НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ МИФИ

ЛЕКЦИЯ 11.
СИНТЕЗ И РАСПОЗНАВАНИЕ РЕЧИ

Демидов Д.В.

Обработка аудиовизуальной информации
Бакалавры, 6 семестр. Магистры, 9 семестр

План лекции

- Синтез речи
  - Фонемы и морфемы
  - Методы синтеза
- Распознавание речи
  - Поиск звуковых фрагментов
  - Голосовая биометрия
  - 4-х факторная авторизация
  - Распознавание голоса

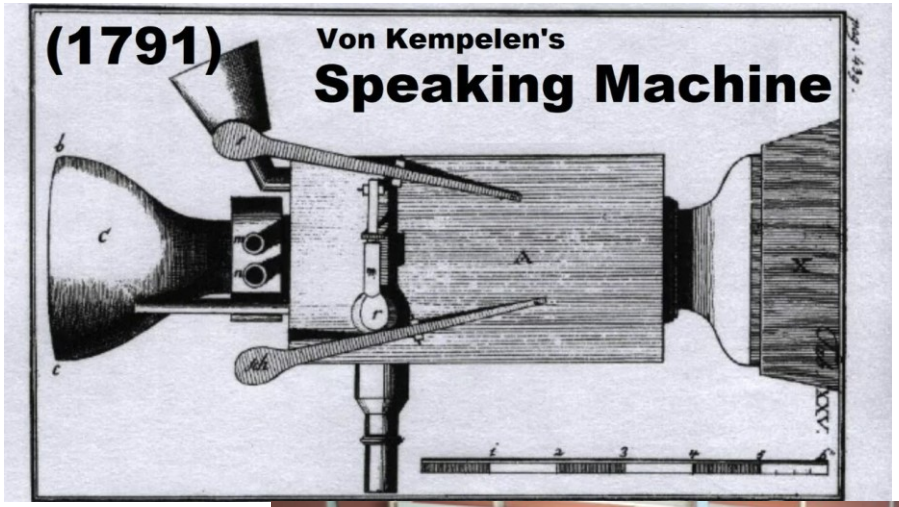
Синтез речи

Основные задачи

Некоторые технологии

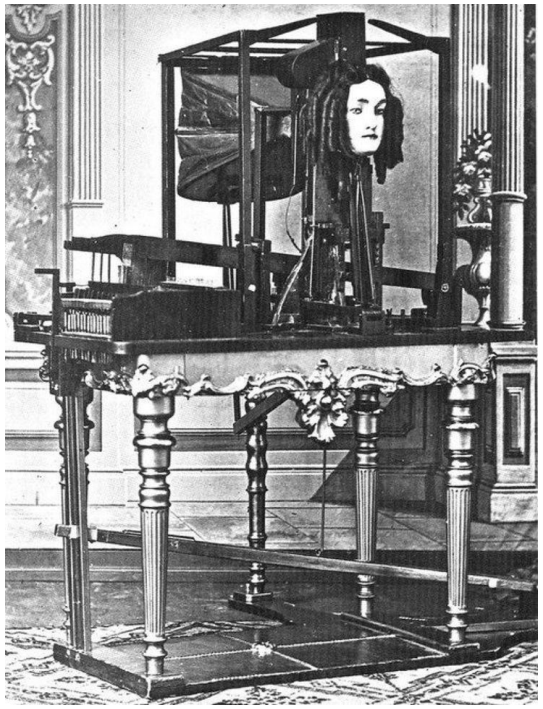
Говорящая машина Кемпелена (1791)

- Изобретена в Австрии в 1787 году Вольфгангом фон Кемпеленом.
- Опубликовано в 1791 «Mechanismus der menschlichen Sprache nebst der Beschreibung einer sprechenden Maschine von W. v. Kempelen».
- Воспроизводила ряд гласных и согласных звук.



Эуфония — механическая говорящая машина Фабера (1845)

- Воздушный мех, приводимый в движение ножной педалью — «лёгкие».
- Вытесняемый из меха воздух при помощи ряда клавиш направляется в различные по объёму трубки — разные положения голосовой щели и полости рта.

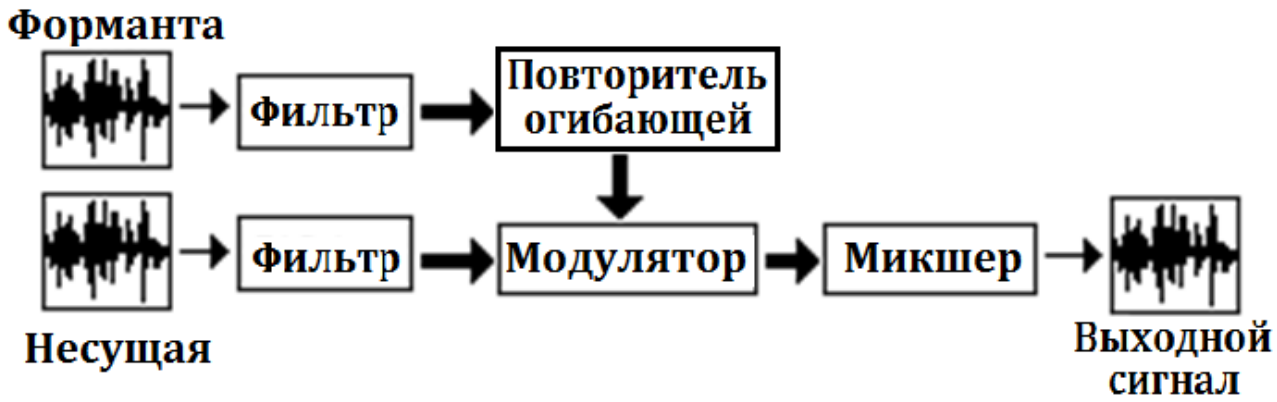


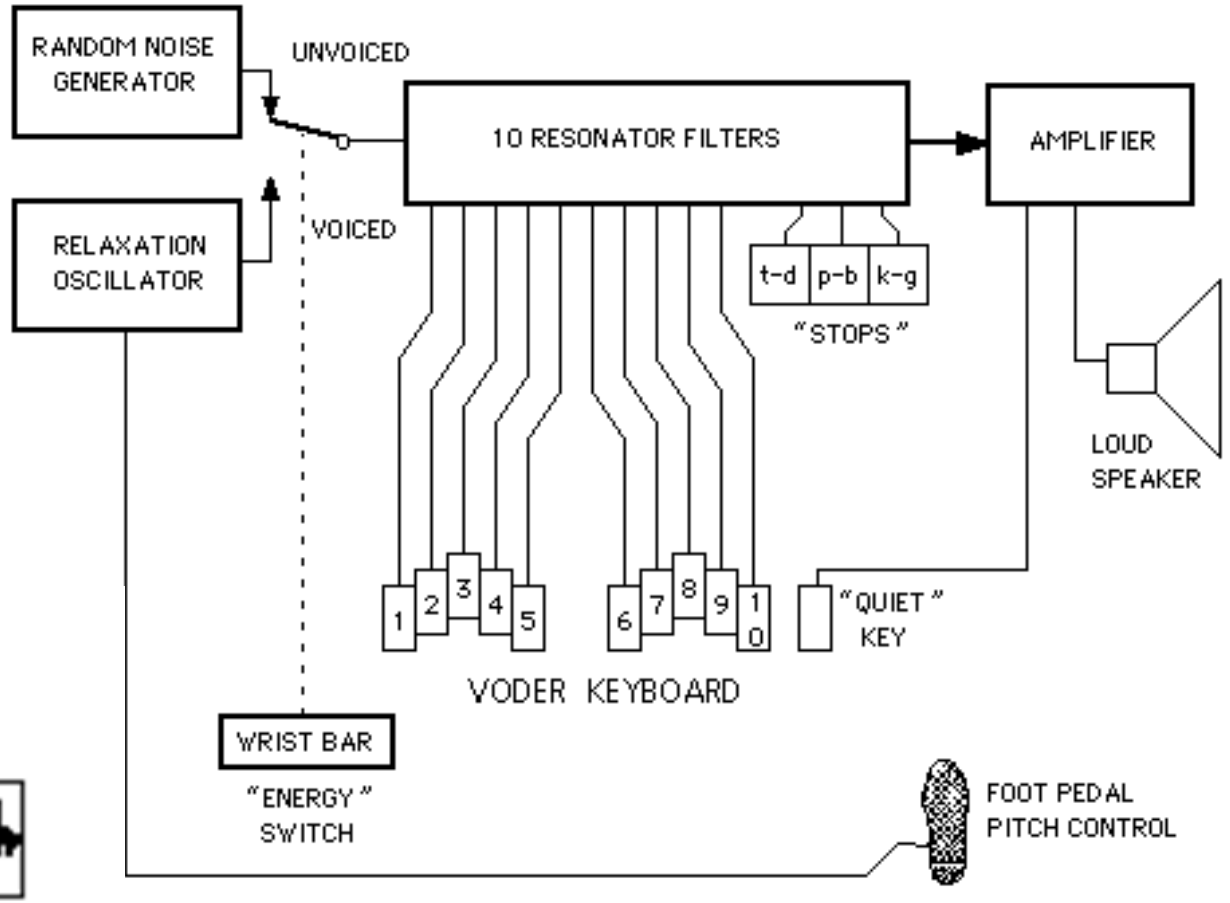
Вокодер — Хомер Дадли, Bell Labs (1928)

- ***Вокодер*** — voice coder — синтезатор речи на основе произвольного сигнала с богатым спектром.
- Изначально вокодеры были разработаны в целях экономии частотных ресурсов радиолинии системы связи при передаче речевых сообщений.
- Вместо собственно речевого сигнала передают только значения его определённых параметров (амплитуды), которые на приёмной стороне управляют синтезатором речи.

Синтезатор речи в вокодере

- Основу синтезатора речи составляют три элемента:
  - Генератор тонального сигнала для формирования гласных звуков;
  - Генератор шума для формирования согласных;
  - Система формантных фильтров для воссоздания индивидуальных особенностей голоса.



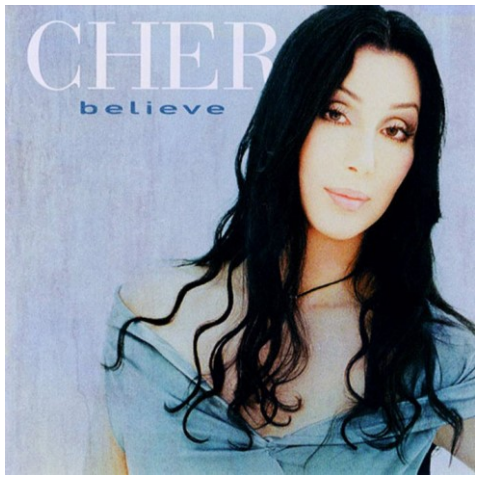


Электроника ЭМ-26

- Вокодер предназначен для синтезирования голоса и других внешних сигналов, имеет внутренний полифонический тон-генератор стрингс, анализатор-синтезатор, модуляционные и вокодерные эффекты. Состоит из 4-х секций управления звуком: Вибрато, Вокодер, Стрингс и Уровень, а также панели слева от клавиатуры для управления высотой тона и др. параметрами.
  - ВИБРАТО имеет регуляторы: акцент, глубина и скорость
  - ВОКОДЕР — хорус (вкл, вид), микрофон/вокодер
  - СТРИНГС — вкл., уровень, скрипка/виолончель, спад
  - УРОВЕНЬ — мкф, линия А, линия С
  - ПАНЕЛЬ слева — громкость, строй, интервал, октава, колесо питч-бенд

Вокодер в музыке

- Pink Floyd — Dogs (лай 9:16) Animals ‘1977
- Pink Floyd — Sheep (6:33 «the Lord is my shepherd…») Animals ‘1977
- Electric Light Orchestra — Sweet talking woman (Out of the blue ‘1977)
- Electric Light Orchestra — Confusion ‘1979
- Cher — Believe ‘1998 (0:33)
- Daft Punk — Get lucky ‘2013 (2:20 эффекты)



Фонема

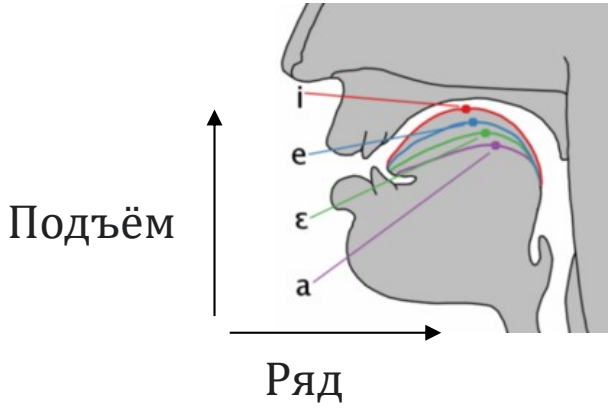
- **Фонема** — звук (греч.), абстрактная единица языка, соответствующая звуку речи как конкретной единице, в которой фонема материально реализуется.
- **Фоны** (звуки речи) бесконечно разнообразны: один человек никогда не произносит одинаково один и тот же звук.
- **Аллофон** — вариант реализации фонемы, обусловленный конкретным фонетическим окружением. Таких окружений гораздо меньше, чем число возможных звуков, поэтому удобно представлять фонему основными аллофонами, а не всеми вариантами фонов.
- Все варианты произношения звука, позволяющие правильно опознавать и различать слова с этим звуком, будут являться реализацией одной и той же фонемы.

Назначение фонем

- Фонема исполняет две ключевые функции, которые характеризуются наличием тесной связи друг с другом:
  - **конститутивная функция** (constitute) состоит в предоставлении фонемного инвентаря, своеобразного строительного материала для конструирования морфем и иных вышестоящих единиц языка;
  - **дистинктивная функция** (distinct) состоит, в свою очередь, в обеспечении различения отдельных морфем.
- Для записи транскрипций слов используется международный фонетический алфавит (но не только он).

Гласный = тон. Фонологические признаки

- Подъём языка к нёбу
  - Гласный верхнего подъёма (закрытый) — язык максимально приближен к нёбу
  - Гласный нижнего подъёма — язык должен быть максимально отодвинут от нёба.
- Удаление от зубов (ряд)
  - Гласный переднего ряда — язык находится как можно ближе к зубам, но без сужения
  - Гласный заднего ряда — язык отодвигается к основанию, не касаясь нижних зубов.
- Огублённость
  - Неогублённый гласный — губы не принимают округлое положение при произнесении.
  - Огублённый гласный — губы округляются или выпячиваются вперёд при артикуляции.



Шва [ə] — нейтральный гласный, лишённый признаков

Гласные в русском языке

- Гласные звуки, фонемы: А И О У Ы Э
- Буквы Е Ё Ю Я соответствуют либо сочетаниям согласный + гласный, либо смягчению согласного + гласный.
- В безударных слогах гласные редуцируются и произносятся кратко и нечётко. Это порождает аллофоны.
- Всего 23 гласных аллофонов и фонем.
- Виды транскрипций:
  - Фонетическая — запись с точностью до аллофонов, в квадратных скобках [ ]
  - Фонематическая — запись с точностью до фонем, в косых чертах / /
  - Морфонологическая — в прямых скобках | |

Подъём \ Ряд	Передний	Средний	Задний
Верхний	i /и/	(ɨ) /ы/	ʊ /ʏ/
Средний	e̝ /э/		o̝ /о/
Нижний		ä /а/	

МФА. Ударные гласные в русском языке

№	Фонема	Аллофон	Огубленность	Ряд	Подъём	Использование	Примеры
1	/а/	[ɑ]	Нет	Передний	Нижний	После твёрдых	трав**а́**, **я**хта
2	/а/	[æ] [ä]	Нет	Передний	Нижний	После мягких	п**я́**ть, в**я́**зь, **я**рь
3	/а/	[ɑ] [ɑ̟]	Нет	Задний	Нижний	После твёрдых перед л (/ɬ/)	п**а́**лка
4	/а/	[ə] [ъ]	Нет	Средний	Средний	Шва	чт**о́**б тебя!
5	/е̟ / (/э/)	[e]	Нет	Передний	Средне-верхний	Между мягкими	п**е**нь; л**е**с, п**е**пел
6	/е̟ / (/э/)	[ɛ]	Нет	Передний	Средне-нижний	Перед твёрдыми	ц**е**ль; ж**е**ст; **э**тот
7	/и/	[i] [и])	Нет	Передний	Средний	Перед мягкими	с**и́**него
8	/ы/	[ɨ] [ы]	Нет	Средний	Верхний		т**ы**
9	/о/	[o]	Да	Задний	Средне-верхний		**о́**блако
10	/о/	[ɵ̟]	Да	Средний	Средне-верхний	Огубленный шва между мягкими	т**ё**тя
11	/у/	[ʊ]	Да	Задний	Верхний		п**у́**ля
12	/у/	[ʉ]	Да	Средний	Верхний	Между мягкими	ч**у́**ть; тр**ю**м

МФА. Безударные гласные в русском языке

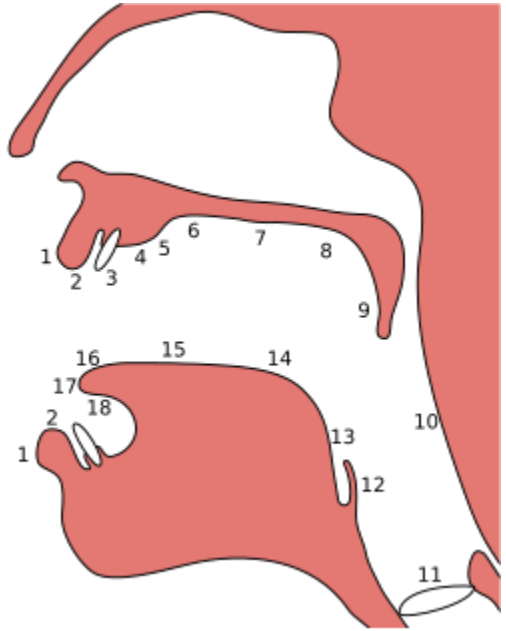
№	Фонема	Аллофон	Огубленность	Ряд	Подъём	Использование	Примеры
13	/а/	[ɐ] [ʌ]	Нет	Средний	Нижний	В начале, перед ударным слогом, в заударном в конце	олова́нный [ɐɬɐˈvʲɶnːɨ]; па́ро́м; с**оо**бра**а**жа́ть; ст**о**па́
14	/а/	[ə]	Нет	Средний	Средний	Шва	ко́ж**а**; о́бла**к**о; се́рдц**е**; гля́д**я**
15	/и/	[ɪ] [и^{e}] [ь]	Нет	Передний	Верхний	Не перед мягкими	т**я**жёлый; **э**та́п; ч**е**ты́ре; дер**е**во; п**я**так; **и**кра
16	/ы/	[ɨ̞]	Нет	Средний	Верхний		т**ы**кать
17	/ы/	[ɯ̞]	Нет	Средний	Верхний	Дифтонгизируется после л (/ɬ/)	пл**ы**ть [pɬɯ̟ɨ̟ʲ]
18	/ы/	[ɬ] [ɨ̞] [ы^{э}]	Нет	Средний	Верхний		д**ы**ша́ть; ц**е**на́, ж**е**на́
19	/ы/	[ɵ]	Нет	Средний	Верхний	После /ц/	на танц**е**; на танц**ы**
20	/ʏ/	[ʊ]	Да	Задний	Верхний		м**у**жчи́на, с**у**хой [sʊˈxoj]
21	/ʏ/	[ʊ̟]	Да	Средний	Верхний	В начале или между мягкими	**ю**ти́ться [jʊ̟ˈtʲitsə], л**ю**би́ть
22	/и/	[ɪ̯]				Неслоговый	мо**й**, тво**й**, по́**й**ман
23	/ʏ/	[ʊ̯]				Неслоговый	Ивано́**в**, рома́но**в**, Козла́доев**в**

Согласный = шум + тон. Фонологические признаки

- По способу образования:
  - Пульмонические — с участием лёгких
    - Шумные
      - Смычные
        - Взрывные
        - Аффрикаты
      - Щелевые (фрикативные, спиранты)
    - Сонорные
      - Дрожащие
        - многоударные
        - одноударные
      - Носовые
      - Аппроксиманты
        - боковые
        - скользящие
  - Непульмонические — без участия лёгких (в русском нет)
    - Щелчки
    - Имплозивы

По месту образования:

1. Внешнелабиальные,
2. Внутрилабиальные,
3. Дентальные,
4. Альвеолярные,
5. Постальвеолярные,
6. Препалатальные,
7. Палатальные,
8. Велярные,
9. Увулярные,
10. Фарингальные,
11. Глоттальные,
12. Эпиглоттальные,
13. Радикальные,
14. Постдорсальные,
15. Предорсальные,
16. Ламинальные,
17. Апикальные,
18. Субапикальные



Согласные. Классификация

По месту образования

По способу образования

	Губно-губные	Губно-зубные	Зубные	Альвео-лярные	Постальвео-лярные	Ретрофлексные	Палатальные	Задне язычные	Увулярные	Фарингальные	Глотталь ные
Взрывные	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Носовые	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Дрожащие	ʙ			r					ʀ		
Одноударные		ⱱ		ɾ		ɽ					
Фрикативные	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Латеральные фрикативные				ɬ ɮ							
Аппроксиманты		ʋ		ɹ		ɻ	j	ɰ			
Латеральные аппроксиманты				l		ɭ	ʎ	ʟ			

Согласные. Классификация

По способу образования

По месту образования

Способ \ Место	Губно-губные	Губно-зубные	Зубные	Альвео-лярные	Пост-альвео-лярные	Ретро-флексные	Палатальные	Задне-язычные	Увулярные	Фарин-гальные	Глоттальные
Взрывные	p b		t d				c	k g	q		
Носовые	m		n						N		
Дрожащие	ʙ		r						R		
Одноударные											
Фрикативные		f v		s z				x			h
Латеральные фрикативные											
Аппроксиманты							j				
Латеральные аппроксиманты			l					L			

Согласные в русском языке

- 40 согласных звуков:
  - 18 твёрдых: Б В Г Д Ж З К Л М Н П Р С Т Ф Х Ц Ш
  - +2 твёрдых в «ага» и «дзен»
  - 19 мягких (палатализованных): Б В Г Д Ж З Й К Л М Н П Р С Т Ф Х Ч Щ
  - +1 мягкий в «джаз»

МФА. Согласные в русском языке. Б В Г Д Ж З Й

Буква	Фонема	тв/м, зв/гл	Способ образования	Место образования	Примеры
б	/b/	тв. зв.	Шумный взрывной	Губно-губной	о**б**увь
б	/b^{j}/	м. зв.	Шумный взрывной	Губно-губной	о**б**ь, **б**играмма
в г	/v/	тв. зв.	Шумный фрикативный	Губно-зубной	**в**олос, е**г**о
в	/v^{j}/	м. зв.	Шумный фрикативный	Губно-зубной	**в**еник
г	/g/	тв. зв.	Шумный взрывной	Заднеязычный	**г**ном
г	/g^{j}/	м. зв.	Шумный взрывной	Заднеязычный	**г**ений
д	/d/	тв. зв.	Шумный взрывной	Переднеязычный зубной	**д**ом
д	/d^{j}/	м. зв.	Шумный взрывной	Переднеязычный альвеолярный	**д**еньги
ж	/ʐ/	тв. зв.	Шумный фрикативный	Переднеязычный ретрофлексный	**ж**ест,
ж	/ʑː/	м. зв.	Шумный фрикативный	Переднеязычный ретрофлексный	**ж**юльен, визжать, во**жж**и
з	/z/	тв. зв.	Шумный фрикативный	Переднеязычный зубной	**з**онт
з с	/z^{j}/	м. зв.	Шумный фрикативный	Переднеязычный зубной	**з**елень, про**сь**ба
й ё я	/j/	м.	Сонорный скользящий	Палатальный	**й**од, **ё**жик, ко**й**от

МФА. Согласные в русском языке. К Л М Н П Р С

Буква	Фонема	тв/м, зв/гл	Способ образования	Место образования	Примеры
к	/k/	тв. гл.	Шумный взрывной	Заднеязычный	**к**ол
к	/k^{j}/	м. гл.	Шумный взрывной	Заднеязычный	**к**ьянти, **к**ельвин
л	/ɬ/	тв.	Сонорный боковой	Переднеязычный зубной	**л**ошадь
л	/l^{j}/	м.	Сонорный боковой	Переднеязычный альвеолярный	**л**ягушка, мыс**ль**
м	/m/	тв.	Сонорный носовой	Губно-губный	**м**анго
м	/m^{j}/	м.	Сонорный носовой	Губно-губный	**м**яукать
н	/n/	тв.	Сонорный носовой	Переднеязычный зубной	**н**ога
н	/n^{j}/	м.	Сонорный носовой	Переднеязычный альвеолярный	**н**ега
п	/p/	тв. гл.	Шумный взрывной	Губно-губный	**п**оле
п	/p^{j}/	м. гл.	Шумный взрывной	Губно-губный	**п**ётр, **п**иво, ко**п**ьё
р	/r/	тв.	Сонорный дрожащий	Переднеязычный альвеолярный	**р**ог
р	/r^{j}/	м.	Сонорный дрожащий	Переднеязычный альвеолярный	**р**ифма, **р**ьяный
с з	/s/	тв. гл.	Шумный фрикативный	Переднеязычный зубной	**с**он, гры**з**ть
с	/s^{j}/	м. гл.	Шумный фрикативный	Переднеязычный зубной	**с**еть

МФА. Согласные в русском языке. Т Ф Х Ц Ч Ш Щ

Буква	Фонема	тв/м, зв/гл	Способ образования	Место образования	Примеры
т	/t/	тв. гл.	Шумный взрывной	Переднеязычный зубной	**т**он
т	/tʲ/	м. гл.	Шумный взрывной	Переднеязычный альвеолярный	**т**ень
ф в	/f/	тв. гл.	Шумный фрикативный	Губно-зубной	**ф**он, выста**в**ка
ф в	/fʲ/	м. гл.	Шумный фрикативный	Губно-зубной	**ф**ишка, чер**вь**
х	/x/	тв. гл.	Шумный фрикативный	Заднеязычный	**х**рен
х г	/xʲ/	м. гл.	Шумный фрикативный	Заднеязычный	**х**имия, бо**г**, лё**г**кий
г хг	[ɣ]	тв. зв.	Шумный фрикативный	Заднеязычный	а**г**а, **г**осподи, бу**хг**алтер
дз	[ʣ]	тв. зв.	Шумная аффриката	Переднеязычный альвеолярный	кин-**дз**а-**дз**а
ч	/tɕ/	м. гл.	Шумная аффриката	Переднеязычный ретрофлексный	**ч**ак-**ч**ак
дж жд	[ʤ]	м. зв.	Шумная аффриката	Переднеязычный ретрофлексный	до**ч**ь бы, **дж**аз, до**жд**и
ц	/ʦ/	тв. гл.	Шумная аффриката	Переднеязычный альвеолярный	**ц**апля
ш	/ʂ/	тв. гл.	Шумный фрикативный	Переднеязычный ретрофлексный	**ш**мель
щ сч жч	/ɕː/	м.	Шумный фрикативный	Переднеязычный ретрофлексный	**щ**ель, **сч**астье, му**жч**ина

Супрасегментные средства (просодии)

- Просодии — описывают супрасегментные средства организации речи:
  - Длительность, ударение, ритм
  - Интонация
  - Тоны (нежно, гневно, хвастливо, иронично)
- Определяют длительность гласных и согласных звуков, ударение, основной тон и ритм языка.
- Используются для передачи интонации.

Проклитики и энклитики

- Проклитики — безударные слова, фонетически примыкающие к следующему слову:
  - Я д**у**мал
  - Под б**е**регом [плд͡б’э’р’ьгьм]
  - Не сл**ы**шу [н’и^{э}͡слы’шу]
- Энклитики — безударные слова, фонетически примыкающие к предшествующему слову
  - Зн**а**ю я
  - Н**а** слово
  - Уход**и** же [ухлд’и’͡жь]
  - Зн**а**ешь ли [зна`jьш͡л’ь]

Морфема

- **Морфема** — наименьшая единица языка, значимая часть слова.
- Деление морфем на части приводит к выделению незначимых элементов — фонем.
- **Классы** морфем:
  - Корни
  - Аффиксы
    - Префиксы (приставки),
    - Постфиксы (после корня)
      - Суффиксы,
      - Флексии (окончания)
      - Постфиксы (после корня)
    - Интерфиксы (между двух корней)
    - Некоторые более экзотические

Сложность сегментирования речи

- Устная речь являет собой практически непрерывный звуковой поток, представляющий определенные сложности для сегментирования.
- В большинстве случаев для выделения фонемы необходимо знание языка.
- Выделяемость фонемы некоторым образом сопряжена со смыслом и со значением, хотя сама по себе она не является значащей единицей.

Методы синтеза речи

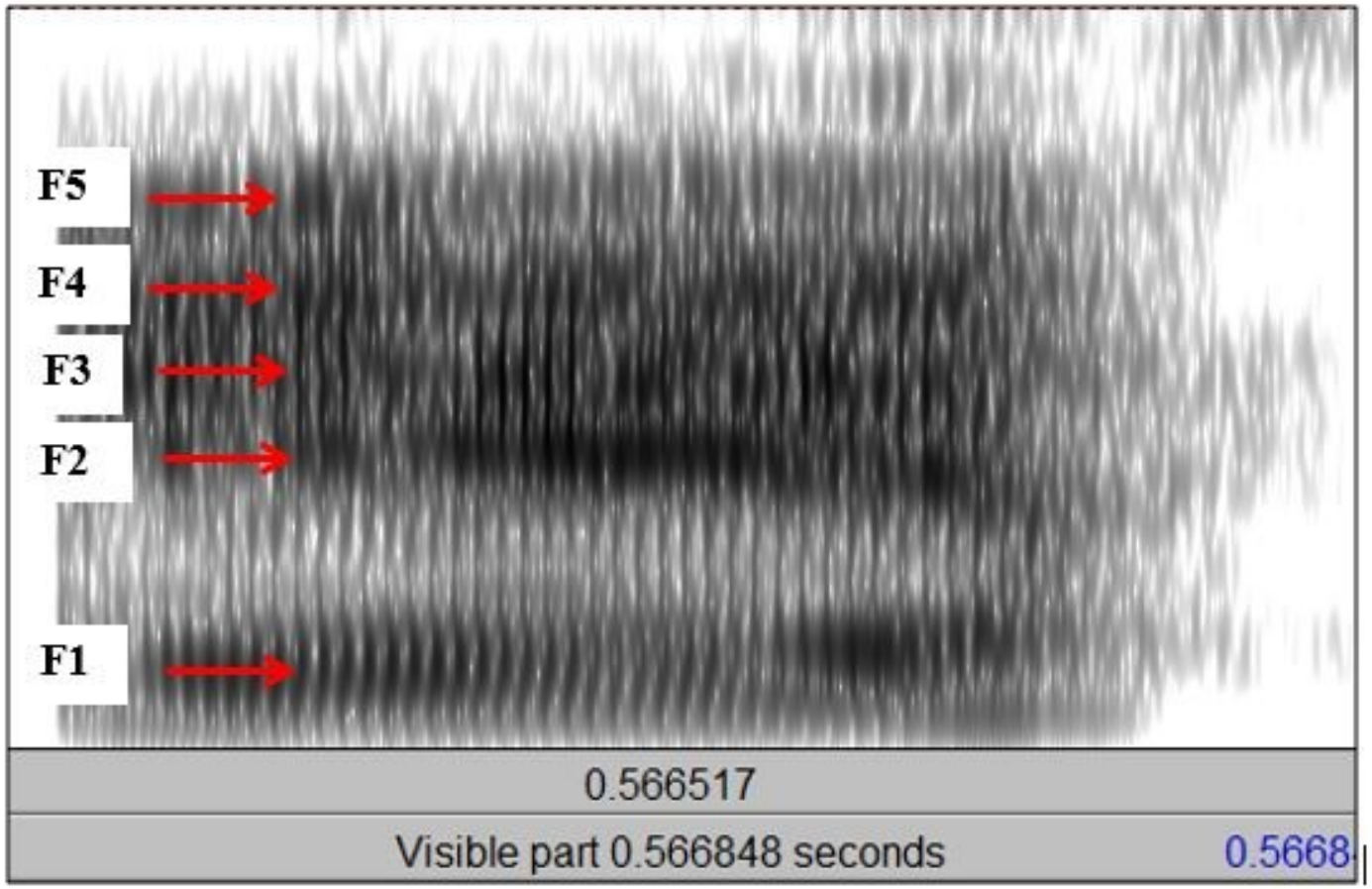
Таблица формант для гласных

- Чтобы синтезировать речевой сигнал, соответствующий определённой фонеме, необходимо настроить центральную частоту каждого полосового фильтра системы на соответствующую частоту форманты.
- Таблица частот формант для некоторых фонем

Фонема	Первая форманта, Гц	Вторая форманта, Гц	Третья форманта, Гц
«и»	270	2300	3000
«е»	400	2000	2550
«а»	660	1700	2400
«у»	640	1200	2400

Спектрограмма гласной [е]

- Видны пять частот с наибольшей энергией — формант

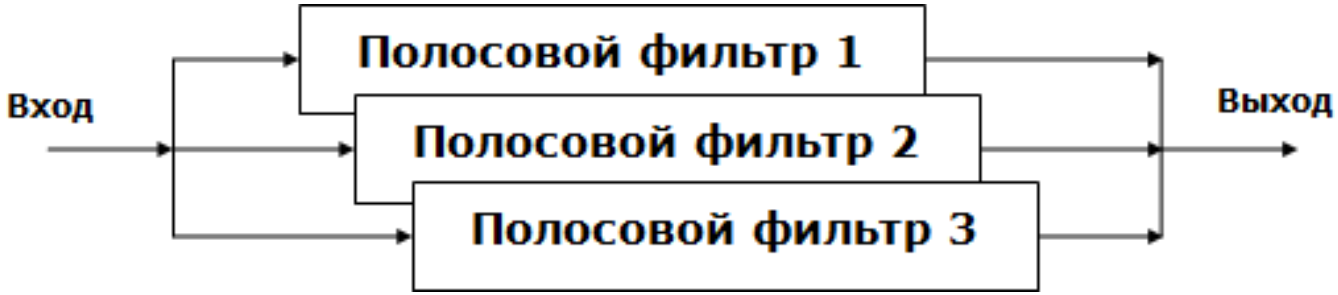


Формантный фильтр

- В основу структуры формантного фильтра заложена упрощённая модель голосового тракта.
- В соответствии с моделью, голосовой тракт представляет собой резонатор с несколькими пиками АЧХ, частоты которых определяют вид произносимой фонемы. Эти пики АЧХ получили название формант.
- Пример спектра фонемы «А»:

Формантный фильтр (2)

- Формантный фильтр создаёт формантные области в спектре входного сигнала с помощью нескольких параллельно соединённых полосовых или фазовых фильтров.
- Количество звеньев в схеме определяет порядок формантного фильтра.
- Схема формантного фильтра третьего порядка:



Определения синтеза речи

- ***Синтез речи*** — это
  - Формирование речевого сигнала по тексту.
  - Искусственное производство человеческой речи.
  - Восстановление формы речевого сигнала по его параметрам.
- Применение:
  - Информирование человека (в т.ч. в системе голосового управления)
  - Акустический диалог человека и компьютера

Способы синтеза речи

- Параметрический синтез
- Компиляционный синтез (конкатенативный, компилятивный) синтез
- Полный синтез по правилам
  - Формантный
  - Артикуляторный
  - На основе записанных отрезков речи

Параметрический синтез

- При параметрическом синтезе звуковой сигнал представлен определённым числом непрерывно изменяющихся параметров.
- Для формирования гласных звуков используется генератор тонального сигнала, для согласных — генератор шума.
- Метод обычно применяют для записи голоса в музыкальных композициях, и чаще речь идет даже не о чистом синтезе голоса, а, скорее, о модуляции.
- Является последней фазой в вокодерных системах.
- Параметрический синтез целесообразно применять в тех случаях, когда набор сообщений ограничен и изменяется не слишком часто.

Компиляционный синтез

- Метод компиляционного синтеза основывается на составлении текстов из заранее записанного "словаря" элементов.
- Размер элемента системы должен быть не менее слова.
- Обычно запас элементов ограничивается несколькими сотнями слов, а содержание синтезируемых текстов — объёмом словаря.
- Используется в различных справочных службах (продажа билетов, прогноз погоды) и технике, требующей оснащения системами речевого ответа: говорящие часы, навигаторы и др.

Полный синтез по правилам

- Полный синтез речи по правилам может воспроизводить речь по заранее неизвестному тексту. Базируется на запрограммированных лингвистических и акустических алгоритмах. Элементы человеческой речи не используются.
- Реализуется путём моделирования речевого тракта, применения аналоговой или цифровой техники. Причём в процессе синтеза значения параметров и правила соединения фонем вводят последовательно через определённый временной интервал, например 5–10 мс.
- Подходы:
  - ***Формантный метод*** базируется на формантах — частотных резонансах речевой акустической системы. Моделируется работа речевого тракта человека, работающего как набор резонаторов. Универсальная и перспективная технология, но понимание результата синтеза требует подготовки.
  - ***Артикуляторный метод*** пытается доработать недостатки формантного путем добавления в модель фонетических особенностей произнесения отдельных звуков.

Синтез речи по правилам на основе отрезков речи

- Базисные единицы речи:
  - Микросегменты;
  - Аллофоны;
  - Дифоны — участки речевого сигнала, включающие в себя переходы между звуками;
  - Полуслоги — сегменты, содержащие половину согласного и половину примыкающего к нему гласного;
  - Слоги;
  - Единицы произвольного размера.
- Есть возможность синтеза речи по не заданному заранее тексту (чтение книги на лету).
- Трудно управлять интонационными характеристиками речи, так как характеристики отдельных слов могут изменяться в зависимости от контекста или типа фразы.
- Качество синтезированной речи несопоставимо с качеством речи естественной (на границах сшивки элементов могут возникать искажения).

Распознавание речи

Основные задачи

Некоторые технологии

Основные задачи

- Системы Interactive Voice Response (IVR) в колл-центрах
  - Биометрическая идентификация
  - Автоматическая маршрутизация звонка
  - Аналитика речи, поиск пауз, тормозов в интерфейсе, аналитика эмоционального состояния
  - Сбор оценок операторов колл-центра
- Персональные помощники
  - Распознавание поисковых запросов
  - Распознавание команд управления

Голосовая биометрия

- 4 фактора:
  - "кто вы"
  - "что у вас есть"
  - "что вы знаете"
  - "что вы делаете"
- Пассивный режим, пассивные системы — не зависят от текста, не проявляют себя (слушают).
- Активный режим, активные системы — зависят от текста, взаимодействуют с пользователем.

Зачем нужна голосовая биометрия?

- Сокращение времени на аутентификацию пользователя с 23 секунд в ручном режиме в центре обработки вызовов (Call Center) до 5 секунд в автоматическом.
- Повышение лояльности пользователей (и, как следствие, доходов от них) в результате отказа от необходимости запоминать всем известные ответы на "секретные" вопросы, помнить PIN-код для входа в систему или отвечать на вопросы назойливого сотрудника банка (ваши ФИО, дата вашего рождения, номер карты и т.п.).
- Снижение числа сотрудников центра обработки вызовов за счет автоматической обработки многих простых вопросов (время работы офиса в праздники, ближайший офис или банкомат, тарифы и т.п.).
- Снижение числа мошеннических операций.
- Снижение времени на ожидание правильного сотрудника, который поможет ответить звонящему.
- Рост продуктивности работников компании и центра обработки вызовов.

Поставщики решений голосовой биометрии

- Некоторые поставщики голосовой биометрии:
  - Agnitio
  - Auraya Systems
  - Authentify
  - KeyLemon
  - Nuance
  - ValidSoft
  - Verint Systems
  - VoiceTrust
  - VoiceVault
  - ЦРТ — Центр речевых технологий (https://www.speechpro.ru/)

Распознавание говорящего по свободной речи

- Пассивный режим
- Идентификация мошенников в режиме тихого прослушивания звонящего/говорящего и идентифицирующие его речь, ничем себя не выдавая.
- Поэтому пассивные системы проще в использовании, но и требуют больших ресурсов для своей реализации.

Распознавание говорящего по заранее определённым фразам

- Активный режим (система «выдаёт» себя, управляя диалогом).
- Требуют большего участия пользователя.
- Сложно идентифицировать мошенника.

Голосовой отпечаток

- *Голосовой отпечаток* — некая уникальная для человека запись, характеризующая голос в целом. Длительность записи 1–2 минуты.
- Несколько различных голосовых отпечатков формируют профиль голоса.
- При создании отпечатка могут использоваться опорные точки в речи (переходы между звуками), высота звуков, акцентированные звуки, темп речи; косвенно учитываются физиологические особенности звукового тракта, горла, глотки, носа; особенности произношения слов и звуков, физические характеристики голоса.
- Распознавание на основе отпечатка может длиться 5–15 секунд.

Интеллектуальные голосовые помощники

- Интеллектуальные голосовые помощники:
  - Алиса (Yandex)
  - iOS Siri (Apple/Nuance)
  - Google Assistant
  - Amazon Alexa

— Сири, почему у меня не ладится с женщинами?
— Я Алиса

Речевые API

- Yandex Speech Kit https://tech.yandex.ru/speechkit/
  - Распознавание речи
  - Анализ речи (биометрическая информация)
  - Синтез речи
- Google Speech API https://cloud.google.com/speech/
  - Распознавание речи (110 диалектов языков)
  - Фильтрация неуместной лексики
  - Контекстно-зависимое распознавание

Yandex SK

- Пример удачного распознавания:

```
<recognitionResults success="1">
   <variant confidence="0.69">твой номер 212-85-06</variant>
   <variant confidence="0.7">твой номер 213-85-06</variant>
</recognitionResults>
```

- Пример неудачного распознавания:

```
<recognitionResults success="0"/>
```

Yandex SK (2)

- Оценка биометрических параметров: пол, возраст, язык
- Возрастная группа задается буквой латинского алфавита:
  - 'c' — ребенок (до 14 лет). Для этой группы пол не указывается;
  - 'y' — подросток (14–20 лет);
  - 'a' — взрослый (20–55 лет);
  - 's' — пожилой (старше 55 лет).
- Пол задается буквой 'm' (male) или 'f' (female).

Yandex SK (3)

- Пример результатов распознавания биометрии

```
{
tag: 'gender', class: 'female',
confidence: 0.3632163107395172
}
{
tag: 'language', class: 'ru',
confidence: 0.8913142085075378 // --> Наиболее вероятный язык речи - русский.
}
```

Yandex SK (4)

- Настройки синтеза фразы:
  - Диктор
  - Эмоции в голосе
- Например:

```
tts.speak(
   'Меня зовут Вася',
   {
        speaker: 'zahar',
        emotion: 'neutral',
        stopCallback: function () {…}
   }
)
```

Google Speech API

- Пример распознавания:


```
{
	"result":[
	{
		"alternative":[
		{
			"transcript":"this is a test",
			"confidence":0.97321892
		},
		{
			"transcript":"this is a test for"
		}
		],
		"final":true
	}
	],
	"result_index":0
}
```

Распознавание записи по аудиофрагменту

TODO

Что почитать

- Русская фонетика
  - Попов М.Б. Фонетика современного русского языка: Учебник / Учебно-методический комплекс по курсу «Фонетика современного русского языка». — СПб.: Филологический факультет СПбГУ, 2014. — 303 с.
  - https://ru.wikipedia.org/wiki/%D0%A0%D1%83%D1%81%D1%81%D0%BA%D0%B0%D1%8F_%D1%84%D0%BE%D0%BD%D0%B5%D1%82%D0%B8%D0%BA%D0%B0
  - Учебник по фонетике русского языка http://www.speech.nw.ru/Manual/menu.html
  - Зализняк А.А. Древне-новгородский диалект. — М.: Школа «Языки русской культуры», 1995. — 720 с.
  - Шушарина И. А. Хронология фонетических изменений в русском язык https://rujaz.narod.ru/istgram/IG_hronologya_fonetika.html
  - Ч и Ц сидели на крыльце https://dzen.ru/a/ZPgafor-oW6DOGuD
- Говорящая машина Фабера, или Механизм, воспроизводивший звуки человеческой речи
  - https://nwm.at/vse-ob-avstrii/nauka/govoryashhaya-mashina-fabera-ili-mehanizm-vosproizvodivshij-zvuki-chelovecheskoj-rechi
- Вокодер https://eomi.ru/electronic/vocoder/
- Синтезатор речи с открытым исходным кодом RHVoice http://tiflo.info/rhvoice/
- Алиса. Как Яндекс учит искусственный интеллект разговаривать с людьми https://habrahabr.ru/company/yandex/blog/339638/
  - Распознавание речи от Яндекса. Под капотом у Yandex.SpeechKit https://habrahabr.ru/company/yandex/blog/198556/
- Avery Li-Chun Wang An Industrial-Strength Audio Search Algorithm
  - https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf
  - https://github.com/bmoquist/Shazam
  - https://github.com/leonardltk/Shazam-An-Industrial-Strength-Audio-Search-Algorithm-