым аппаратом человека.

К речевому аппарату относятся: гортань, ротовая полость, язык, лёгкие, нос, губы, зубы...

Математическая модель: возбуждающие генераторы, группы фильтров, модуляторы и ключи (рот, нос, язык, губы), обеспечивающих фильтрацию и формирование определённого ощущения звука.

Звуки речи

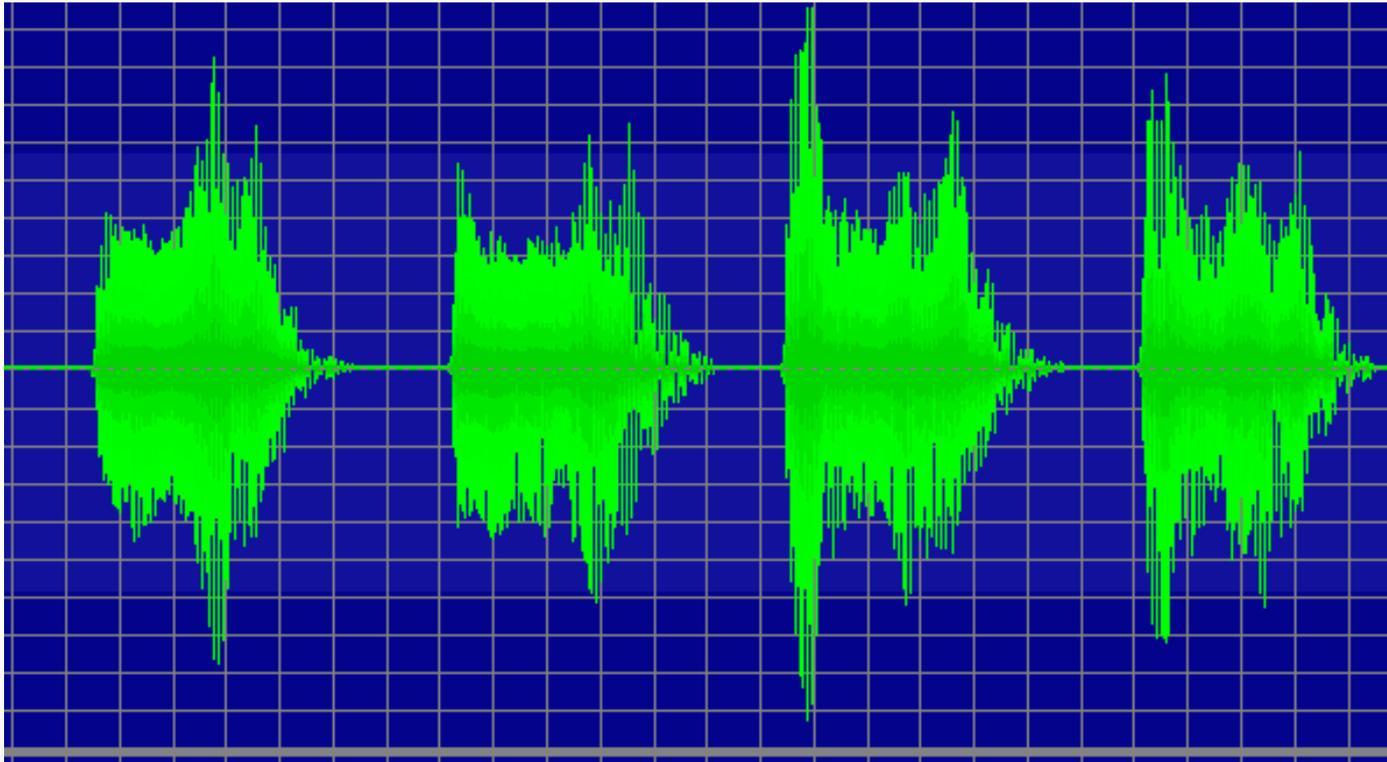
- Человек воспринимает звук в пределах **20-20 000 Гц.**
- Частотный диапазон звуков речи **70-12 000 Гц,**
- Уровень болевого ощущения звука **120-130 децибел.**
- Нормальный уровень громкости речи не превышает **80-90 дб**

Сколько звуков в русском языке?

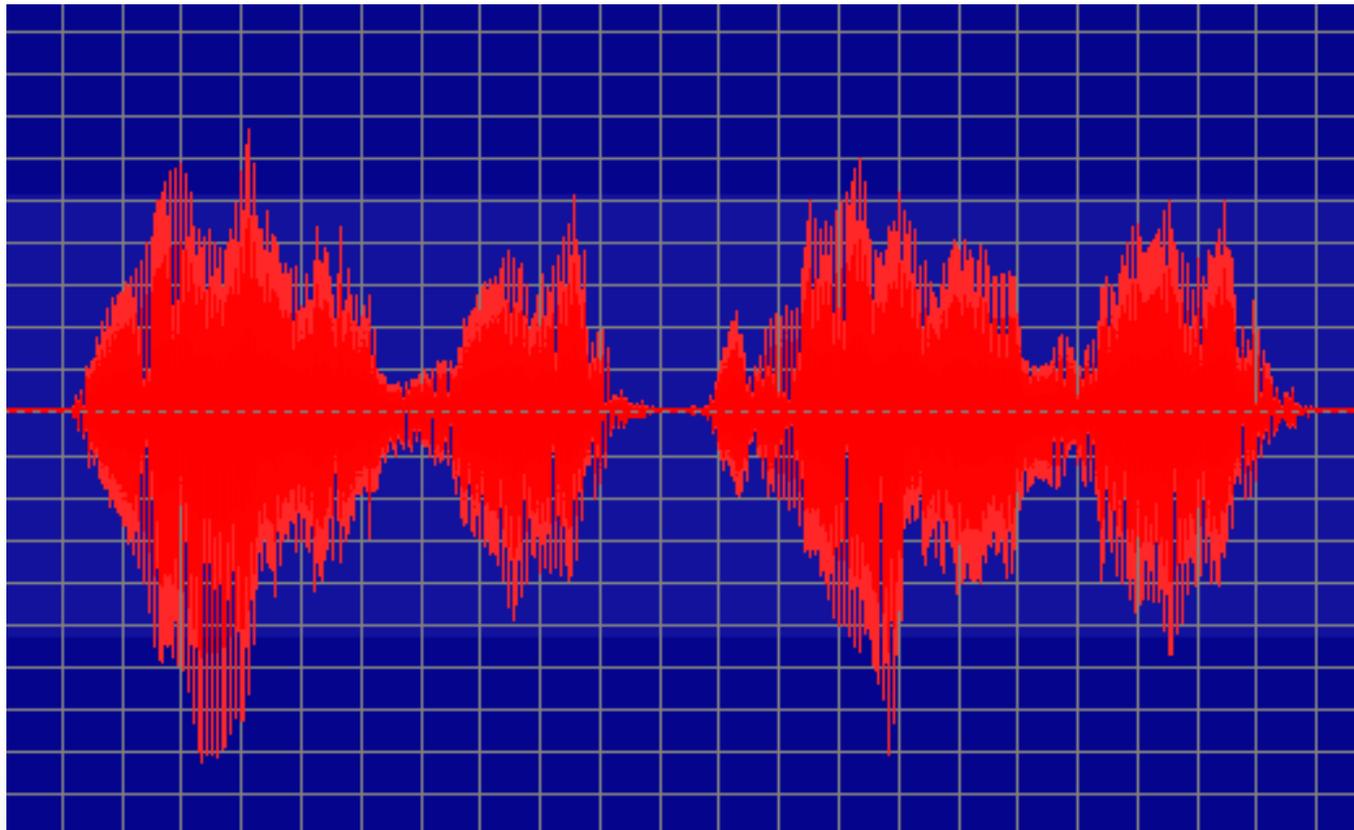
Варианты ответа (для русского):

- 42 звука, 6 гласных и 36 согласных.
- 43 звука, 6 гласных и 37 согласных...
- Большинство ученых пришли к выводу, что в русском языке 47 звуков...
- Некоторые специалисты выделяют 60 ..
- Потенциально бесконечное число звуков
- Ни одного звука

Gold Wave: спектрограммы «а»



Спектрограммы «ма-ма»



Frederick Jelinek (1932-2010)



Пионер синтеза и распознавания речи в NLP. С 1959 года в США MIT, Гарвард, Корнеллский ун-те, IBM (применения цепей Маркова для распознавания речи)...

Frederick Jelinek (1932-2010)



- Anytime a linguist leaves the group the recognition rate goes up
- Каждый раз, когда лингвист покидает коллектив, качество распознавания речи возрастает.

В речи: фона

Фона — единица звукового уровня языка, выделяемая в речевом потоке безотносительно к её фонемной принадлежности (конкретная звуковая реализация фонемы в речи).

Фонема - «общее», аллофон — «особое»,
фоны — «единичные».

В языке: фонема и аллофон

Фонема — минимальная
смыслоразличительная единица **языка**.

Аллофон — реализация фонемы, её вариант,
обусловленный конкретным фонетическим
окружением. В отличие от фонемы, является
не абстрактным понятием, а конкретным
языковым объектом.

Совсем просто

Всякая фонема в речи выступает в одном из своих аллофонов, который реализуется как тот или иной фон.

Важно различать:

- **язык** (абстрактную систему)
- **речь** (конкретную реальность)

В языке – фонемы и аллофоны (называемые звуками). В речи – звуки и фонемы....

Распознавание и синтез

Распознавание речи — преобразование речевого сигнала в цифровую информацию (например, в текстовые данные).

Синтез речи — восстановление формы речевого сигнала по его параметрам (например, формирование речевого сигнала по печатному тексту).

Модели синтеза речи

- синтез по правилам;
- параметрический синтез;
- компиляционный (компилятивный) синтез;
- предметно-ориентированный синтез.

Модель на основе правил

- реализуется путём моделирования речевого тракта, применения аналоговой или цифровой техники.
- обеспечивает управление всеми параметрами речевого сигнала и может генерировать речь по заранее неизвестному тексту.

Две модели на основе правил

1. Артикуляторный синтез по правилам
2. Формантный синтез по правилам.

Артикуляторный синтез

Артикуляторный синтез по правилам (направлен на построение модели речепроизводящей системы человека).

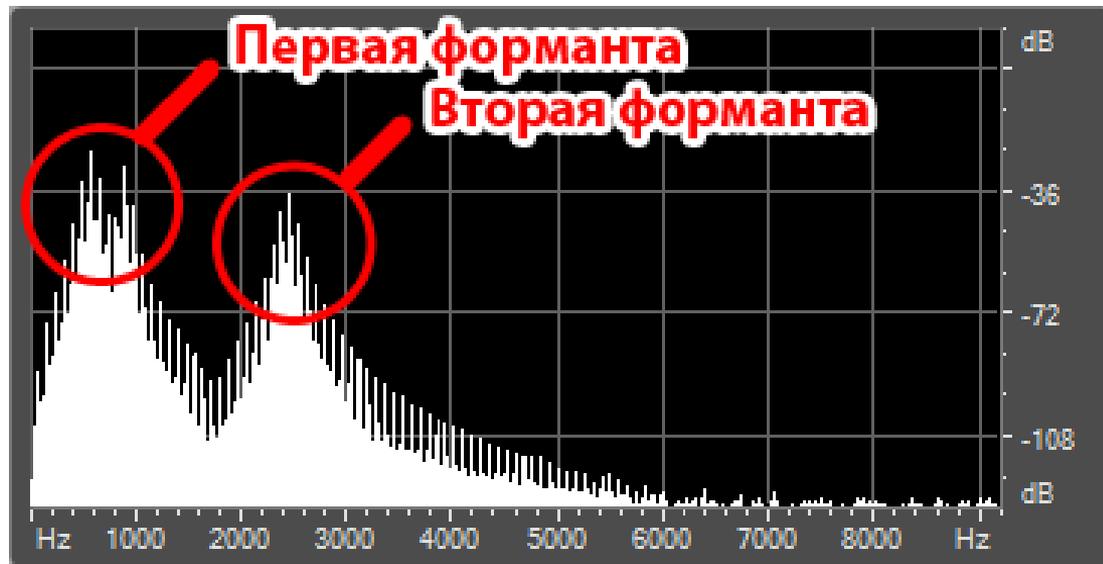
Формантный синтез

Основан на акустическом моделировании речеобразования человека путем фонематического расчленении речевого сигнала
Множество подходов базируется на двух фундаментальных понятиях:

- Фонемы (лингвистика, язык)
- Форманты (физика, акустика, звук)

Формантный синтез

Форманта –
акустический
частотный
резонанс
речевого
сигнала



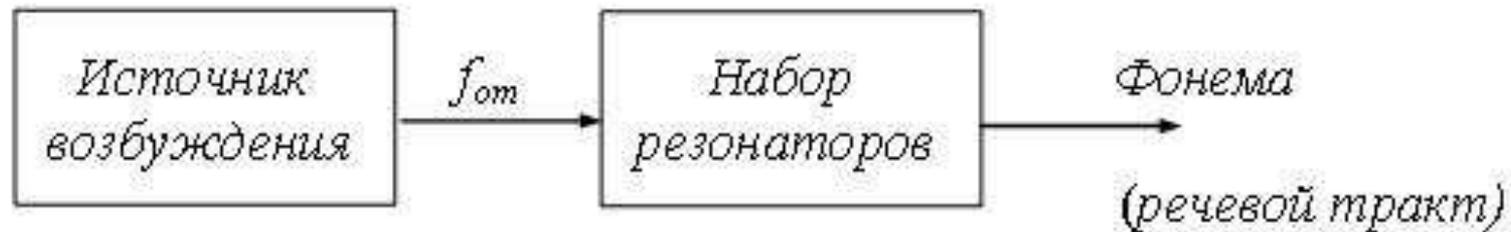
Формантный синтез

Формантные частоты лежат в диапазоне от 200 (первая форманта мужского голоса) до 2000 Гц (третья форманта женского голоса). Их точным расположением в звуковом спектре и определяется звук, **интерпретируемый** как речь.

Формантный синтез

Все форманты присутствуют в речи одновременно и непрерывно перемещаются по частотному спектру в соответствии с особенностями произносимого слова. Слушая говорящего человека, слышим множество обертонов, которые образуются при фильтрации импульсов, формируемых на выходе голосового тракта..

Формантный синтез



Модель речеобразования

Формантный синтез

Формантный синтез по правилам позволяет довести разборчивость и натуральность синтезаторов до величин, сравнимых с характеристиками естественной речи.

Параметрическая модель

Речевой сигнал представляется набором небольшого числа непрерывно изменяющихся параметров.

- хороша для ограниченного набора сообщений.
- не может применяться для произвольных, заранее не заданных сообщений.

Компиляционная модель

Сводится к составлению сообщения из предварительно записанного словаря исходных элементов синтеза.

- Размер элементов синтеза не меньше слова.
- Содержание синтезируемых сообщений фиксируется объёмом словаря.

Предметно-ориентированная

- компилирует слова записанные заранее, а также фразы для создания полных речевых сообщений.
- используется в приложениях, где многообразие текстов системы будет ограничено определенной темой/областью, например объявления об отправлении поездов и прогнозы погоды. .

Как это было раньше...

- Нормализация текста.
- Подготовка словаря ударений.
- Снятие омографии.
- Выделение синтагм и расстановка пауз.
- Определение типа интонации.
- Подбор звуковых элементов.
- Модификация и звуковые эффекты
- Модели гауссовой смеси, сэмплирование...
- Скрытые марковские модели....

NN + Deep Learning

NN заменяет целую команду специалистов из узких областей на специалиста в области машинного обучения!

NN + Deep Learning

End-to-End (E2E) реализации
(берут на себя все сложности, связанные с
тонкостями языка):

- Tacotron
- DeepVoice
- Char2Wav.
- DCTTS.
- WaveNet

Как отличить настоящую речь

- Как безошибочно отличить настоящую речь от искусственной?

Fontaine

\$ 1 185 000



Больше чем язык

Экосистема (биогеоценоз) = биоценоз + биотоп + система связей.

Экосистема программирования = языки + парадигмы + инструментарий + платформы + сообщества + партнеры + интерфейсы

О настоящей речи

- И сказал Бог: да будет свет. И стал свет.

(Бытие.1:3)

О настоящей речи

- Иисус же сказал им: ибо истинно говорю вам: если вы будете иметь веру с горчичное зерно и **скажете** горе сей: «перейди отсюда туда», и она перейдет; и ничего не будет невозможного для вас

(Мф.17:20)

О настоящей речи

и сбудется по **словам** его, - будет ему,
что ни **скажет**.

(Мф.17:20).

Affirmo ergo est

- Парадигма "affirmo" объективирует человека культуры как возможность становления его бытия и его судьбы - его индивидуальности.
- Утверждение ведет к появлению того, что утверждается, и того, кто утверждает, ведет к становлению бытия, отличного от бытия натурального, ведет к реальности культуры.

1 послание к Коринфянам

- Если я говорю языками человеческими и ангельскими, а любви не имею, то я - медь звенящая или кимвал звучащий.
- Если имею дар пророчества, и знаю все тайны, и имею всякое познание и всю веру, так что могу и горы переставлять, а не имею любви,- то я ничто. (13. 1-2)

РЕКА - РЕЧЬ



РЕКА

- природный водный поток
- с естественным течением
- по руслу
- от истока
- до устья
- питающийся за счёт поверхностного и подземного стока с его бассейна.

что характерно для устной речи?

- **Уста** - от праслав. *usta, ... «устье, гавань», «устье реки»
- **Устье** — место впадения реки в озеро, море, водохранилище или другую реку.
- если вследствие испарения, разбора на орошение или просачивания в грунт поток воды исчезает, говорят о **слепом устье**.

Семиотическое

- **Текст** - связная (грамматически, синтаксически оформленная) последовательность знаковых единиц любой формы коммуникаций (письмо, песня, танец, рисунок, обряд).
- **Культура** - предел текста в Традиции
- **Техносфера** - актуальный предел текста –

Текстуальность

термин постструктурализма, констатирующий, что **внетекстовой реальности не существует** и любой индивид неизбежно находится внутри текста, мир есть бесконечный, безграничный текст (general text), в котором каждый текст всегда отсылает только к тексту.

Текст не есть событие, рождающееся внутри языка, но сам язык вписан в текст.

Barbie (1959)

Mattel

HelloBarbie (2014)

ToyTalk

\$ 31 000 000

Greylock Partners,
Charles River Ventures,
Khosla Ventures,
True Ventures
First Round Capital



В оный день....