

## СПЕКТРАЛЬНЫЙ АНАЛИЗ РЕЧИ

*Есманчик И. А., Лапич В. Д.*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Иванов М.А. – канд. физ.-мат. наук, доцент*

Предложен простой признак, позволяющий различить мужские и женские голоса. Нужно иметь усреднённые спектры мужского и женского голосов. Для спектра неизвестного голоса предлагается построить вектор амплитуд в шестимерном пространстве опорных частот и вычислить углы между этим вектором и усреднёнными мужскими и женскими векторами. Значения этих углов позволяют надёжно различить два типа голосов.

Искусственный интеллект помогает управлять предприятиями, строить дома и промышленные объекты, выращивать урожаи, лечить и учить людей. В медицине он помогает выявлять редкие заболевания, в астрономии – изучать далёкие галактики. Искусственный интеллект открывает невероятные возможности, берёт на себя множество рутинных функций, которые отнимают драгоценное время у специалистов, делая множество процессов проще и быстрее. Среди актуальных задач в этой области – обеспечение распознавания и синтеза речи. Мы поставили перед собой узкую задачу: найти простой способ, который можно было бы использовать для распознавания мужских и женских голосов.

Мужской голос, как правило, имеет более низкие частоты, соответствующие звучанию малой октавы (130,81 Гц (включительно) до 261,63 Гц), в то время как у женского голоса частоты более высокие (261,63 Гц (включительно) до 523,25 Гц). Это помогает идентифицировать различия и классифицировать голоса по двум категориям.

Используя программу Adobe Audition, мы записали по 10 файлов небольшой длительности с мужскими и женскими голосами. Для каждого из этих файлов мы выполнили Быстрое преобразование Фурье и получили 2 набора по 10 спектров.

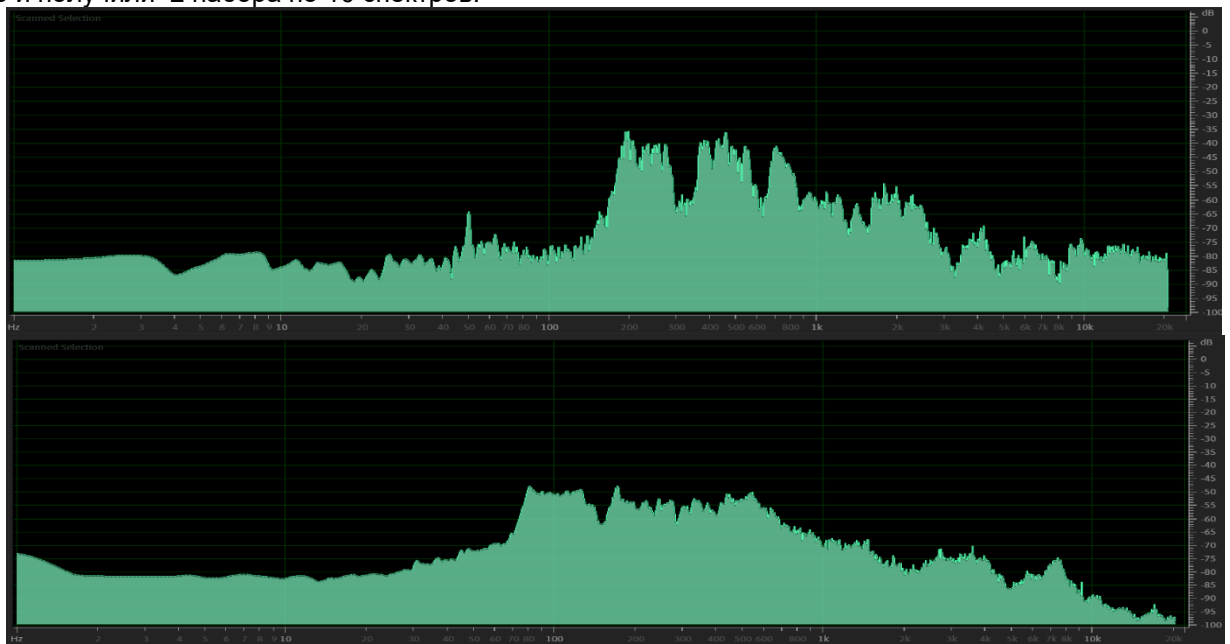


Рисунок 1 – Спектрограмма женского (вверху) и мужского (внизу) голоса, амплитуды показаны в дБ

Наша основная идея была в том, чтобы анализировать небольшой набор амплитуд для базовых частот, принадлежащих полосе частот, включающей обе эти октавы. Мы выбрали 6 базовых частот: 100, 130, 170, 210, 260 Гц – для мужского голоса и 260, 340, 420, 520, 700 Гц – для женского.

Из амплитуд на этих частотах мы сформировали 6-мерные векторы:  $\vec{a}_k = (a_1, a_2, \dots, a_6)$ ,  $k = 1, 2, \dots, 10$ , где  $a_k$  – значение амплитуды в дБ. Для работы нам понадобилось преобразовать их к линейному масштабу по следующей формуле:

$$a_i \rightarrow 10^{\frac{a_i}{10}}$$

Взяв данные из наших спектров, каждый из этих векторов мы нормировали на единичную длину (чтобы не зависеть от громкости сигнала) и определили средний вектор для спектра мужских и женских голосов ( $\langle \vec{a}_M \rangle$ ,  $\langle \vec{a}_Ж \rangle$ ), их также нормировали на единичную длину.

В этом 6-мерном пространстве косинус угла между единичными векторами определили как:

$$\cos \alpha = \vec{A} \cdot \vec{B} = A_1 \cdot B_1 + A_2 \cdot B_2 + \dots + A_6 \cdot B_6$$

Вычислили углы между отдельными векторами из группы женских голосов и средним вектором для этой же группы, а также между средним единичным вектором из мужской группы. То же самое проделали для группы мужских голосов. Результаты занесены в таблицы 1 и 2, f - женский, m - мужской.

Таблица 1 – значения косинуса угла для мужских голосов

№	<Am>	<Af>
Am1	0.838502	0.673121
Am2	0.788392	0.734875
Am3	0.984774	0.713500
Am4	0.839292	0.691576
Am5	0.882821	0.774307
Am6	0.838932	0.568122
Am7	0.792553	0.615862
Am8	0.892477	0.535614
Am9	0.875091	0.591640
Am10	0.927251	0.602261

Таблица 2 – значения косинуса угла для женских голосов

№	<Af>	<Am>
Af1	0.850276	0.531451
Af2	0.898994	0.467377
Af3	0.968035	0.613548
Af4	0.904857	0.412263
Af5	0.983135	0.526979
Af6	0.936733	0.476838
Af7	0.980094	0.356607
Af8	0.966084	0.657035
Af9	0.924664	0.637154
Af10	0.859676	0.652968

Найдем средний угол для каждой серии:

$$\langle \alpha_{mm} \rangle = 29.023^\circ; \langle \alpha_{fm} \rangle = 49.205^\circ$$

$$\langle \alpha_{ff} \rangle = 20.800^\circ; \langle \alpha_{mf} \rangle = 57.524^\circ$$

Видим, что углы  $\langle \alpha_{mm} \rangle$  и  $\langle \alpha_{ff} \rangle$  заметно меньше углов  $\langle \alpha_{fm} \rangle$  и  $\langle \alpha_{mf} \rangle$ . Это означает, что значения этих углов являются достаточно надежным признаком для различения мужских и женских голосов.

Такой подход к анализу голосов может быть полезным для таких задач, как автоматическое распознавание голоса, классификация аудиофайлов по типу голоса (женскому или мужскому), в медицинской диагностике, а также в анализе аудиорекламы и в маркетинговых исследованиях. Описанный способ является простым в реализации, не требует большого времени вычислений и большого объема памяти.

**Список использованных источников:**

1. Преобразование Фурье: самый подробный разбор [Электронный ресурс]. – Режим доступа: <https://proglib.io/p/fourier-transform>.
2. Савельев И.В. Курс общей физики : учеб. пособие. Том 1 / И.В. Савельев. – М. : Изд-во Моск. ун-та : Наука, 1988. – 289 с
3. Нейросеть "Яндекса" научилась распознавать эмоции по голосу [Электронный ресурс]. – Режим доступа: <https://www.vesti.ru/hitech/article/3844216>.
4. Физический смысл БПФ [Электронный ресурс]. – Режим доступа: <https://slonpts.narod.ru/algo/math/furie/06.html>.