

УДК 621.3

О. В. Рыбальский¹, В. И. Соловьев²

¹Национальная академия внутренних дел
Соломенская площадь, 1, 03034 Киев, Украина

²Восточноукраинский национальный университет
Молодежный квартал, 20-А, 91034 Луганск, Украина

Разработка и исследования пригодности экспертного инструментария «фрактал-м» для идентификации диктора по параметрам голосовых сигналов

Рассмотрен процесс исследований, разработки и проверки пригодности программы и методики для идентификации диктора по параметрам сигналов речи. В разработке программы использован фрактальный метод представления идентификационных признаков.

Ключевые слова: диктор, идентификация, частота основного тона, фонограмма, фрактал, экспертиза.

Введение

Используемые в настоящее время во многих странах аппаратно-программные средства и методики идентификационных исследований сигналов речи вызывают справедливые нарекания многих работников судов, следствия и адвокатуры, высказывающих обоснованные сомнения в объективности результатов экспертиз, проводимых с их применением.

Мы полагаем, что в основе подобного отношения к такому инструментарию лежит, возможно подсознательное, ощущение несоответствия протекания процесса распознавания человеком звуков речевой информации и используемых методов и средств идентификации этому процессу.

Цель статьи и постановка заданий

Изучив состояние и принципы работы современного инструментария идентификации диктора по сигналам речи, мы пришли к выводу, что высказываемые сомнения вполне обоснованы [1–10]. Этот вывод основан главным образом на современных представлениях о механизме восприятия и преобразования звуковых колебаний слуховым аппаратом человека. Как известно, звуковые колебания воспринимаются большим количеством рецепторных волосковых клеток человеческого уха. Каждая из таких клеток выделяет сигналы своей частоты (а точнее, узко-

© О. В. Рыбальский, В. И. Соловьев

го диапазона частот). Таким образом, волосковые клетки как бы образуют «гребенку» полосовых фильтров, воспринимающих сигналы «своих» частот, а также их уровни и изменения этих уровней [11–15]. Современными нейрофизиологическими исследованиями последних лет установлен и механизм работы нервных волокон слухового нерва [11–15]. Этими исследованиями доказано, что основные процессы передачи информации, содержащейся в звуковых сигналах, в мозг носят импульсный характер, а длительности этих импульсов лежат в пределах от единиц до сотни миллисекунд [15].

Исходя из этих данных, мы пришли к выводу, что способность анализатора к анализу сигналов на столь малых временных интервалах является определяющим фактором при идентификации и анализе речи, а этим требованиям наиболее полно отвечают вейвлет-преобразования.

Следовательно, в современной фоноскопии существует проблема создания нового инструментария для идентификационных исследований речевых сигналов, адекватного современным представлениям о работе слухового аппарата человека и обеспечивающего высокую достоверность результатов идентификации диктора по сигналам речи.

Разрешение этой проблемы потребовало решения ряда научных и технических задач.

Цель статьи — показать процесс постановки и нахождения путей решения этих задач, позволивший в результате, создать необходимый экспертный инструментарий, обеспечивающий высокую достоверность идентификации диктора.

Основной задачей было исследование факторов, влияющих на достоверность результатов идентификации. А решение этой задачи требовало изучения тонкой структуры речевых сигналов. Следовательно, требовалось создать новые программные средства, необходимые для проведения этих исследований. Но в первую очередь было необходимо осмыслить результаты современных научных исследований в области нейрофизиологии слуховой системы и адаптировать эти знания к задаче криминалистической идентификации диктора по голосу с точки зрения возможностей технической реализации поставленных задач. Решение этой задачи обеспечивало выбор необходимого математического аппарата, позволяющего сначала создать, а затем и отработать необходимые программы и методики. Таким образом, требовался комплексный системный подход к разрешению возникшей проблемы, объединяющий в своих рамках нейрофизиологические, математические, технические (в том числе акустические) и правовые аспекты.

Основная часть

Как было показано выше, изучение и осмысление нейрофизиологического аспекта проблемы продемонстрировало необходимость исследования сигналов на интервалах длительностью до ста миллисекунд. При этом мы предположили, что высокая достоверность результатов идентификации может быть обеспечена только при исследовании отдельных фонем и их компонент во временной и частотной области представления сигналов. Изучение сигналов во временной области необходимо потому, что все фонемы имеют четко выраженный фрактальный характер, сохраняемый и индивидуальный для каждой из фонем, т.е. форма сигнала фонемы

во временной области одинакова во всех языках и примерно одинакова при ее произношении любым индивидуумом. При этом количество таких фрактальных образований весьма ограничено (см., например, рис. 1 и рис. 2). Более того, путем применения аффинных преобразований фрагмента, показанного на рис. 2, можно получить фонему, идентифицируемую слушателями как звук [a], что и показано на рис. 3.

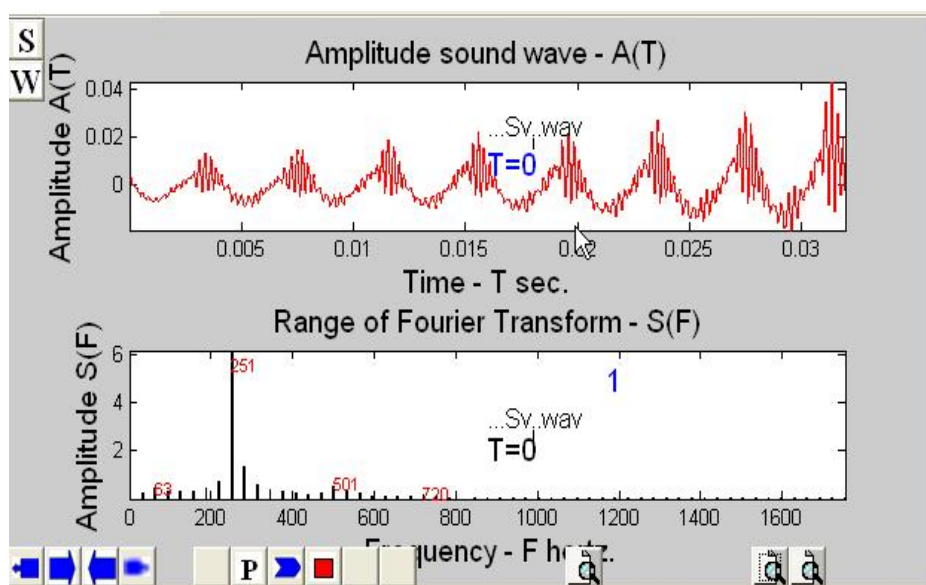


Рис. 1. Фрагмент акустической волны (фонема [и])

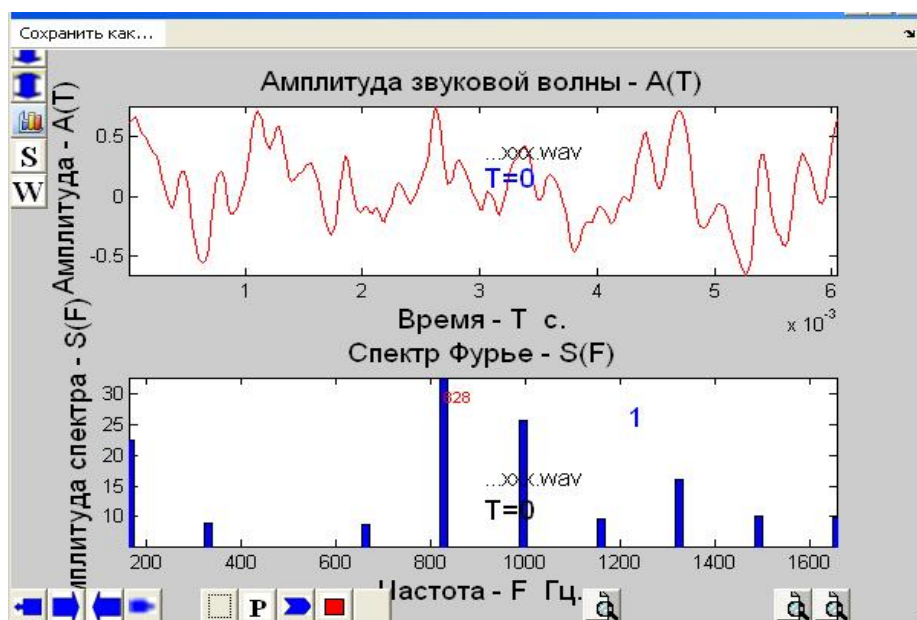


Рис. 2. «Атомарный» фрагмент фонемы [a]

Мы полагаем, что именно это мультифрактальное подобие позволяет нам распознавать речь любого человека. Основное различие, определяющее индивидуальность диктора, состоит в индивидуальности частотного состава сигналов фрагментов фонем, образующих этот звук при его произношении конкретной личностью. Эта индивидуальность, по нашему мнению, определяется частотой основного тона (ЧОТ) и модулирующими параметрами этой частоты. И ЧОТ, и эти параметры определяются индивидуальностью составляющих голосового тракта любого человека [16].

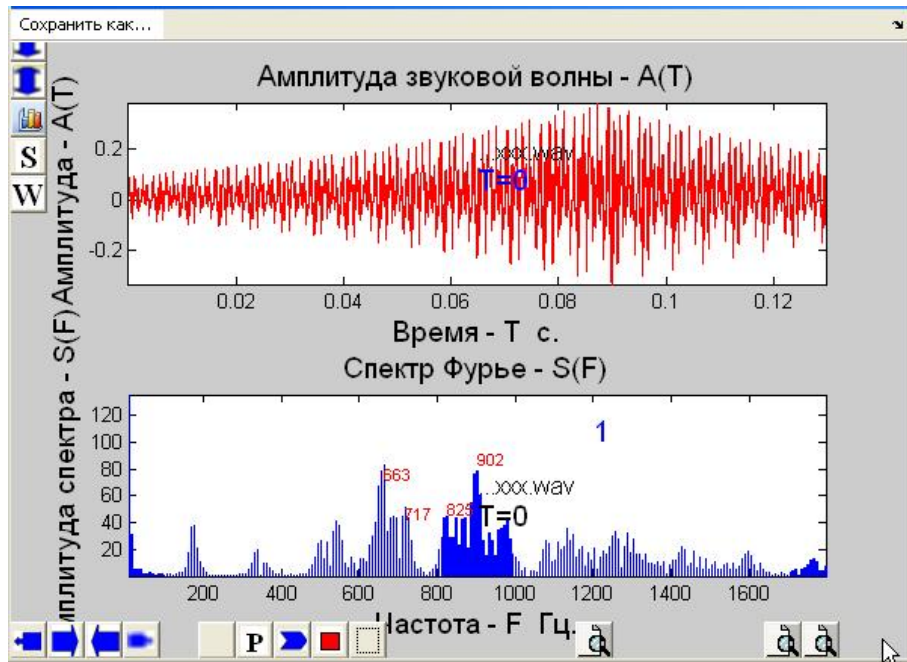


Рис. 3. Фрагмент модифицированной звуковой волны (фонема [а])

Исходя из этих особенностей построения и слухового восприятия фонем, мы предположили, что для создания инструментария для идентификационных исследований речи фрактальный подход является оптимальным, что привело к применению математического аппарата вейвлет-преобразования. Как было показано в наших работах, для реализации этого подхода оптимальным является вейвлет Морле [17].

Реализация программного обеспечения потребовала решения целого ряда разноплановых научно-технических задач теоретического и экспериментального характера.

Так, например, возникли две взаимосвязанные задачи — автоматической сегментации фонограммы на фонемы и выделения, подсчета и определения меры близости фрактальных образований, содержащихся в исследуемых сигналах спорной и образцовой фонограмм. Обе эти задачи были нами решены как в теоретическом, так и практическом аспектах [17–19].

При разработке программы и проведении исследований пришлось решить еще ряд математических задач, связанных с оптимизацией и ускорением вычис-

лений. Решение этих задач связано с вычислительной сложностью, обусловленной использованием двухпараметрического вейвлета Морле, и необходимостью обеспечения приемлемого времени проведения экспертизы. Потребовалось также решить ряд задач исследовательского и технического характера, связанных с определением особенностей проявления и выделения идентификационных признаков речи диктора при использовании фрактального подхода и вейвлет-анализа. Особого внимания, с нашей точки зрения, заслуживает разработанный метод выделения ЧОТ, при котором благодаря применению фрактального подхода, эта частота выделяется на «атомарном» уровне, т.е. по отдельным компонентам фонем. Решения этих задач частично показаны в работах [18; 19], а некоторые еще не освещенные аспекты этих решений являются темами отдельных публикаций.

Далее начался процесс исследований речевых сигналов с одновременной отработкой программы и методики ее применения при экспертизе. В результате этих исследований получены новые интересные возможности для идентификации голоса диктора, в том числе, возможность идентификации голоса на реализациях продолжительностью нескольких секунд, идентификации поющего или говорящего шепотом и т.д.

Основной результат этих исследований свелся к тезису: *доминирующей характеристикой индивидуальности диктора является частота его основного тона, выделяемого из фрактальных фрагментов отдельных фонем*. Описание проведенных теоретических и экспериментальных исследований, позволивших получить этот вывод, является предметом отдельной публикации и здесь не рассматривается.

Для установления пригодности сравнительных исследований по характерным фрактальным фрагментам фонем для проведения идентификации личности по голосу нами был проведен ряд экспериментов. Для этого была разработана первая версия программы, на которой проводились исследования возможности идентификации диктора путем сравнения ЧОТ, выделенной из фрагментов фонем. Использовались фонограммы экспертиз предыдущих лет, в которых диктор был установлен, что подтверждено результатами судебных заседаний, где рассматривались экспертные заключения по этим фонограммам. Эти эксперименты проиллюстрированы на рис. 4 и рис. 5.

В процессе проведения экспериментов была использована следующая методика.

1. В компьютер вводятся две фонограммы — образцовая и спорная.

2. Образцовая фонограмма сегментируется на фрагменты и вычисляется распределение частот основного тона по всей длине фонограммы на основе распределений частот основного тона, полученных для сегментов. Поскольку известно, что в образцовой фонограмме голос принадлежит конкретному лицу, то, с учетом использованных принципов построения программы, данные операции не вызывают затруднений.

3. В спорной фонограмме производится сегментация на фрагменты и рассчитывается распределение частот основного тона для каждого из этих фрагментов. Проверяется принадлежность к одному распределению для частот основного тона каждого фрагмента, выделенного из спорной фонограммы, с распределением, по-

лученным для образцовой фонограммы. Определяется степень близости распределения частот основного тона каждого из фрагментов с распределением этих частот для всей образцовой фонограммы. На основании проведенного анализа определяются фрагменты, относящиеся к голосу идентифицируемой личности. Производится проверка принадлежности распределений частоты основного тона в отобранных фрагментах к одному распределению. Устанавливается степень близости между распределениями частоты основного тона для каждого из отобранных фрагментов, с распределением для частоты основного тона, полученным в образцовой фонограмме. Затем производится установление степени близости между распределением частоты основного тона в образцовой фонограмме, с общим распределением этой частоты для всех отобранных фрагментов спорной фонограммы.

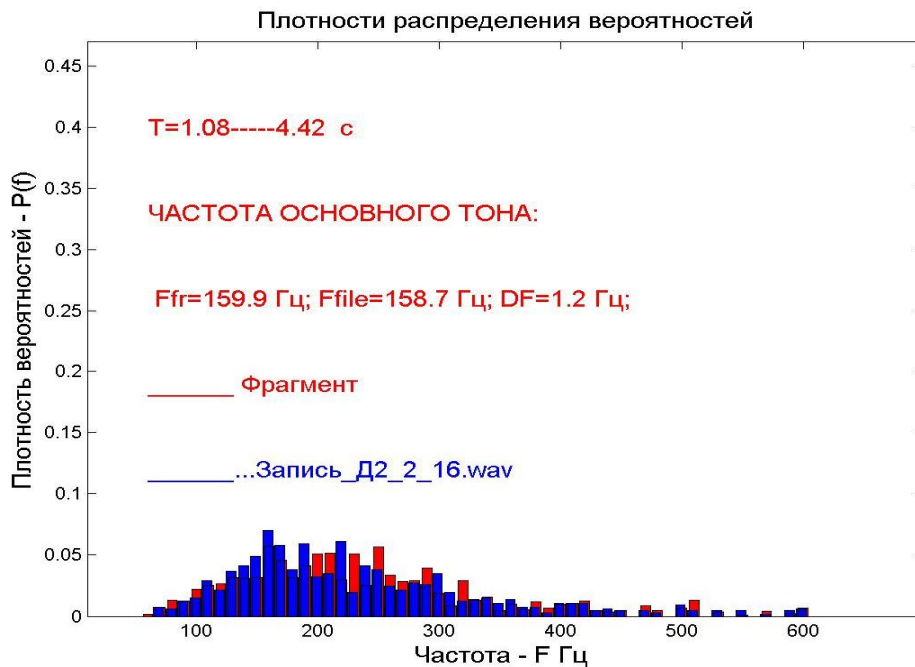


Рис. 4. Сравнение распределения плотности вероятностей частоты основного тона, выделенной из одного из фрагментов спорной и всей образцовой фонограмм. Речь на двух фонограммах принадлежит одному диктору

Выводы об идентичности и различии дикторов, полученные при проведении экспериментов (рис. 4 и рис. 5), обусловлены различием плотностей вероятности ЧОТ и математических ожиданий, что показано на гистограммах. При этом мера близости этих частот DF между ЧОТ, выделенной из фрагмента спорной (F_{fr}) и образцовой (F_{file}) фонограмм, определяет принятие решения экспертом.

Априорное знание принадлежности диктора при проведении эксперимента обеспечило точность принятия общего решения о принципиальной пригодности фрактального метода идентификации диктора. Но при идентификационной экспертизе никаких априорных знаний о принадлежности голоса диктора не может существовать по определению. Более того, в фонограммах, полученных, напри-

мер, оперативным путем, как правило, присутствуют голоса двух и более собеседников. Следовательно, получение подтверждения принципиальной пригодности метода привело к пониманию необходимости создания усовершенствованной программы, позволяющей автоматизировать процесс проведения экспертизы. Эти усовершенствования выразились в следующих требованиях.

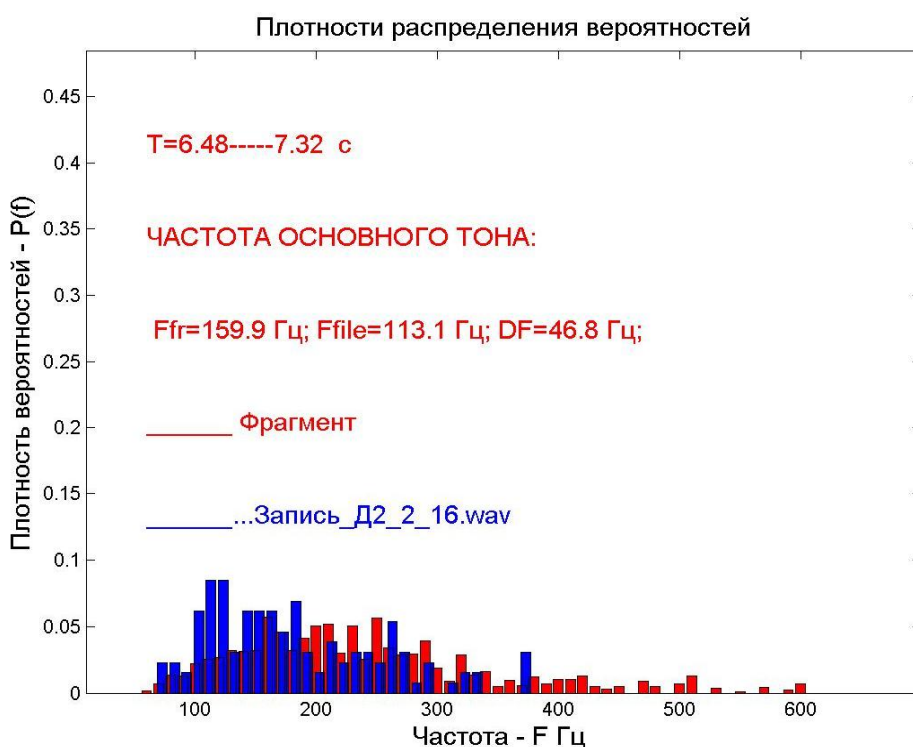


Рис. 5. Сравнение распределения плотности вероятностей частоты основного тона, выделенной из одного из фрагментов спорной и всей образцовой фонограмм. Речь на двух фонограммах принадлежит разным дикторам

1. Образцовая фонограмма должна содержать запись голоса только одного диктора — человека, чей голос предположительно записан на спорной фонограмме.
2. Речь на образцовой фонограмме должна сегментироваться на фонемы в автоматическом режиме.
3. Сегментированные фонемы должны автоматически сортироваться по произносимым звукам, поскольку наличие в них определенных типов фракталов, характерных для фонемы каждого звука, известны.
4. Для каждой из одинаковых фонем (или их характерных фрагментов) в автоматическом режиме выделяется ЧОТ и рассчитываются плотности вероятностей ЧОТ для каждой из отсортированных групп фонем.
5. В спорной фонограмме также производят автоматическую сегментацию по фонемам и сортируют фонемы по произносимым звукам.
6. В каждой из отсортированных фонем выделяют ЧОТ и сравнивают с ЧОТ для данной группы фонем, выделенной из образцовой фонограммы. Из отобран-

ных фонем автоматически отбирают близкие к образцовым по ЧОТ для данных фонем и формируют из них группы для последующего сравнения.

7. Производится сравнение ЧОТ по группам и определяется мера близости, а на основании результатов сравнения делается вывод относительно идентичности параметров голосовых сигналов идентифицируемой личности.

Проведенные исследования позволили разработать программу, отвечающую этим требованиям.

Конкретное описание принципов построения программы, ее последующей отработки и создания экспертной методики являются предметом последующих публикаций.

ВЫВОДЫ

1. В процессе исследований возможности создания методов и программ для достоверной идентификации диктора по параметрам сигналов речи было установлено, что речевые сигналы носят фрактальный характер.

2. Установлено, что для обеспечения высокой достоверности идентификации личности по параметрам речевых сигналов целесообразно выделять частоту основного тона на уровне отдельных частей фонем, содержащих характерные для них фрактальные образования.

3. Разработана программа, реализующая возможность идентификации личности по фрагментам фонем, выделяемым из сигналов речи, записанных на образцовой и спорной фонограммах.

1. *Forensic speech and audio analysis forensic linnguistics / A. Review: 2001 to 2004 / A.P.A. Broeders MA PhD Chief Scientist Netherlands Forensic Institute Ministry of Justice / P.O. Box 24044 2490 AA THE HAGUE. The Netherlands.*

2. *Судебные экспертизы и исследования. Видео-фоноскопические экспертизы и исследования [Электронный ресурс]. — Режим доступа: http://www.forenex.ru/forensic_audio_video/*

3. *On the Theory and Practice of Voice Identification / Bolt R.H. [et al.] — National Academy of Sciences: Washington DC, 1979.*

4. *Saks M.J. Merlin and Solomon: Lessons from the Laws Formative Encounters with Forensic Identification Science / M.J. Saks // Hastings Law Journal. — 1998. — 49(4). — P. 1069–1141.*

5. *Ormerod D. Sounding Out Expert Voice Identification / D. Ormerod // Criminal Law Review. — 2002. — P. 771–790.*

6. *Meuwly D. Reconnaissance de Locuteurs: L'Apport D'une Approche Automatique / D. Meuwly. — PhD Thesis — University of Lausanne, 2001.*

7. *Marescal F. (1999) The Forensic Speaker Recognition Method Used by the French Gendarmerie/ F. Marescal. — Internal Publication. — IRCGN: Paris.*

8. *Pfister B. Personenidentifikation anhand der Stimme / B. Pfister // Kriminalistik. — 2001. — 55(4). — P. 287–292.*

9. *Gonzalez-Rodriguez J. Identivox: A PC Windows Tool for Text-Independent Speaker Recognition in Forensic Environments / Gonzalez-Rodriguez J., Ortega-García J. & J. Lucena-Molina. — in J. WUjciakiewicz (ed.) Problems of Forensic Sciences: Proc. of EAFS Meeting. — Cracow, Poland, 2001. — Vol. XLVII. — P. 246–253.*

10. *Галяшина Е.И.* Проблемы развития судебной фоноскопической экспертизы / Галяшина Е.И. // Сучасні проблеми розвитку судової експертизи: зб. матер. засід. круглого столу, присвяч. 10-річчю створення Севастопольського від. Харк. наук.-дослід. ін-ту суд. експертиз ім. І.С. Бокаріуса (Севастополь, 10–11 черв. 2010 р.). — Х.: ХНДІСЕ, 2010. — С. 89–91.
11. *Бибиков Н.Г.* Нейрофизиологические механизмы слуховой адаптации. I. Адаптация в течение действия стимула / Н.Г. Бибиков // Успехи физиологических наук. — 2010. — Т. 41, № 3. — С. 72–91.
12. *Бибиков Н.Г.* Нейрофизиологические механизмы слуховой адаптации. II. Эффекты последовательного действия / Н.Г. Бибиков // Успехи физиологических наук.— 2010. — Т. 41, № 4. — С. 76–492
13. *Dynamic Range Adaptation to Sound Level Statistics in the Auditory Nerve* / Wen B, Wang G.I., Dean I., Delgutte B. // J. Neurosci. — 2009. — Vol. 29. — P. 13797–13808.
14. *Time Course of Dynamic Range Adaptation in the Auditory Nerve* / Wen B., Wang G.I., Dean I., Delgutte B // J. Neurophysiology. — 2012. — Vol. 107. — in press.
15. *Бибиков Н.Г.* Механизмы адаптации в нейронных сетях слухового анализатора / Н.Г. Бибиков; сб. тр. XXV сессии Российского акустического общества: сессия Научного совета по акустике РАН, медицинская и биологическая акустика. — 2013. — С. 51–56.
16. *Сапожков М.А.* Электроакустика: учеб. для вузов / М.А. Сапожков. — М.: Связь, 1978. — 272 с.
17. *Соловьев В.И.* Идентификация характеристик голоса на основе максимумов вейвлет-преобразования / В.И. Соловьев // Реєстрація, зберігання і оброб. даних. — 2013. — Т. 15, № 2. — С. 39–47.
18. *Рыбальский О.В.* Сегментация звукового сигнала в задачах выявления монтажа в аудио-файлах / О.В. Рыбальский, В.И. Соловьев // Вісник Київського національного ун-ту ім. Т. Шевченка. Військово-спеціальні науки. — 2011. — № 33. — С. 213–220.
19. *Соловьев В.И.* Метод анализа главных компонент в задачах обработки аудиоинформации / В.И. Соловьев // Сучасна спеціальна техніка. — № 1 (32), 2013. — С. 64–69.

Поступила в редакцию 27.01.2014