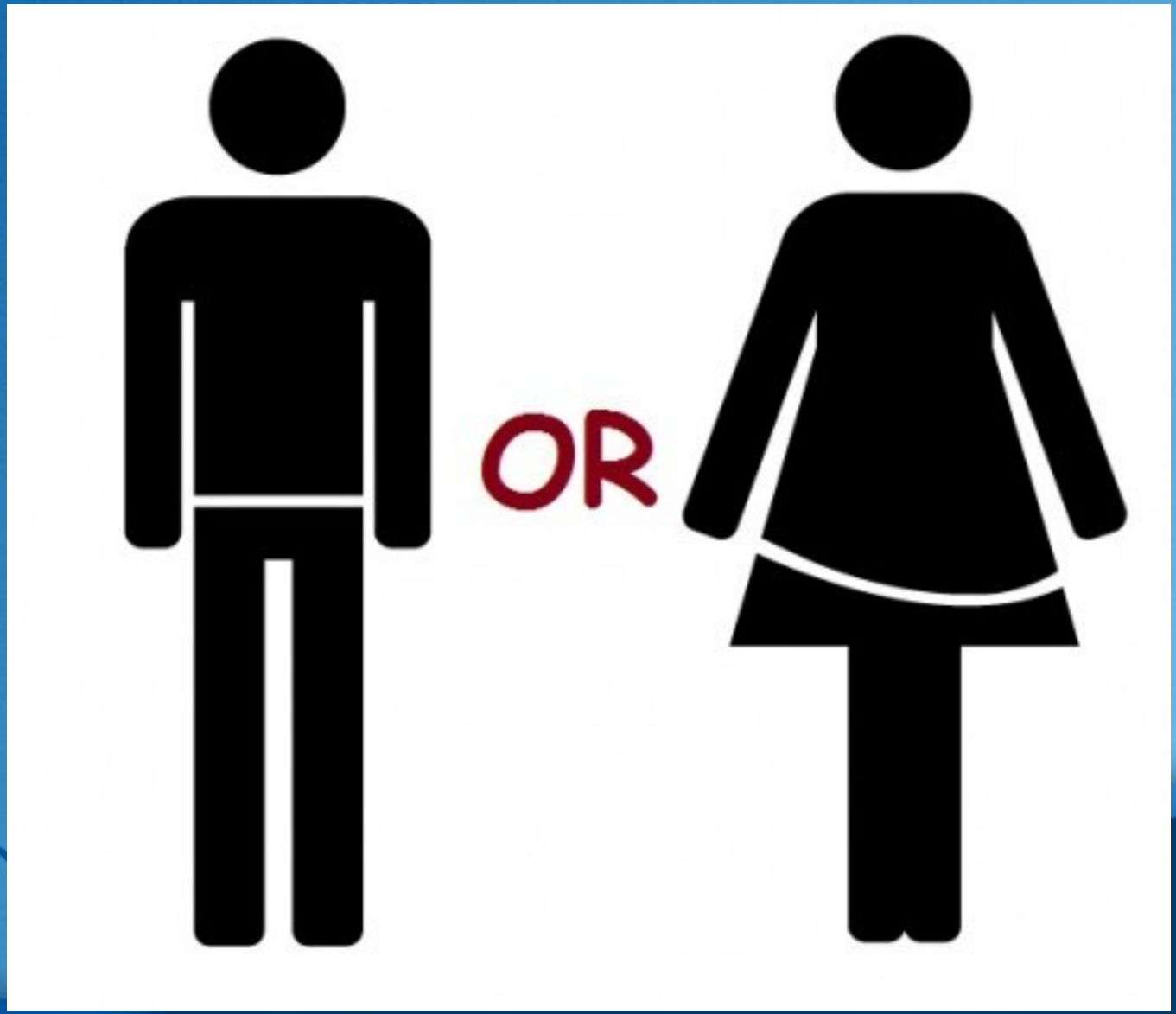
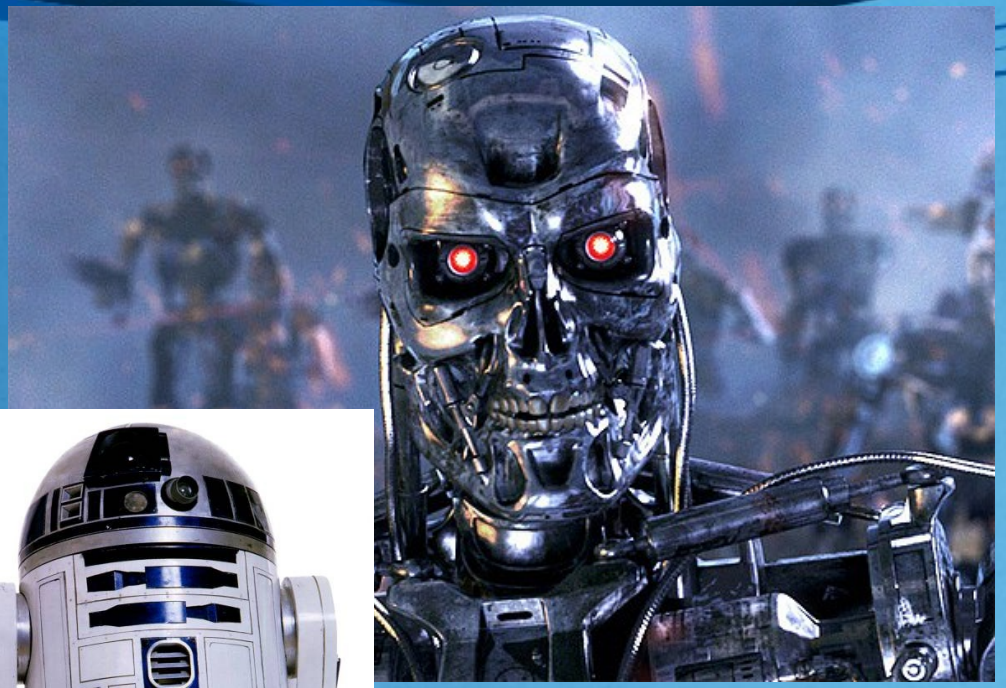
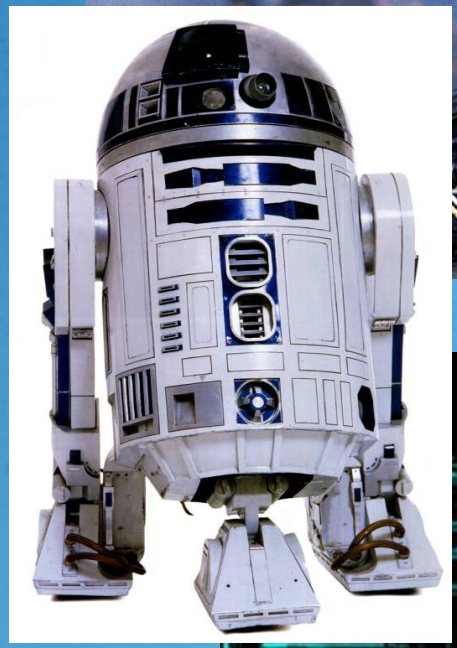
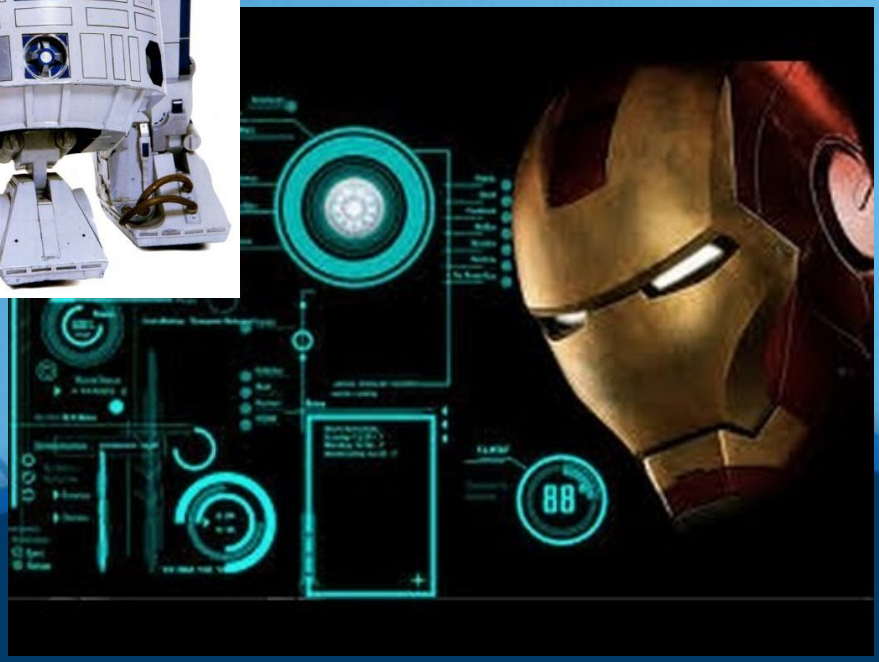
Идентификация пола человека по его голосу



- В наши дни все большую популярность набирает перспектива создания дружественного ИИ
- Не исключено, что уже через пару десятков лет различные роботы будут такими же жителями Земли как и люди
- В идеале они смогут видеть, слышать и ощущать все вокруг себя подобно человеку

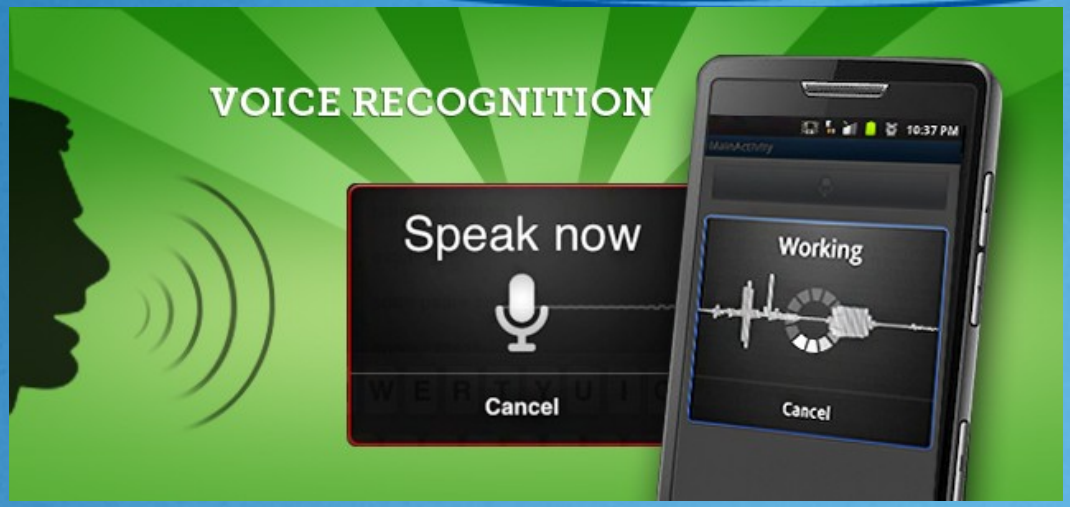


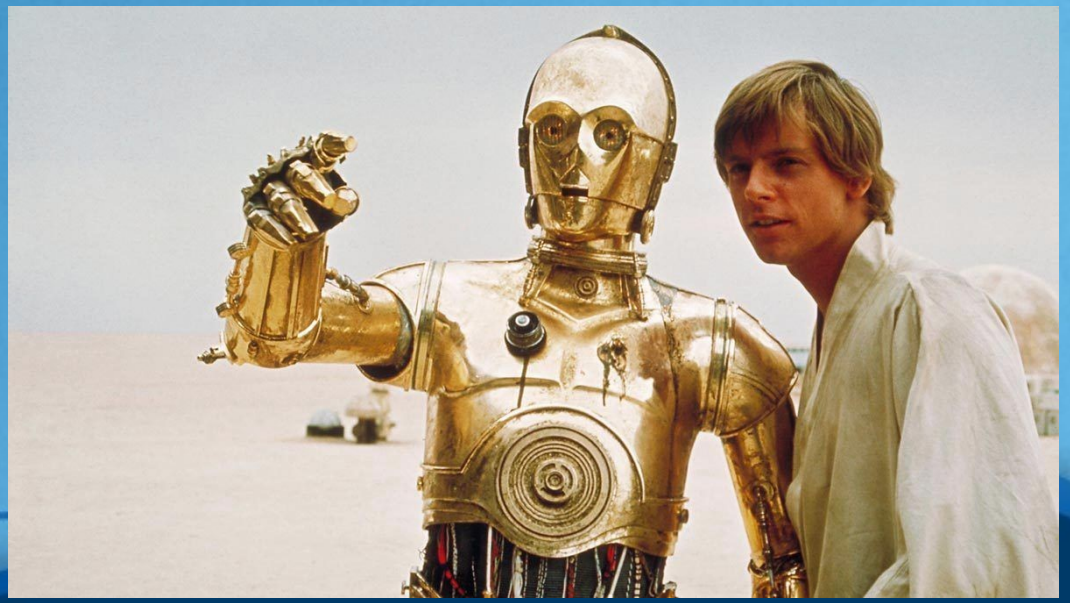




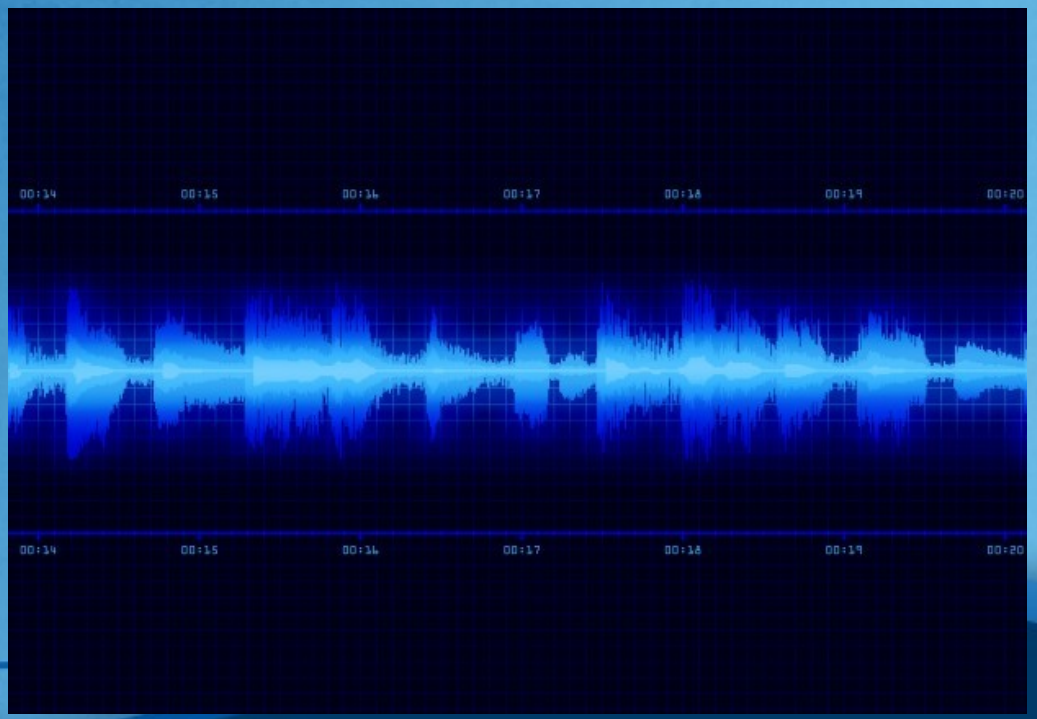
- **Одной из ключевых тем в этом вопросе является способность роботов распознавать голос**

- **Возможность установить пол говорящего человека значительно повышает точность распознавания эмоций, возраста, а также улучшает работоспособность систем идентификации личности**

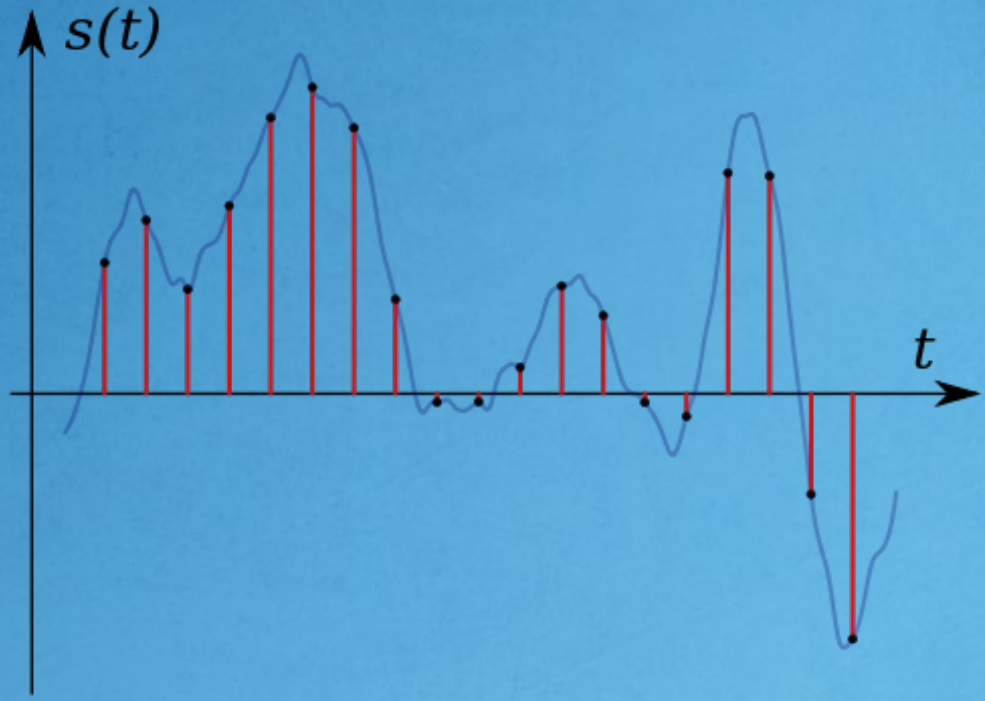




- **Все, что связано с машинным распознаванием звука да и вообще с работой со звуком стало доступно благодаря такой вещи как цифровая обработка сигналов**



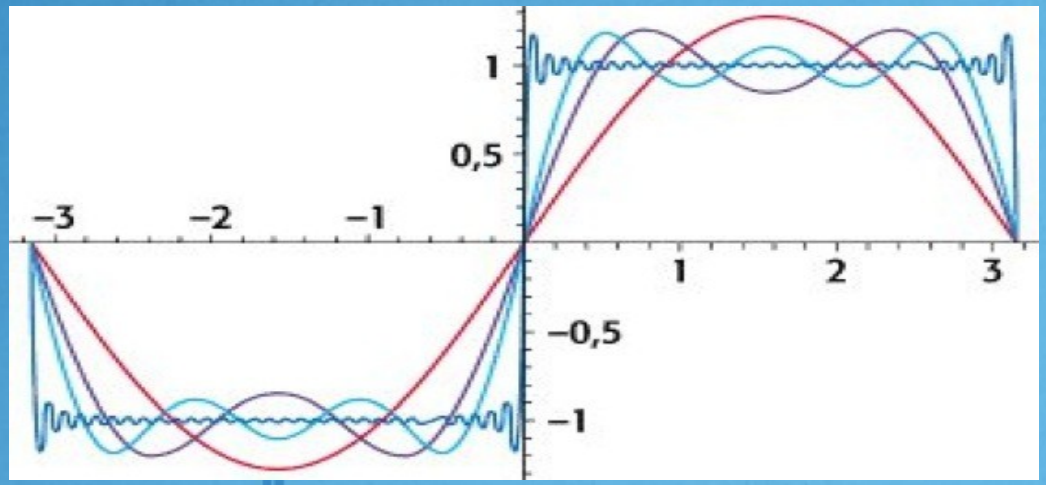
- **Чтобы записать сигнал без потерь нужно брать значение его амплитуды с частотой в два раза превышающей самую высокочастотную составляющую**



- **Чтобы лучше понять вышесказанное достаточно вспомнить про то, что для удобства анализа сигналов их раскладывают на синусоиды с помощью быстрого преобразования Фурье**

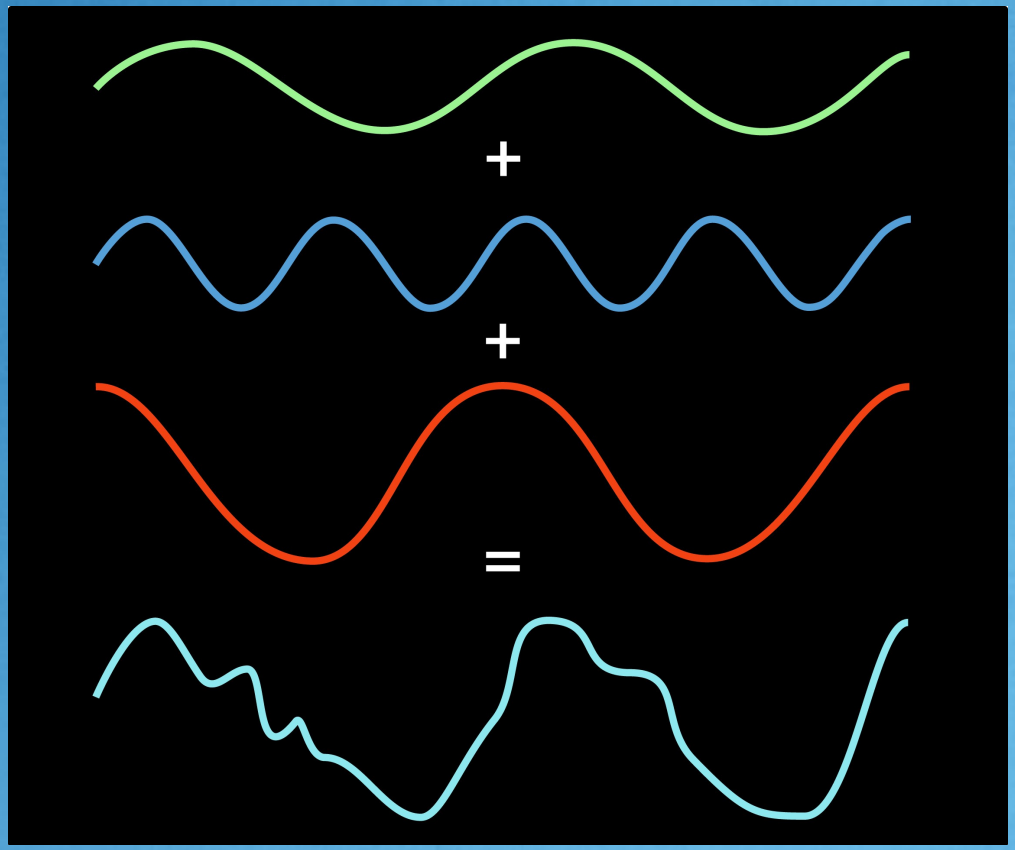
$$g(t) = a_0 + \sum_{m=1}^{\infty} a_m \cos\left(\frac{2\pi mt}{T}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi nt}{T}\right)$$

$$= \sum_{m=0}^{\infty} a_m \cos\left(\frac{2\pi mt}{T}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi nt}{T}\right)$$



- Именно таким образом работает человеческое ухо

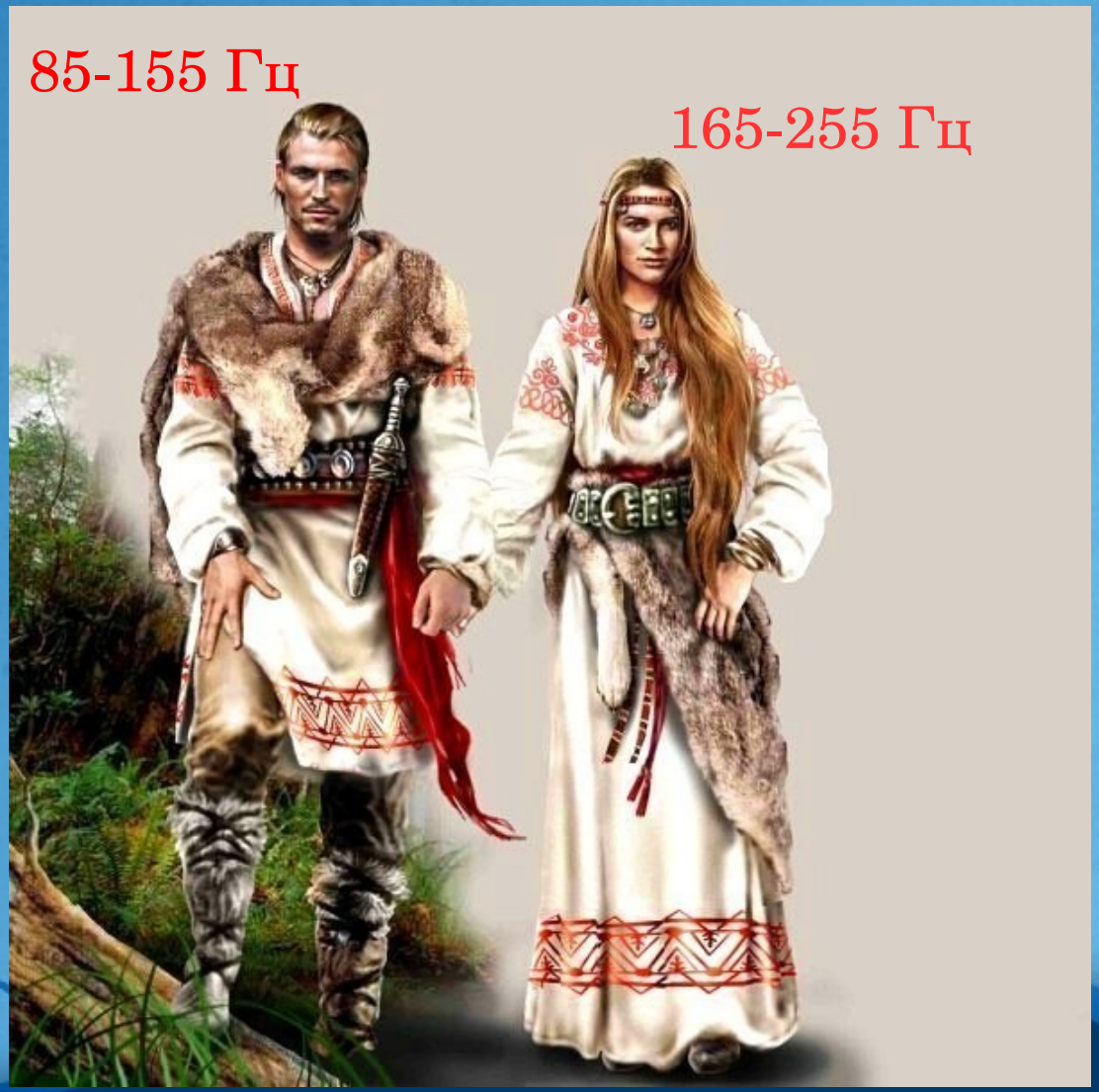
- Кроме того, в результате мы получаем спектр сигнала(значения амплитуд синусоид), который описывает множество нужных для классификации признаков

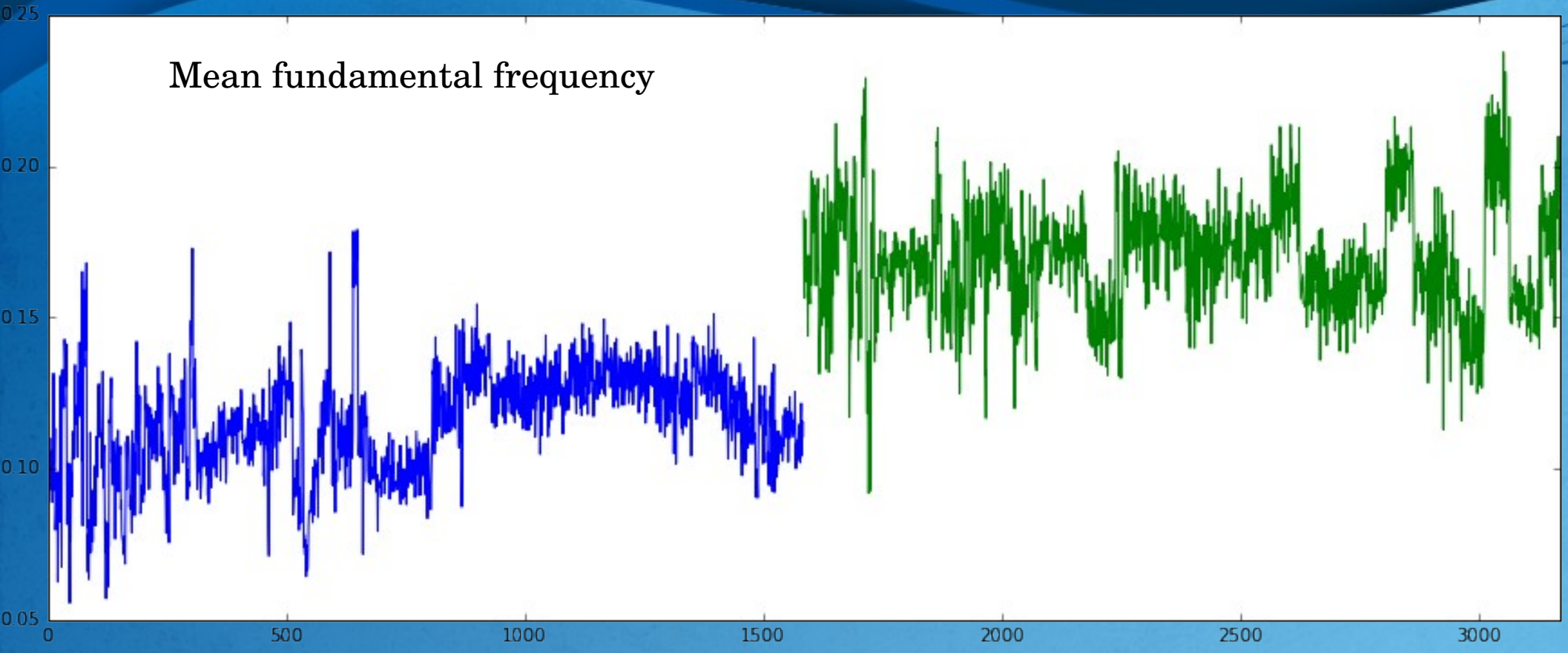


- **meanfreq - среднее арифметическое всех частот присутствующих на аудиодорожке**
- **sd - среднеквадратическое отклонение частоты**
- **median - медиана частот**
- **Q25 - первый квартиль частот**
- **Q75 - третий квартиль частот**
- **IQR - диапазон частот между первым и третим квартилем**
- **skew - ассиметрия спектра частот**
- **kurt - эксцесс спектра частот**
- **sp.ent - спектральная энтропия**
- **sfm - спектральная плоскостность**
- **mode - самая часто встречаемая частота(мода спектра)**
- **centroid - "центр масс" спектра частот**
- **meanfun - среднее значение фундаментальной частоты**
- **minfun - минимальное значение фундаментальной частоты**
- **maxfun - максимальное значение фундаментальной частоты**
- **meandom - среднее значение доминантной частоты**
- **mindom - минимальное значение доминантной частоты**
- **maxdom - максимальное значение доминантной частоты**
- **dfrange - диапазон доминантной чатоты**
- **modindx - индекс частотной модуляции**

Перейдем к анализу имеющихся данных

- Как говорит Википедия: "Голос типичного взрослого мужчины имеет фундаментальную частоту (нижнюю) от 85 до 155 Гц, типичной взрослой женщины от 165 до 255 Гц."





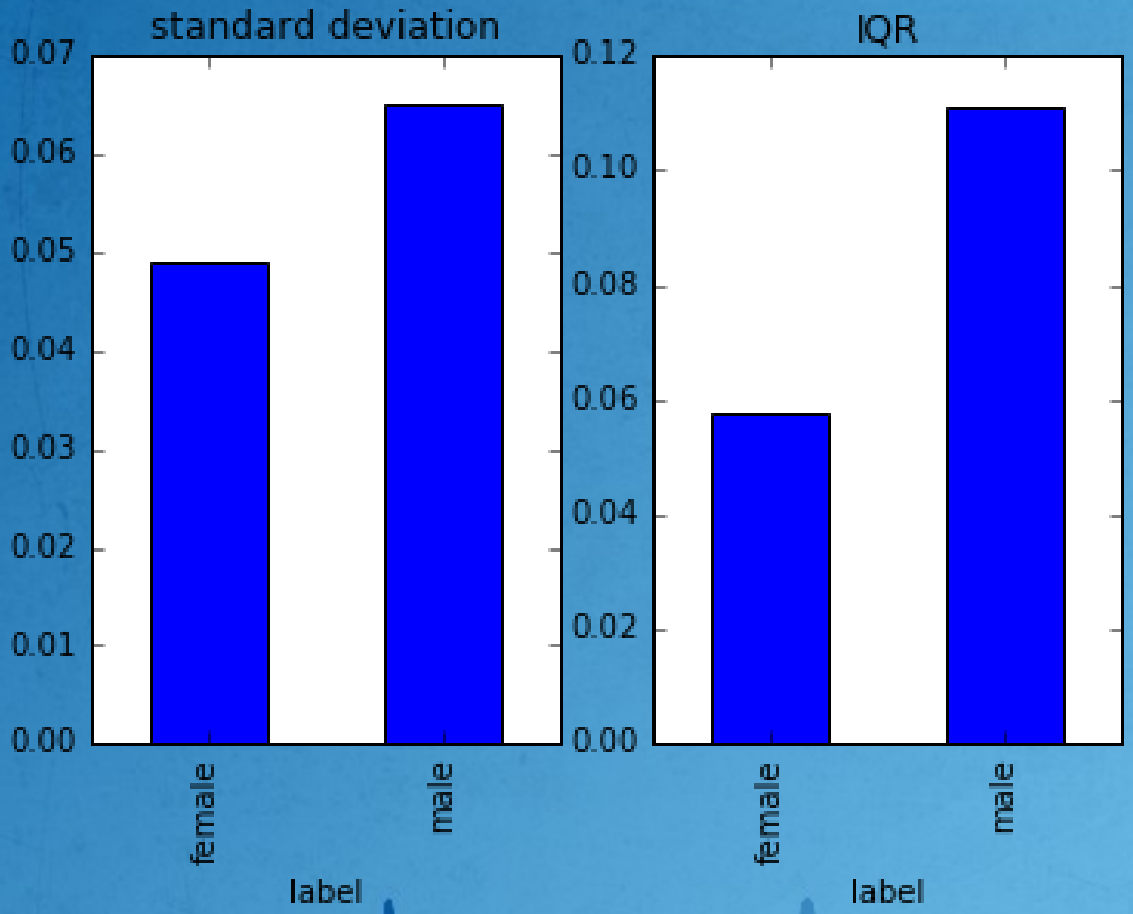
male

female

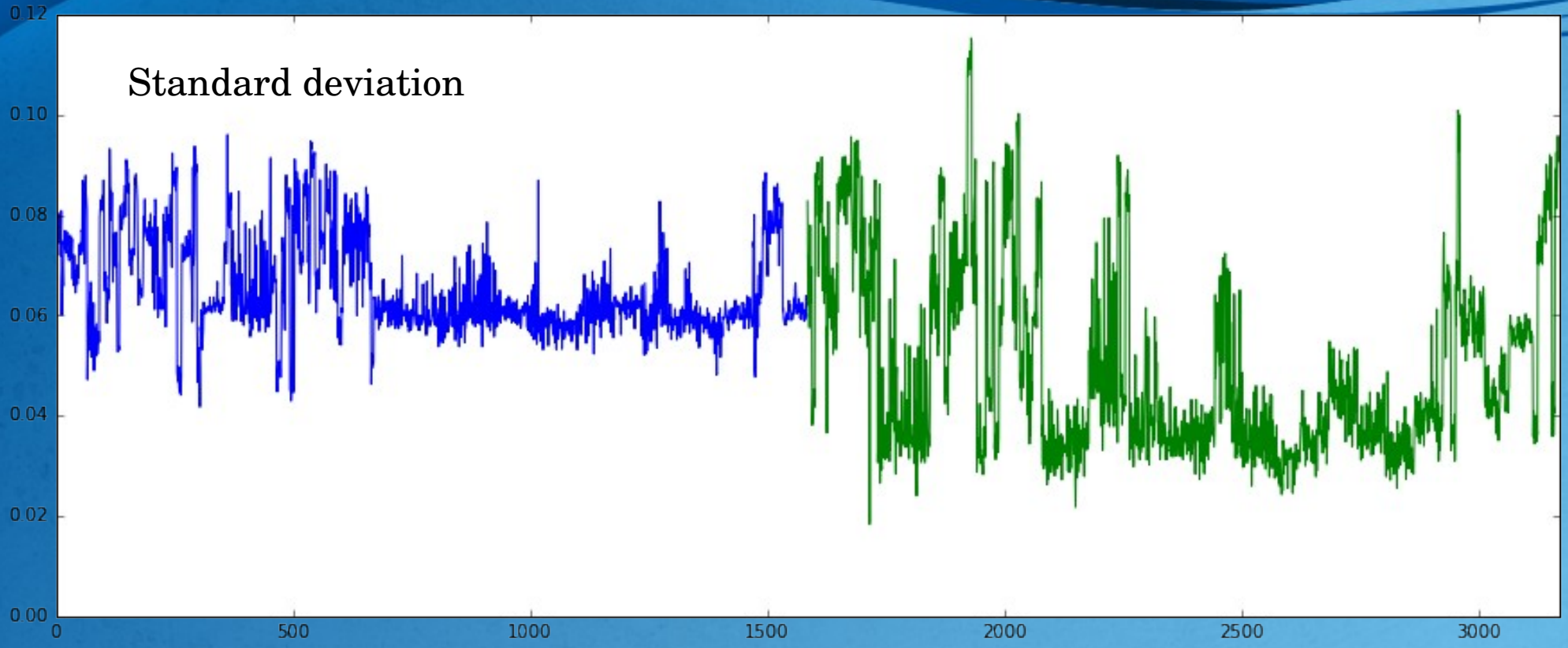
Предположение подтвердилось очень хорошо!

Известно также, что у мужчин гораздо шире частотный спектр. Это может быть связано с тем, что во время добывания еды древним охотникам нужно было копировать голоса различных птиц и зверей, из-за чего в процессе эволюции они получили богатый набор воспроизводимых звуков.

Это наводит на мысль о том, что **межквартильный диапазон** и **стандартное отклонение** также должны играть важную роль при классификации.



Межквартильный диапазон(IQR) ведет себя так как и предполагалось. На счет стандартного отклонения нельзя быть так уверенным, давайте построим график.



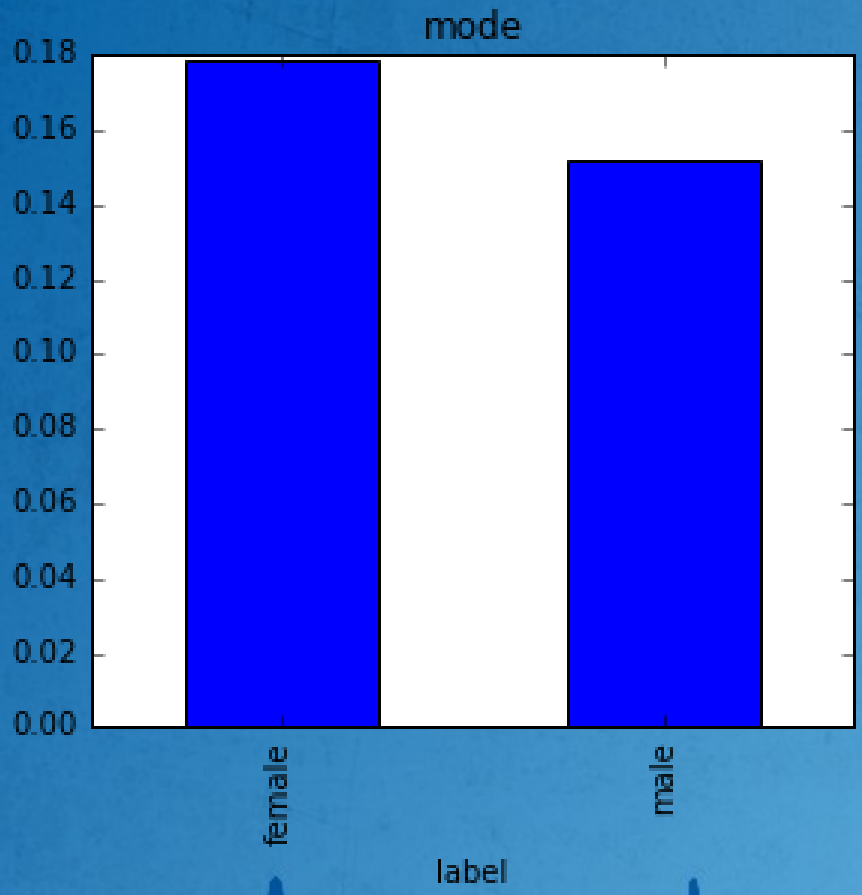
male

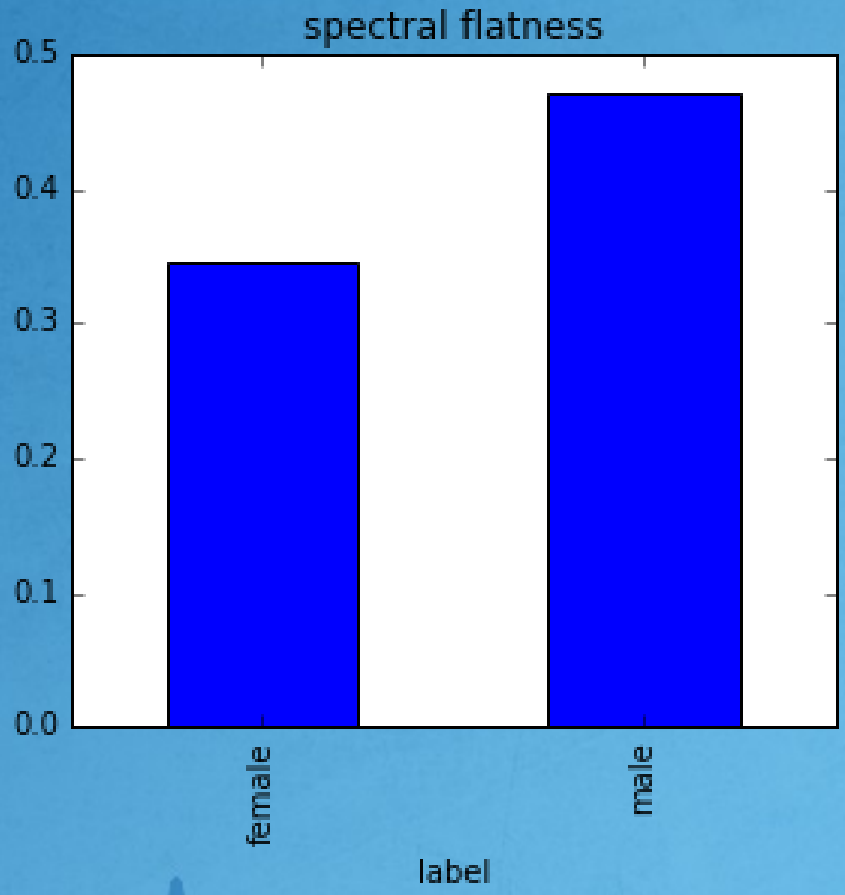
female

В основном результаты для разных полов отличаются, но среди женщин присутствует много нетипичных показателей.

- Также хотелось бы обратить внимание на такие характеристики, как **мода** и **спектральная плоскостность**. Известно, что в голосе женщин преобладают более высокие частоты, а значит по значению моды можно отличить мужской голос от женского
- Спектральная плоскостность характеризует чистоту голоса

- Именно поэтому для различных объявлений намного чаще выбирают женщин из-за того что их легче услышать в толпе.

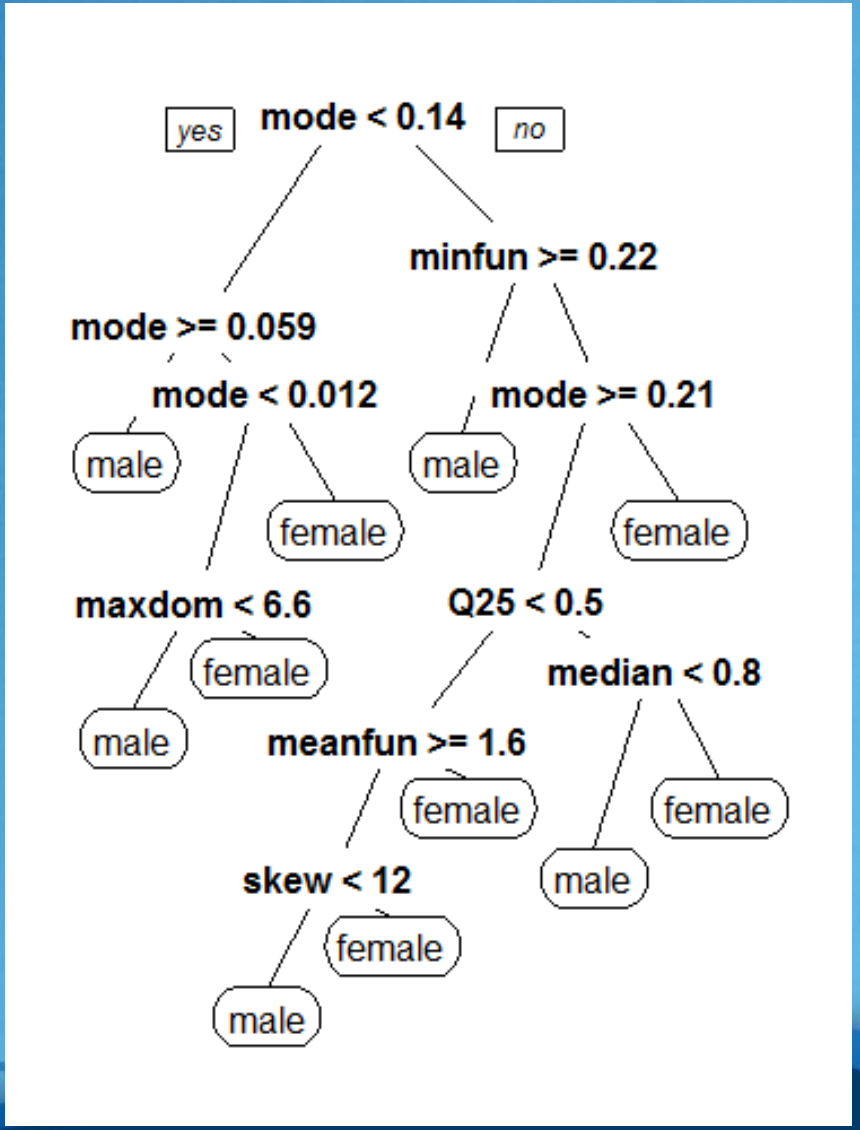


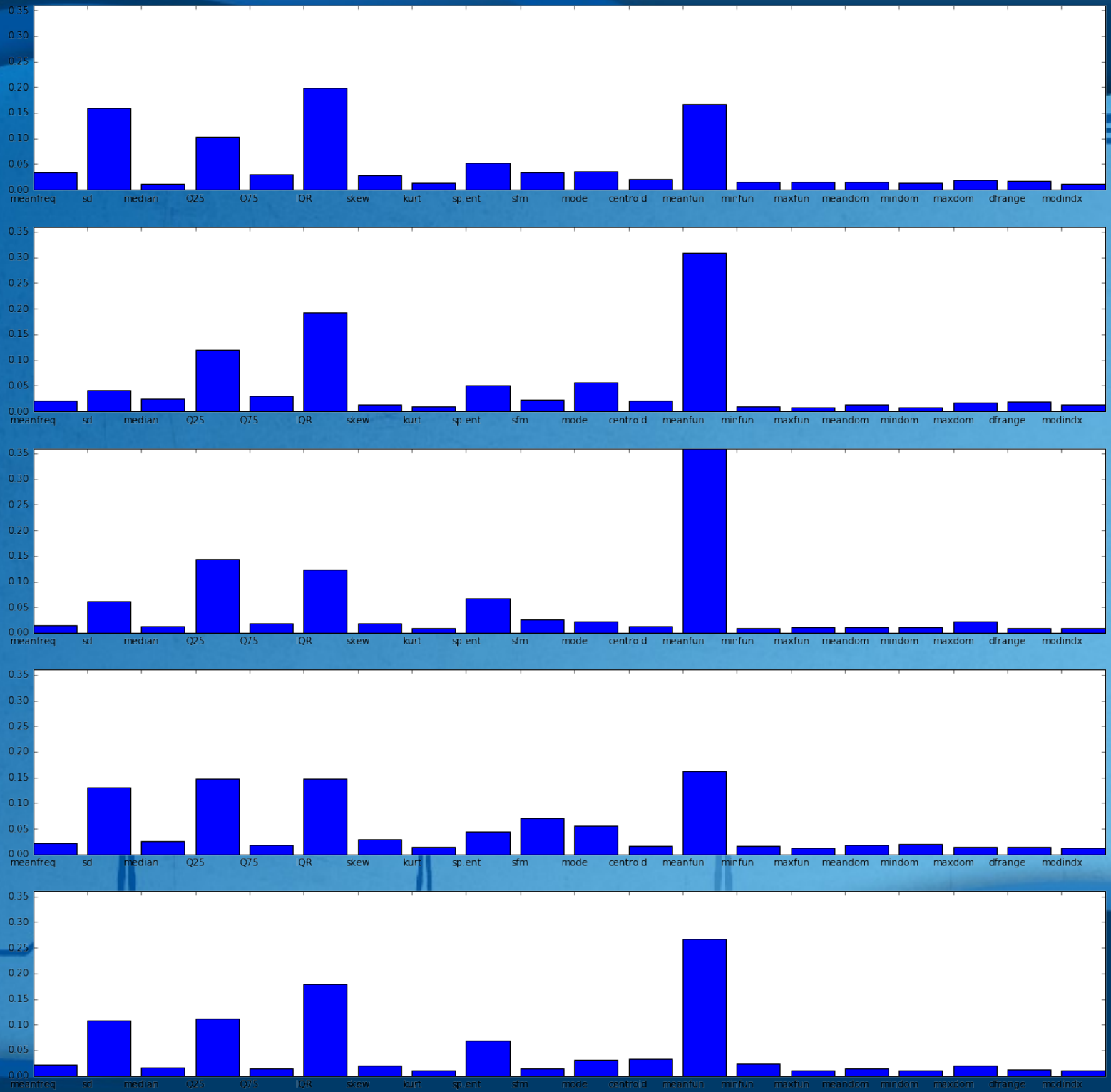


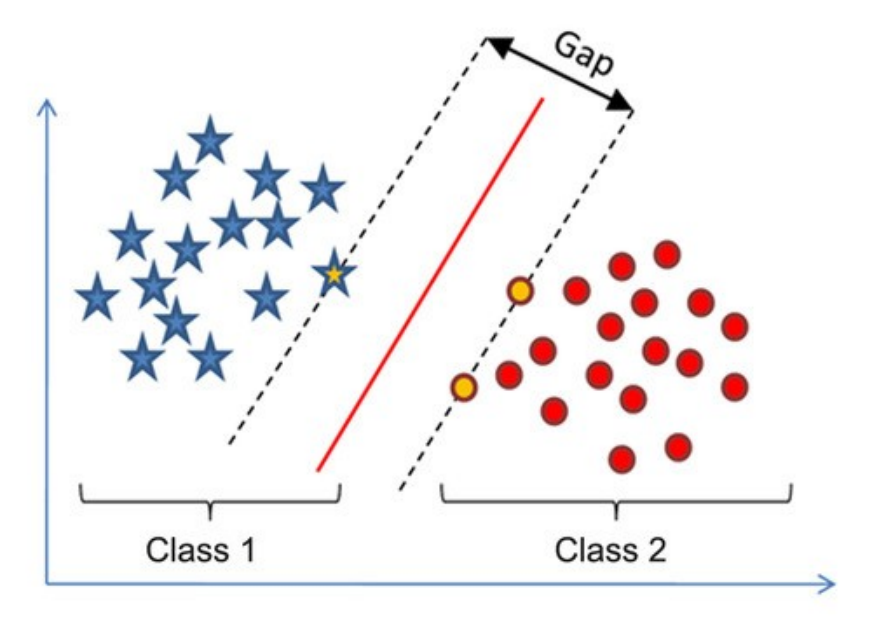
Как видите статистика говорит то же самое.

- **Итак, мы только что исследовали как влияют на результат часто упоминаемые в литературе характеристики голоса**
- **Но имеется еще около десятка признаков которые могут быть важными, но с которыми нельзя быть так уверенным, как с предыдущими**
- **Для того чтобы узнать какие признаки важные, какие - нет и насколько одни важнее других давайте воспользуемся одним из методов оценки важности каждого признака и посмотрим что получилось**

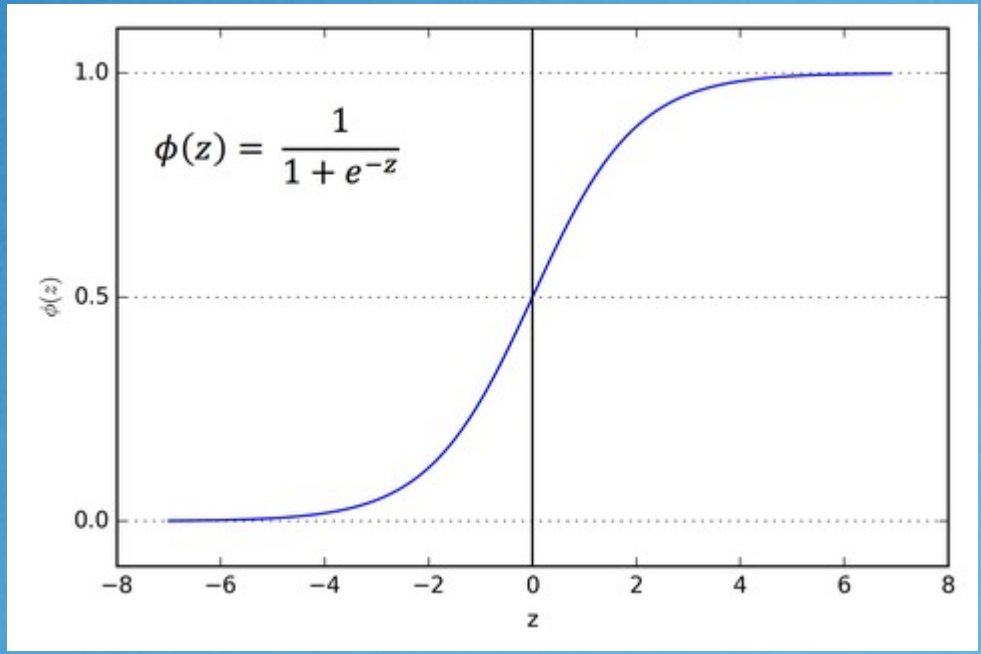
- **Для поставленной цели я решил использовать метод оценки, предоставляемый классом ExtraTreesClassifier**
- **Вкратце этот метод использует деревья решений, в которых каждый узел представляет собой условие, где значение определенного признака сравнивается с каким-то пороговым значением**
- **Для оценки важности посчитывается через узлы с какими признаками классификатор проходит чаще всего**



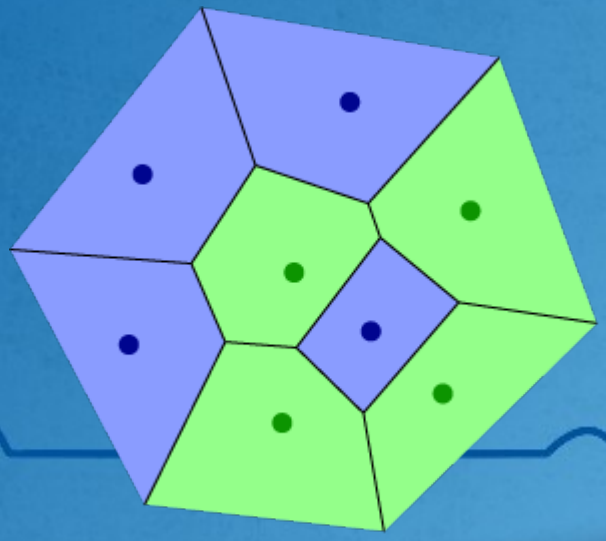

0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
meanfreq
sd
median
Q25
Q75
IQR
skew
kurt
sp.ent
sfm
mode
centroid
meanfun
minfun
maxfun
meandom
mindom
maxdom
dfrange
modindx



Support Vector Machine
(linear and rbf kernel)

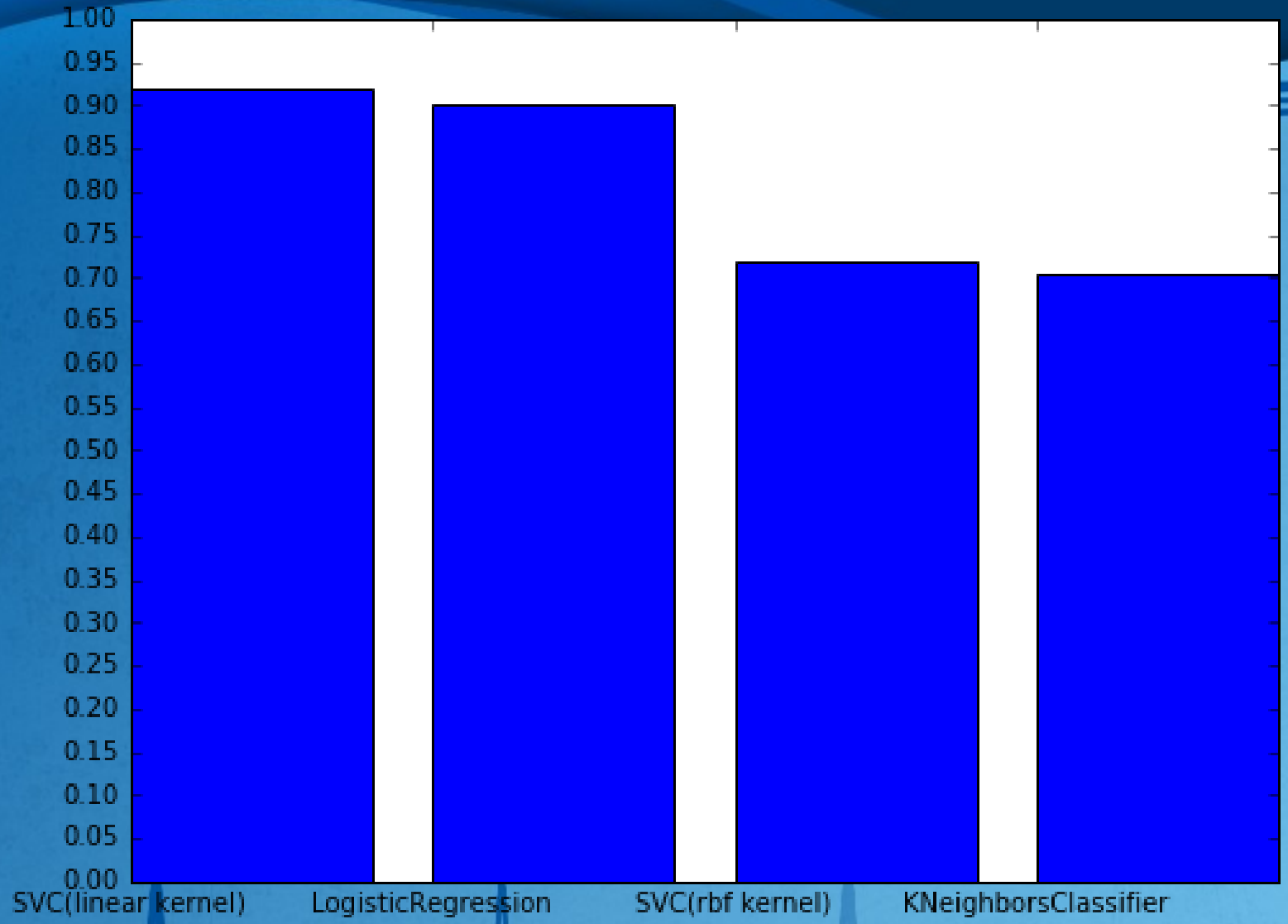


Logistic Regression



kNN

Эти алгоритмы машинного обучения я использовал для классификации

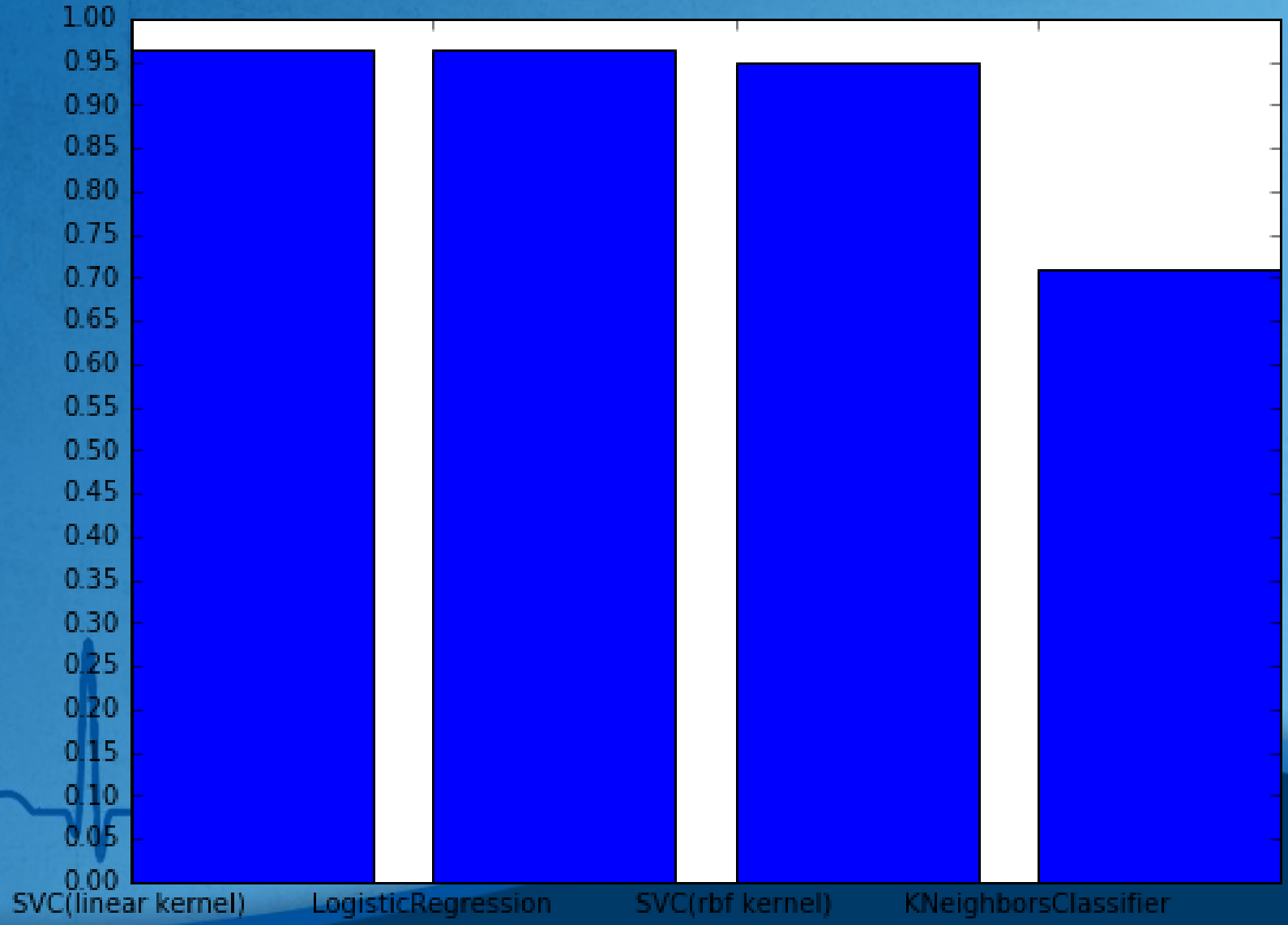


А это результаты скользящего контроля алгоритмов на тренировочной выборке. Вероятно, даны линейно разделимы. Этим можно объяснить такой отрыв линейных разделителей от нелинейных.

- **В предыдущем способе есть один недостаток, который обязательно нужно исправить. Дело в том, что в вопросе производительности таких алгоритмов как kNN и SVC немаловажную роль играют гиперпараметры, значения которых нужно подобрать в ходе эксперимента. Я же в этом случае использовал их значения по умолчанию.**

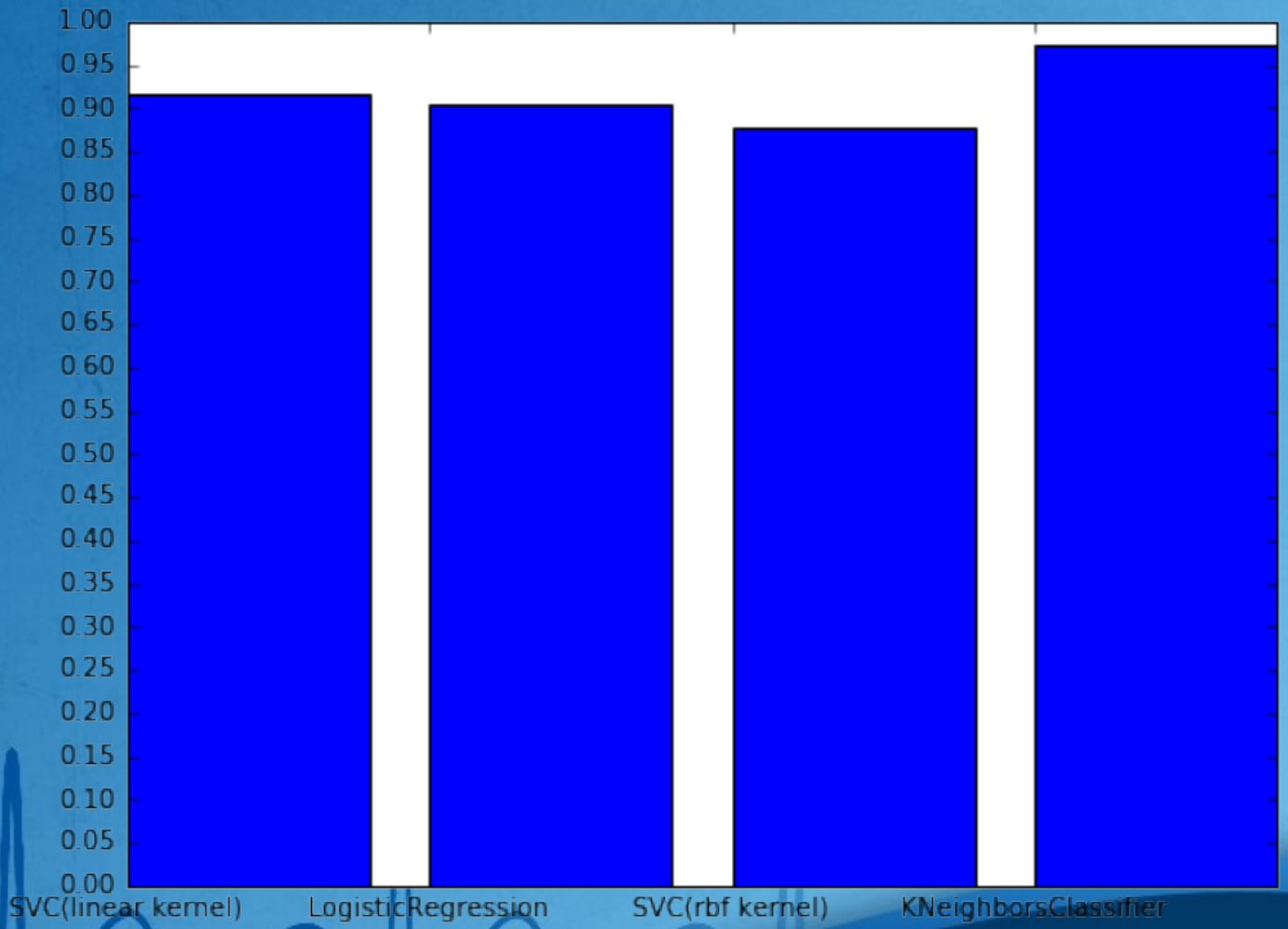
Модель	Стандартные значения	Оптимальные значения
SVC(linear)	C = 1.0	C = 12.0
SVC(rbf)	C = 1.0, gamma = 0.005	C = 10000.0, gamma = 0.004
kNN	k = 5	k = 7
Logistic	penalty = 'l2'	penalty = 'l1'

После подбора оптимальных значений гиперпараметров имеем следующие результаты на тестовой выборке:

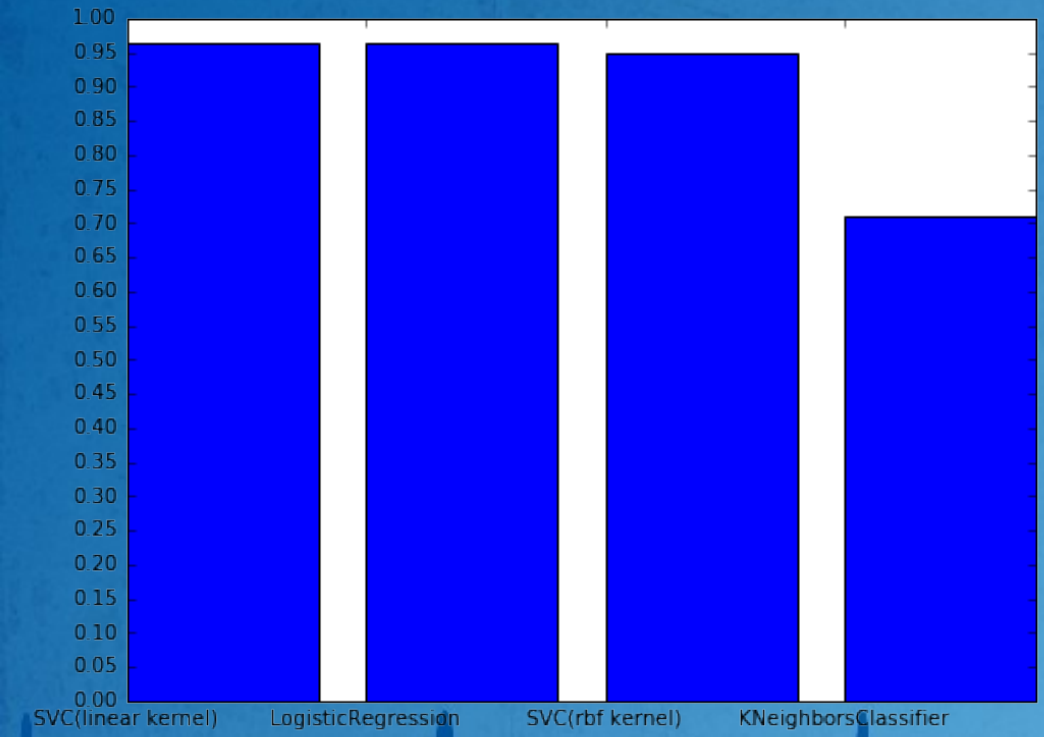


- Вроде бы все проверено и на этом можно заканчивать, но настораживает такая низкая эффективность kNN-модели.
- А что если малозначимые признаки в спорных ситуациях не облегчают выбор а только мешают и из-за похожести их значений "ближайшими соседями" становятся образцы разных классов.
- Чтобы проверить это утверждение давайте отбросим те признаки, которые ExtraTreesClassifier посчитал малозначимыми и посмотрим что получится.

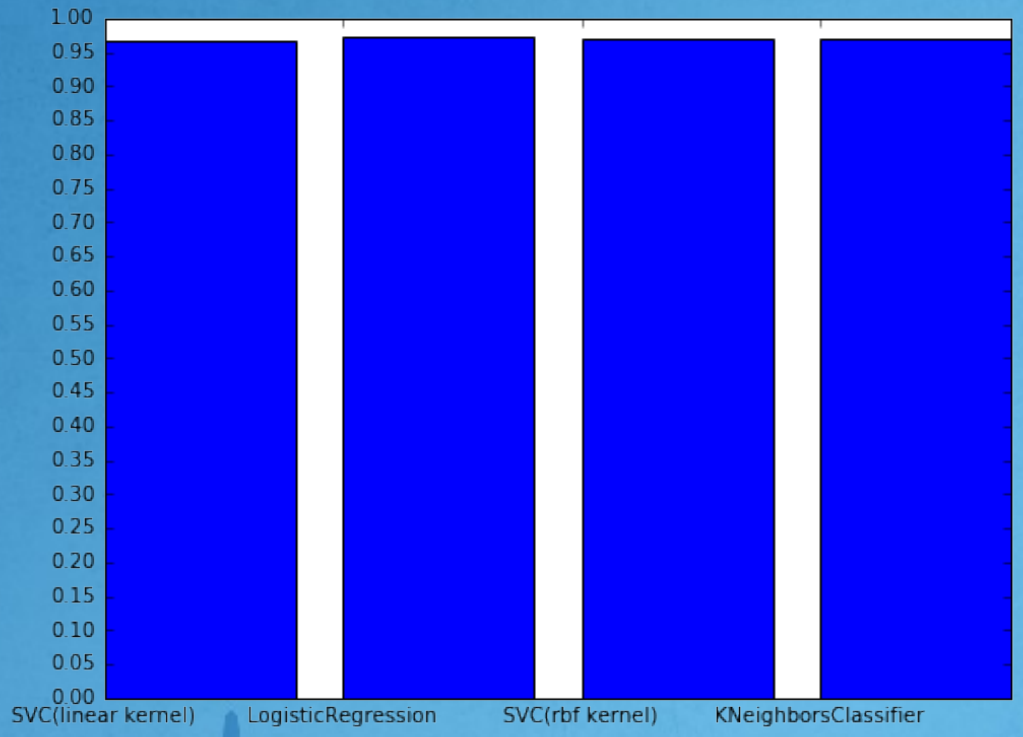
И вот что получаем в итоге:



Подобрав оптимальные гиперпараметры для моделей получаем следующее(на тестовой выборке):



20 признаков



7 самых важных признаков

Итоги

- Модели, которую можно считать победителем как таковой нету. Но ради точности можно заметить что лучшее поведение показала логистическая регрессия(с очень маленьким отрывом)
- Удалось достигнуть точности в 97% на тестовой выборке