

УДК 004.93.12

## НЕЙРОСЕТЕВОЕ РАСПОЗНАВАНИЕ МУЗЫКАЛЬНЫХ ИНСТРУМЕНТОВ С ИСПОЛЬЗОВАНИЕМ МЕЛ-ЧАСТОТНЫХ КЕПСТРАЛЬНЫХ КОЭФФИЦИЕНТОВ

Станкевич Ф.В., Спицын В.Г.

ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»,  
Томск, e-mail: stankevichfv@tpu.ru

В данной статье рассматривается распознавание образов с применением нейронных сетей на примере распознавания музыкальных инструментов. Проблема распознавания звука музыкальных инструментов рассматривается во многих работах, но большинство из них используют большое количество признаков. Целью нашей работы является минимизация количества признаков при условии получения высокой точности распознавания. В качестве признаков используются мел-частотные кепстральные коэффициенты с применением метода главных компонент. Для классификации признаков используется нейронная сеть прямого распространения. Для обучения сети были рассмотрены два подхода: метод обратного распространения ошибки и нейроэволюционный алгоритм Enforced Subpopulations (ESP). Метод обратного распространения ошибки показал лучшие результаты. Данная работа посвящена распознаванию отдельно звучащих нот музыкальных инструментов, однако предложенный алгоритм показал хорошие результаты и при анализе сложных сигналов.

**Ключевые слова:** распознавание образов, музыкальные инструменты, мел-частотные кепстральные коэффициенты, метод главных компонент, нейроэволюция

## NEURAL NETWORK RECOGNITION OF MUSICAL INSTRUMENTS USING MEL-FREQUENCY CEPSTRAL COEFFICIENTS

Stankevich F.V., Spitsyn V.G.

National Research Tomsk Polytechnic University, Tomsk, e-mail: stankevichfv@tpu.ru

In this paper the task of automatic musical instrument recognition is considered. A lot of papers have been published on this topic, but most of them use a great number of features for recognition of musical instrument sound. This work is devoted to minimization the number of features together with getting of high recognition accuracy. The Mel-Frequency Cepstral Coefficients are considered as the main set of features. Additionally, the number of MFCC has been optimized using Principle Component Analysis. A feedforward neural network was used as a classifier. Two approaches were considered for classifier training: backpropagation method and Enforced Subpopulations neuroevolution approach. The backpropagation method was determined as the best one. Basically, this work is devoted to recognition of isolated notes, but also the developed algorithm was applied to a complex signal and as a result solo parties of musical instruments were successfully identified.

**Keywords:** pattern recognition, musical instruments, mel-frequency cepstral coefficients, principle component analysis, neuroevolution

Человеческий мозг способен распознавать различные визуальные и звуковые образы, в том числе и отличать звучание одного музыкального инструмента от другого, даже на фоне звучания других инструментов. В данной работе рассматривается проблема распознавания звука отдельного музыкального инструмента на основе его спектральных особенностей с применением нейронных сетей.

Задача распознавания музыкальных инструментов востребована при обработке сложных музыкальных сигналов, а именно аудиозаписей музыкальных произведений различных жанров, таких как классическая музыка, эстрадная музыка и другие. Данная задача относится к классу задач Music Information Retrieval (MIR) и может быть использована для аннотирования медиаконтента, сегментации музыкальных сигналов и идентификации музыкальных объектов (нот).

Этой теме посвящено множество исследований, например [2, 3 и 7], но боль-

шинство из них используют большое количество признаков для распознавания музыкальных инструментов. Целью данной работы ставится уменьшение количества признаков с получением высокой точности при распознавании отдельно звучащих нот музыкальных инструментов.

### Извлечение признаков

Музыкальный звук имеет характерную структуру в спектральном диапазоне (рис. 1.) В его составе есть основной тон, как правило, с наибольшей амплитудой, и сопутствующие гармоники – обертоны, наиболее значимыми являются первые несколько гармоник. Именно эти обертоны определяют тембр звука и соответственно музыкальный инструмент.

Однако спектр представляет собой большой набор данных, которые нецелесообразно использовать в исходном виде для решения задачи распознавания. В связи с этим необходимо определить значимый

набор признаков. В качестве такого набора было решено использовать мел-частотные кепстральные коэффициенты (MFCC). Данные коэффициенты были определены как лучшие признаки для распознавания музыкальных инструментов в работе [5].

Мел-частотные кепстральные коэффициенты представляют собой нелинейный спектр спектра, хорошо аппроксимируют слуховую систему человека, а также успешно используются для решения задач распознавания речи.

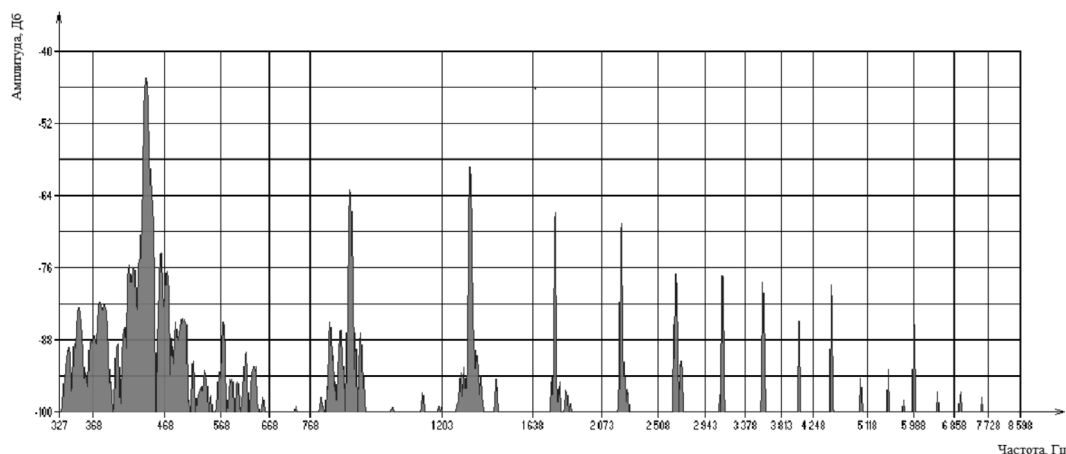


Рис. 1. Структура звука музыкального инструмента в спектральном диапазоне [1]

Алгоритм вычисления MFCC можно описать следующим образом [4]:

- вычисление оконного преобразования Фурье;
- нелинейное разбиение спектра на  $n$  частей с применением мел-шкалы;
- вычисление энергии сигнала для каждого интервала с применением треугольных фильтров (с перекрытием);
- вычисление логарифма энергии сигнала для каждого интервала;
- выполнение дискретного косинусного преобразования.

Для снижения сложности полученного пространства признаков мы использовали метод главных компонент (РСА). Это позволило нам уменьшить корреляцию признаков и удалить наименее значимые из них. Алгоритм вычисления главных компонент может быть описан следующим образом:

- определение матрицы корреляции;
- нахождение собственных значений и соответствующих собственных векторов;
- упорядочивание собственных векторов по соответствующим им собственным значениям (по убыванию);
- нахождение проекций входных данных на собственные векторы;
- отбрасывание последних  $m$  проекций.

Первые проекции представляют наиболее значимые компоненты в исходном векторе данных, и, соответственно, последние проекции представляют наименее значи-

мые. Более подробное описание алгоритма может быть найдено в [8].

### Классификация

Для обучения классификатора мы использовали набор маркированных примеров изолированных нот музыкальных инструментов Университета Айовы [9]. Для каждого инструмента были отобраны ноты в их эффективном рабочем диапазоне. Длительность звучания каждой ноты составляла 1–2 секунды. Мы использовали ноты, сыгранные в обычном стиле (деташе) на форте (громко), для обучения классификатора, а для тестирования – ноты, сыгранные на меццо-форте (довольно громко). Данный подход позволил приблизить процесс распознавания музыкальных инструментов к реальному сценарию.

В качестве классификатора мы использовали искусственную нейронную сеть (ИНС) прямого распространения и метод обратного распространения ошибки для обучения сети. Нейронная сеть содержала один скрытый слой. Параметры сети представлены ниже:

- число нейронов в скрытом слое: 24;
- допустимая ошибка классификации: 1%;
- скорость обучения: 0,05;
- максимальное количество эпох обучения: 250.

Для оценки процесса обучения мы использовали кросс-валидацию. Валидационная часть составляла 30% от обучающей выборки.

В качестве альтернативы методу обратного распространения ошибки, мы применили нейроэволюционный подход обучения сети, а именно метод Enforced Subpopulations (ESP), предложенный Фаустино Гомесом. Метод адаптирует значения весов ИНС посредством их генетической эволюции. Данный метод использует прямое кодирование и нейронную сеть прямого распространения с одним скрытым слоем. Одна из особенностей данного метода – это использование механизма взрывной мутации на основе распределения Коши для вывода процесса эволюции из локального экстремума. Кратко данный алгоритм может быть описан следующими шагами [6]:

- инициализация – создание  $h$  подпопуляций с  $n$  нейронами, где  $h$  – число нейронов в скрытом слое,  $n$  – размер популяции;
- оценка – выбирается случайная комбинация нейронов (по одному из каждой подпопуляции) и формируется нейронная сеть, затем оценивается ее приспособленность. Данная приспособленность добавляется кумулятивно к каждому нейрону этой сети. Оценка продолжается до тех пор, пока

каждый нейрон не примет участие как минимум в десяти оценках;

- проверка вырождения – если приспособленность лучшей сети не улучшается в течение  $b$  поколений, то выполняется взрывная мутация. Если после двух мутаций не происходит улучшение приспособленности, то выполняется адаптация размера сети;

- рекомбинация – вычисляется средняя приспособленность каждого нейрона, затем они сортируются (в пределах подпопуляции). 25 % лучших нейронов скрещиваются с использованием одноточечного кроссинговера. Для нейронов с низкой приспособленностью выполняется мутация с распределением Коши. В конце происходит выбор лучших  $n$  нейронов.

- этапы оценки-рекомбинации повторяются до тех пор пока не найдена сеть с требуемой приспособленностью (качеством).

### Алгоритм распознавания

Алгоритм распознавания музыкальных инструментов состоит из шести блоков (рис. 2).

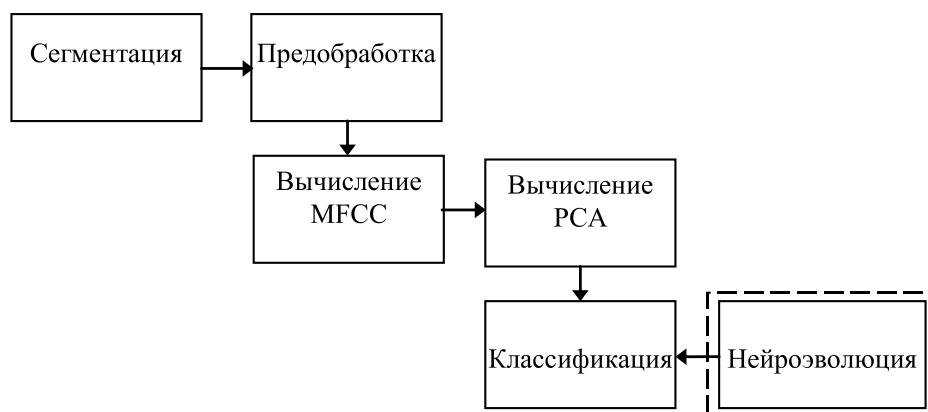


Рис. 2. Схема алгоритма распознавания

Сегментация разбивает сигнал на интервалы по 100 мс, предполагая, что спектр сигнала на этих участках постоянен в статистическом смысле. Предобработка включает увеличение амплитуды частот в верхней части спектра, а также применение оконной функции Хэмминга для устранения высоких частот полученных при обрезании сигнала на границах интервала. Предобработка происходит во временной области сигнала. Следующими этапами являются вычисление мел-частотных кепстральных коэффициентов и применение метода главных компонент. Нейроэволюционный метод обучения ИНС является опциональным этапом алгоритма.

### Полученные результаты

В результате тестирования алгоритма были найдены его оптимальные параметры. Это количество MFCC – 14 и количество главных компонент – 7. Эти параметры обеспечивают высокую точность (более 90%) при минимизации числа используемых признаков.

Метод главных компонент позволил нам сделать входные данные более компактными, уменьшить пространство признаков (с 14 до 7) и соответственно найти наиболее оптимальную поверхность решения. На рис. 3 показаны пространства признаков MFCC (а) и PCA (б) (три первых составляющих, 3D-вид). Можно заметить, что данные располагаются более компактно и через

них можно построить более простую разделяющую поверхность.

Использование нейроразволюционного алгоритма Enforced Subpopulation не смог-

ло дать более высокую точность в сравнении с методом обратного распространения ошибки. Результаты сравнения отражены в табл. 1.

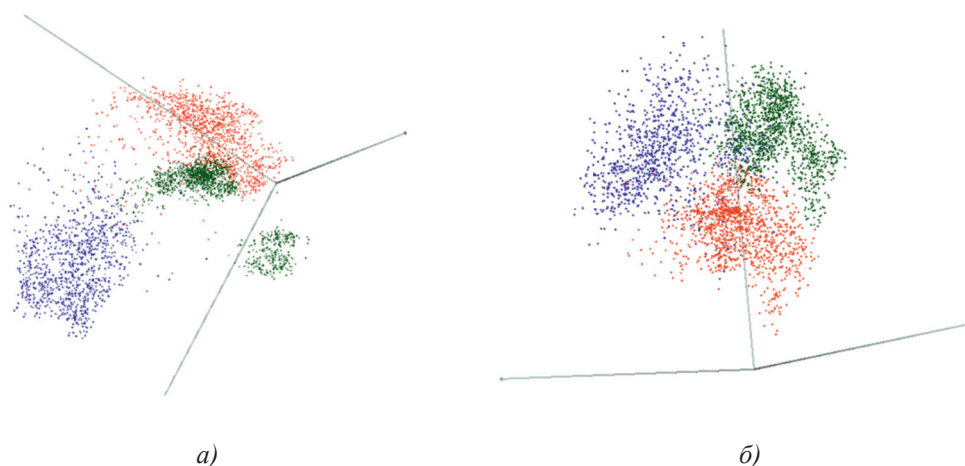


Рис. 3. Проекция данных: а – MFCC, б – PCA

Таблица 1

Сравнение ESP и метода обратного распространения ошибки

Метод	Кол-во инструментов	Время обучения, мин	Точность, %
ESP	2	10	99,10
Метод обратного распр. ошибки		0,01	100,00
ESP	3	21	97,12
Метод обратного распр. ошибки		0,01	97,46
ESP	4	30	95,10
Метод обратного распр. ошибки		0,05	96,15
ESP	5	45	91,76
Метод обратного распр. ошибки		0,4	92,84
ESP	6	54	92,07
Метод обратного распр. ошибки		0,9	92,89

Точность распознавания по инструментам, полученная с помощью метода обратного распространения ошибки показана в табл. 2.

Таблица 2

Точность распознавания (по инструментам)

Инструмент	Точность, %
Труба	99,47
Скрипка	92,95
Фортепиано	94,35
Флейта	95,15
Кларнет	74,31
Маримба	96,50
Средняя	92,89

Средняя точность распознавания для шести музыкальных инструментов составила 92,89%. Ошибки классификации по каждому инструменту показаны в табл. 3.

Зависимость точности распознавания от числа используемых в алгоритме музыкальных инструментов представлена на рис. 4. Инструменты добавлялись в следующем порядке: труба, фортепиано, скрипка, флейта, кларнет, маримба, тромбон, гобой, гитара, контрабас, саксофон и виолончель.

Описанный выше алгоритм был проанализирован относительно применимости к сложному сигналу. Для анализа использовалась Прелюдия Шостаковича № 5, в которой одновременно звучат два инструмента: скрипка и фортепиано. Сольные партии данных инструментов были успешно идентифицированы с помощью разработанного алгоритма. Результат анализа сложного сигнала представлен на рис. 5. Внизу вручную обозначены фактические сольные партии, выше цветом показан результат работы программы.

Таблица 3

Ошибки классификации

	Труба	Скрип.	Фо-но	Флейта	Клар.	Маримба
Труба	113	-	1	-	-	-
Скрипка	-	41	1	-	-	2
Фо-но	-	-	21	2	-	-
Флейта	-	-	-	68	-	-
Кларнет	-	-	-	12	52	1
Маримба	-	-	1	-	-	70

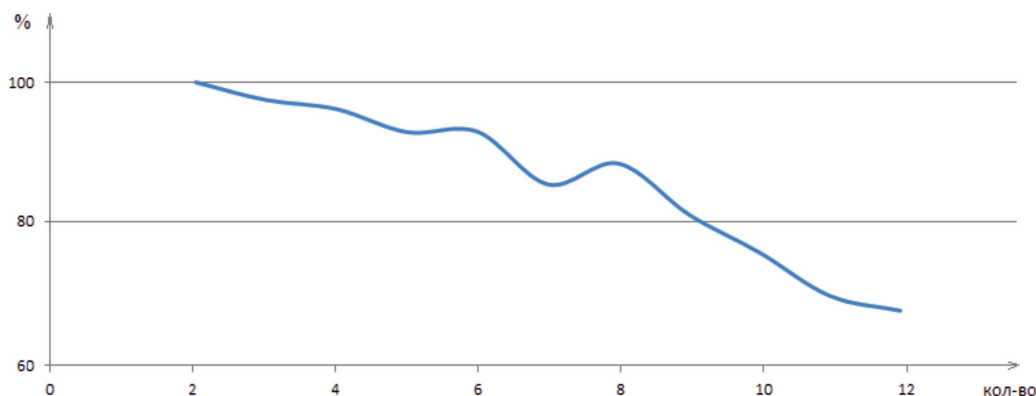


Рис. 4. Зависимость точности распознавания от числа музыкальных инструментов

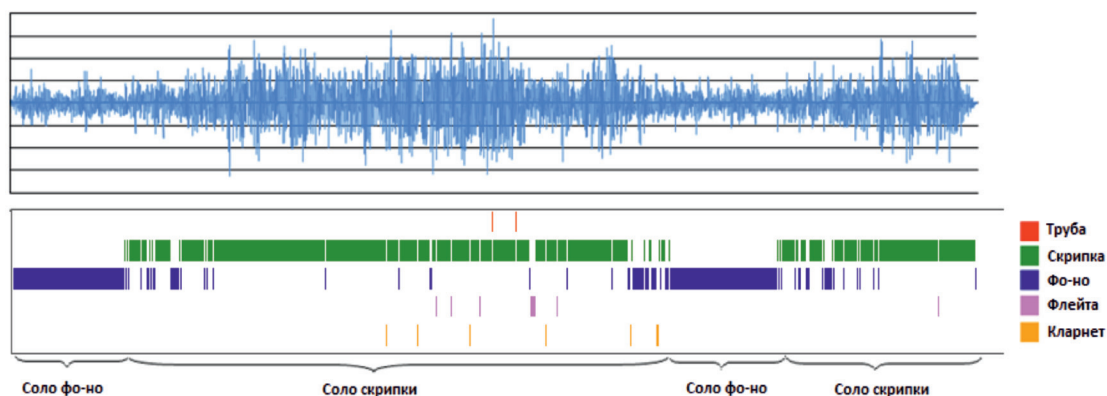


Рис. 5. Результат анализа сложного сигнала

### Заключение

В данной работе мы проанализировали использование нейросетевого подхода для распознавания музыкальных инструментов на основе мел-частотных кепстральных коэффициентов. Для снижения числа признаков был использован метод главных компонент. В результате была получена высокая точность распознавания музыкальных инструментов при использовании семи признаков, подаваемых на вход классификатора на базе искусственной нейронной сети.

*Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научного проекта № 12-08-000296.*

### Список литературы

1. Фадеев А.С. Идентификация музыкальных объектов на основе непрерывного вейвлет-преобразования. Диссертация, Томский политехнический университет, 2008 – 192 с.
2. Brown J.C. Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. J Acoust. Soc Am, 1999. – 30 p.
3. Chandwadkar D.M., Sutaone M.S. Role of Features and Classifiers on Accuracy of Identification of Musical Instruments.

2nd National Conference on Computational Intelligence and Signal Processing, 2012. – P. 66–70.

4. Davis S., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, – 1980. – Vol. 28. – № 4. – P. 357–366.

5. Eronen A. Comparison of features for musical instrument recognition. *Workshop on Signal Processing for Audio and Acoustics*, 2001. – P. 19–22.

6. Gomez F. Robust non-linear control through neuroevolution. PhD thesis, The University of Texas at Austin, 2003. – p. 137.

7. Nielsen A. B., Sigurdsson S., Hansen L. and Arenas-Garcia J. On the relevance of spectral features for instrument classification. *Proc. Acoustics, Speech and Signal Processing*, 2007. – Vol. 2. – P. 485–488.

8. Simon O. Haykin Neural networks: a comprehensive foundation, 2d ed, Person Education, 1999. – 842 p.

9. University of Iowa Electronic Music Studios. Music instruments samples. Available HTTP: <http://theremin.music.uiowa.edu/MIS.html>, access date: 6 Jul 2014.

### References

1. Fadeev A.S. Identifikatsiya muzykal'nykh ob'ektov na osnove nepreryvnogo veyvlet-preobrazovaniya. Dissertatsiya, Tomskiy politekhnicheskii universitet, 2008, 192 p.

2. Brown J.C. Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. *J Acoust. Soc Am*, 1999, 30 p.

3. Chandwadkar D.M., Sutaone M.S. Role of Features and Classifiers on Accuracy of Identification of Musical Instruments.

2nd National Conference on Computational Intelligence and Signal Processing, 2012, pp. 66–70.

4. Davis S., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1980, Vol. 28 No.4, pp. 357–366.

5. Eronen A. Comparison of features for musical instrument recognition. *Workshop on Signal Processing for Audio and Acoustics*, 2001, pp. 19–22.

6. Gomez F. Robust non-linear control through neuroevolution. PhD thesis, The University of Texas at Austin, 2003, p. 137.

7. Nielsen A.B., Sigurdsson S., Hansen L. and Arenas-Garcia J. On the relevance of spectral features for instrument classification. *Proc. Acoustics, Speech and Signal Processing*, Vol. 2, 2007, pp. 485–488.

8. Simon O. Haykin Neural networks: a comprehensive foundation, 2d ed, Person Education, 1999, 842 p.

9. University of Iowa Electronic Music Studios. Music instruments samples. Available HTTP: <http://theremin.music.uiowa.edu/MIS.html>, access date: 6 Jul 2014.

### Рецензенты:

Ким В.Л., д.т.н., профессор, профессор ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», г. Томск;

Авдеева Д.К., д.т.н., профессор, директор ООО «Медприбор», г. Томск.

Работа поступила в редакцию 05.12.2014.