

УДК 004.934.2

Исследование применения метода определения базовых эмоций на основе отношения частот спектра звуковой волны

О. А. Близнюк, А. П. Шкарапута

Пермский государственный национальный исследовательский университет
Россия, 614990, г. Пермь, ул. Букирева, 15
olesiabliznuk77@gmail.com, shkaraputa@psu.ru

Исследуется возможность применения метода определения эмоционального состояния человека на основе сравнения отношений частот спектра звуковой волны. Рассматриваемый метод определения базовых эмоций (радость, печаль) основывается на поиске отношений частот, соответствующих мажорному и минорному ладу. Поиск данных соответствий осуществляется через определенные промежутки времени и их количество суммируется. На основе количественных характеристик делается заключение об эмоциональном состоянии говорящего человека. В результате анализа пятидесяти записей, среди которых (по мнению авторов) 25 были с радостными и 25 с печальными эмоциями, выявлены случаи, при которых базовые эмоции (радость, печаль), наиболее хорошо определяются с помощью рассматриваемого метода. В данной работе исследовались такие звуки как смех, плач, возгласы, печальные вздохи, а также записи из детских аудиосказок. При определении эмоций все записи были классифицированы по принципу: "определено", "не определено" и "определено неверно". "Определено" означало, что у записи с печальными эмоциями превосходил минорный лад, а у радостной – мажорный, "не определено": не было найдено ни одного музыкального лада или количество найденных мажорных и минорных ладов одинаковое, и "определено неверно": у печальных записей превосходил мажор, а у радостных – минор. В результате исследования показано, что печальная эмоция лучше всего определяется для плача, а радостная для смеха. В этих случаях эмоциональное состояние определялось примерно в семидесяти процентах.

Ключевые слова: классификация звуков; определение эмоций; частоты; звуковые волны; минор; мажор.

DOI: 10.17072/1993-0550-2017-4-86-91

Введение

Распознавание эмоционального состояния человека является одной из актуальных проблем современности в связи с развитием цифровых технологий. Решение проблемы распознавания эмоций востребовано в таких областях, как психология, робототехника (в том числе "умный дом"), криминалистика, медицина и прочее.

Эмоциональные состояния имеют очень разнообразные проявления. По степени интенсивности и продолжительности они могут

быть длительными, но слабыми (печаль), или сильными, но кратковременными (радость).

По субъективному переживанию все многообразие эмоций можно разделить на 2 категории: на эмоции положительного порядка, связанные с удовлетворением жизненных потребностей человека и потому доставляющие удовольствие, и эмоции отрицательного порядка, связанные с неудовлетворением жизненных потребностей и потому доставляющие неудовольствие.

Согласно американскому психологу Полу Экману [1], базовых эмоций семь:

1. радость (довольство);
2. удивление;

3. печаль (грусть);
4. гнев (злость);
5. отвращение;
6. презрение;
7. страх.

Существует множество способов определения эмоционального состояния человека. Например, метод опорных векторов [2], который относится к группе граничных методов классификации. Он определяет принадлежность объектов к классам с помощью границ областей. Или метод векторного квантования [3], задача которого ставится как минимизация искажения при замещении каждого вектора соответствующим кодовым вектором.

В работе [4], была показана принципиальная возможность использования определения базовых эмоций на основе отношений частотных характеристик звуковых волн.

Целью данной работы является определение, для каких звуков говорящего метод дает лучший результат.

1. Определение эмоционального состояния на основе характеристик звуковой волны

1.1. Описание метода

Предполагается, что звуковая волна, соответствующая радостному эмоциональному состоянию, содержит отношения, соответствующие мажорному музыкальному ладу, а для печального состояния – отношения соответствующие минорному музыкальному ладу.

Лад в музыке – это система отношений устойчивых и неустойчивых звуков и созвучий, которая работает на определенный звуковой эффект.

Ладов в музыке довольно много, основные из них – это мажор и минор. Мажор обозначает веселый, радостный музыкальный лад, а минор – грустный, печальный и мягкий.

В теории музыки признаком эмоций считается наличие отношения между частотами соответствующих интервалам, не превышающим октаву; это приводит к тому, что отношение составляет меньше двух. Так, мажорный лад составляют три звука (ноты) с соответствующими частотами и отношения между частотой первой ступени (основной звук с минимальной частотой) к частоте третьей ступени: 1,26 (большая терция), обратное соотношение – 0,79 для мажорного лада, для

минорного лада отношение первой ступени к третьей ступени: 1,19 (малая терция), а обратное соотношение – 0,84.

Приведем пример для мажорного и минорного ладов:

Ля - мажор	
Ля	440
до#	554.36
Ми	659.26
Ля - минор	
Ля	440
До	523.25
Ми	164,81

Так, в мажорном ладе присутствуют ноты, например, Ля и До дизь с частотами 440 и 554,36 соответственно, их отношения дают величины: 1,26 или 0,79. А в минорном ладе присутствуют: Ля и До, с частотами 440 и 523,25, отношения которых равны: 1,19 и 0,84.

Таким образом, если отношение частотных характеристик равно 1,26 или 0,79, то на данном отрезке звукового сигнала содержится мажор, а если 1,19 или 0,84 – минор.

Для проверки предположения звуковой сигнал разбивается по времени на некоторое количество отрезков, затем на каждом отрезке находят частоты трех максимальных амплитуд. Далее считаются отношения частот друг к другу внутри каждого блока (отрезка). На основании полученных количественных показателей делается вывод о присутствии мажора или минора.

Из рассмотренного выше примера видно, что данный метод достаточно прост для понимания и не требует больших знаний в области математики, а также не требует существенных временных затрат и дополнительных данных для анализа. Кроме того, он не нуждается в обучении как система, использующая метод опорных векторов.

1.2. Выбор данных

Для частотного анализа, который необходим в данном исследовании, был выбран программный продукт – Adobe Audition CS 2017, так как он позволяет выбрать необходимые участки звукового сигнала для анализа, имеет понятный интерфейс, а также имеет небольшую погрешность определения данных, связанную с переводом сигнала из аналогового в цифровой.

В работе [4] было исследовано 20 записей, в которых звуки не разделялись по группам. В данной работе рассматривается 50 записей, среди которых (на основании оценки авторов) 25 с радостными эмоциями и 25 – с печальными. Записи содержат такие эмоции, как смех, плач, восторг, вздохи, кроме того были проанализированы фразы из детских аудиосказок.

Наилучшая длина отрезков для спектрального анализа была определена опытным путем. Оказалось, что лучше всего брать промежутки длиной 1 сек., так как на них было чаще замечено присутствие соотношений мажорного или минорного музыкального лада.

Ниже приведем примеры спектральных характеристик звуковых волн смеха и плача, взятых в определенный момент времени (см. табл. 1 и 2).

Таблица 1. Спектральные характеристики для смеха

	1100	1400	2800
1100	1	1,272727	2,545455
1400	0,785714	1	2
2800	0,392857	0,5	1

Таблица 2. Спектральные характеристики для плача

	1680	1200	1020
1680	1	0,714286	0,607143
1200	1,4	1	0,85
1020	1,647059	1,176471	1

В данных таблицах представлены частоты волн спектра с максимальными амплитудами и их отношения для смеха и плача соответственно.

Аудиосказки были разбиты на отрезки, соответствующие слогам в произнесенном слове или фразе. Данные снимались в конце слога. С точки зрения звукового сигнала – это участки наибольшей интенсивности звука, то есть с наибольшими амплитудами.

Внутри каждого отрезка проводилось 3 измерения: фиксировались три частоты, которым соответствовали наибольшие амплитуды.

Результаты полученных измерений и являлись данными для дальнейшего анализа.

1.3. Расчеты

Для вычисления погрешностей, возникающих при преобразовании звукового сиг-

нала в цифровой, а также возникающих при снятии данных, применялась формула относительной погрешности

$$\delta_x = \frac{\Delta x}{x},$$

где Δx – абсолютная погрешность измерения, x – значение частоты волны. Тогда относительная погрешность соотношения:

$$\delta_{\frac{x}{y}} = \delta_x + \delta_y.$$

В данном случае абсолютная погрешность измерения равна

$$\Delta x = \begin{cases} 5 \text{ Гц}, & x \leq 1000 \text{ Гц} \\ 10 \text{ Гц}, & x > 1000 \text{ Гц}. \end{cases}$$

Для предотвращения коллизий было принято решение, что если отношение с учетом погрешности соответствует как мажору, так и минору, то необходимо выбрать тот ладовый окрас, значение которого ближе к полученному соотношению.

На основе полученных значений, проводились расчеты: вычислялись отношения частот внутри каждого блока, в которых проводился поиск радостной и грустной эмоций.

Для наглядности удобно представить результаты в виде диаграмм, на которых для каждой фразы столбцами изображены значения, соответствующие количеству соотношений мажорного или минорного лада в записи.

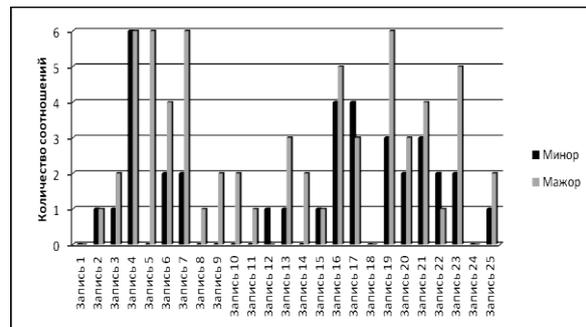


Рис. 1. Диаграмма значений для радостной эмоции

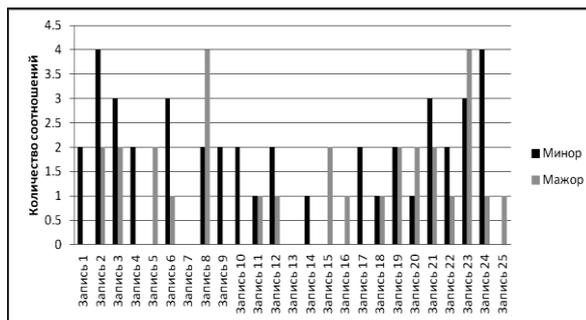


Рис. 2. Диаграмма значений для печальной эмоции

Для радостных эмоций (рис. 1) видно, что шестнадцать записей указывают на явное превосходство соотношений мажорного лада над минорным, три записи содержат больше соотношений минорного лада и шесть одинаковое количество минорного и мажорного.

Практически аналогичные результаты были получены для печальной эмоции (рис. 2), у тринадцати записей преобладают минорные лады, у семи записей преобладает мажор и у пяти выявлено равное количество минорного и мажорного ладов.

На основании представленных результатов можно сделать вывод, что примерно 60 % радостных и грустных эмоций определяются верно. Однако метод, основанный на соотношениях частотных характеристик, не всегда показывал необходимый ладовый окрас соответствующей фразы. Полагаем, это могло произойти по следующим причинам:

1. Человек, голос которого был записан, "не попадал" в необходимые ноты (плохо интонировал), в результате чего погрешность значительно выросла.
2. Значения некоторых частот в произносимых человеком фразах находятся в разных октавах (соотношения больше 2).
3. Погрешность, связанная с кодированием звука: в частотном анализе применялся FFT-алгоритм быстрого вычисления дискретного преобразования ряда Фурье. В данном алгоритме число анализируемых отсчетов сигнала имеет решающее значение для вида спектра. Чем меньше это число, тем реже сетка частот, по которым FFT раскладывает сигнал, и тем меньше деталей по частоте видно на спектре [5]. На рис. 3 представлены различия при разных значениях числа отсчетов (сверху – меньшее, снизу – большее).

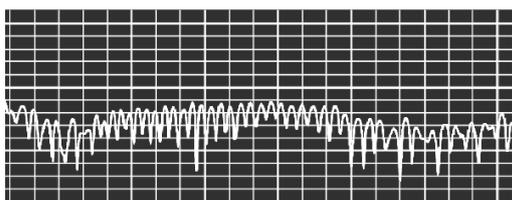
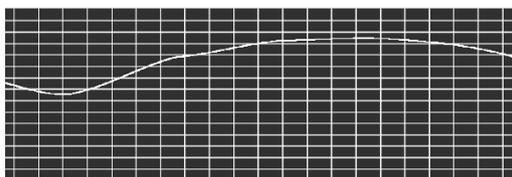


Рис. 3. Частотный анализ при разном числе отсчетов

2. Исследование звуков для различных групп эмоционального состояния

Чтобы определить те звуки, для которых данный метод работает наиболее корректно, все записи были разделены на группы. Записи с радостными эмоциями были разделены на такие группы, как смех, веселые (радостные) разговоры из аудиосказок и возгласы, а записи с грустными (печальными) эмоциями были так же разбиты на 3 группы: плач, грустные разговоры из аудиосказок и грустные вздохи.

Количество записей в каждой группе были следующими:

Таблица 3. Звуки для различных групп эмоционального состояния

Радостные звуки		Грустные звуки	
Смех	12	Плач	10
Радостные (веселые) разговоры из аудиосказок	6	Грустные разговоры из аудиосказок	7
Возгласы	7	Вздохи	8

После анализа было подсчитано, какое количество записей каждой группы было определено верно и неверно, а также сколько не было определено (т. е. не было найдено ни одного музыкального лада или количество найденных мажорных и минорных ладов было одинаково). Полученные результаты представлены в табл. 4, 5.

Таблица 4. Определение эмоций в группах для радостных эмоций

	Определено верно	Не определено	Определено не верно
Смех	9	0	3
Радостные (веселые) разговоры из аудиосказок	2	3	1
Возгласы	4	1	2

Таблица 5. Определение эмоций в группах для печальных эмоций

	Определено верно	Не определено	Определено не верно
Плач	7	1	2
Грустные разговоры из аудиосказок	2	3	2
Вздохи	5	0	3

На основе табличных данных был вычислен процент правильного определения эмоционального состояния внутри каждой группы.

Полученные результаты представлены на диаграммах ниже (рис. 4–5).

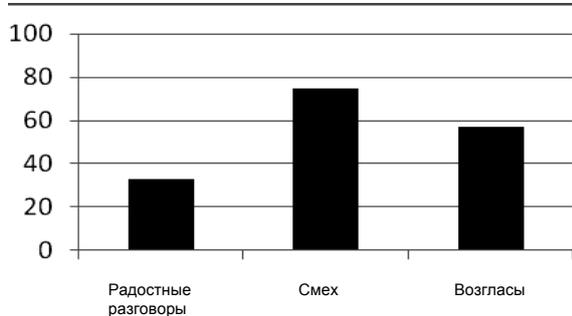


Рис. 4. Процент определения радостных звуков

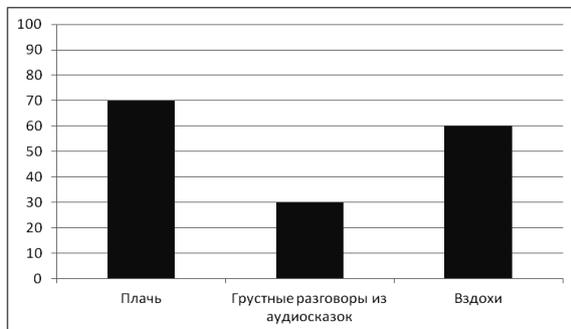


Рис. 5. Процент определения грустных звуков

По данным на диаграммах можно сделать вывод, что наиболее правильно определяется смех – 75 % и плач – 70 %, а разговоры из аудиосказок, как радостные, так и грустные, верно распознаются очень редко (всего 30 %), на наш взгляд, это связано с наигранностью эмоций рассказчика.

Заключение

В заключение необходимо отметить, что предложенный метод, основанный на оценке отношений частотных характеристик звуковой волны, можно с высокой степенью точности использовать для определения эмоционального состояния человека.

Предложенный метод прост для использования и восприятия. Наибольшей точно-

стью метод обладает при определении смеха для радостной эмоции и плача для грустной.

Дальнейшая работа по определению эмоций на основе частотных характеристик звуковой волны имеет большие перспективы, так как есть предположение, что данный метод можно применять для распознавания других эмоций, а также для большего разнообразия звуков, в том числе для звуков животных и даже звуков искусственного происхождения.

Список литературы

1. Романенко В.О. Эмоциональные характеристики вокальной речи и их связь с акустическими параметрами // Terra Humana. 2011. № 124.
2. Прокис Дж. Цифровая связь / пер. Кловский Д.Д. М.: 2000. URL: http://sernam.ru/book_p_net.php?id=38 (дата обращения: 03.10.2017).
3. Савельев И.В. Курс общей физики: в 3 т. М.: Наука, 1970. 517 с. Т. 3.
4. Близнюк О.А., Шкаранута А.П. Определение эмоционального состояния с помощью оценки соотношений частот звуковой волны // Математика и междисциплинарные исследования – 2017: сб. статей. Пермь, 2017. С. 45–49.
5. Лукин А. Спектроанализатор – что мы на нем видим? URL: <http://prosound.ixbt.com/education/spektr-analys.shtml> (дата обращения: 05.10.2017).
6. Бурцев Н.Н. О частотах нот полного звукоряда. Якутск, 2011. URL: <http://vargan.spb.ru/forum/topic/1710-бурцев-нн-о-частотах-нот-полного-звукоряда/> (дата обращения: 27.09.2017).
7. Изард К.Э. Психология эмоций / пер. с англ. СПб: Изд-во "Питер", 1999. 464 с.
8. Музыкальный лад. URL: <http://music-education.ru/muzykalnyj-lad/> (дата обращения: 20.09.2017).
9. Кураев Г.А., Пожарская Е.Н.. Психология человека. Ростов-на-Дону: УНИИ РГУ, 2002. 232 с.
10. Козлов Н.И. Базовые эмоции. URL: http://www.psychologos.ru/articles/view/bazovy_emoicii (дата обращения: 19.09.2017).

Study of application of the method for the basic emotions determination based on the ratio of the sound wave frequencies

O. A. Bliznyuk, A. P. Shkaraputa

Perm State University; 15, Bukireva st., Perm, 614990, Russia
olesiablnuk77@gmail.com, shkaraputa@psu.ru

The paper investigates the possibility of using the method of determining a person's emotional state based on comparison of the sound wave frequencies. The considered method of determining the basic emotions (joy, sadness) is based on the search for the frequencies of the spectrum corresponding to the major and minor mode. The search for compliance is carried out at regular intervals, and the number of matches is summated. The conclusion about the emotional state of the speaking person is made on the basis of quantitative characteristics. Fifty records have been analyzed, including 25 those with joyful emotions and 25 ones with sad emotions. We studied such sounds as laughter, crying, exclamations, sad sighs, as well as recordings from audios for children. When determining emotions, all records were classified according to the principle: "defined", "not defined" and "misdefined". "Defined" meant that in the record with sad emotions the minor mode was superior, while in the joyful record the major mode was superior. "Not defined" meant that no musical harmony was found, or the number of major and minor modes found was the same. "Misdefined" meant that in sad records the major mode was superior, and joyful ones had a minor mode. According to the research results, a sad emotion is best determined for crying and joyous – for laughter. The emotional state was determined in 70 percent of these cases.

Keywords: *classification of sounds; determination of emotions; frequencies; sound waves; minor; major.*