

БИОИНСПИРИРОВАННЫЙ И ПСИХОАКУСТИЧЕСКИ МОТИВИРОВАННЫЙ АНАЛИЗ ЗВУКОВЫХ СИГНАЛОВ

Вашкевич М. И., Азаров И. С.
Кафедра электронных вычислительных средств,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: vashkevich@bsuir.by, azarov@bsuir.by

Биоинспирированный и психоакустически мотивированные методы анализа звуковых сигналов являются основой многих систем, выполняющих обработку речевой и аудио информации. В основе этих методов лежит знание особенностей работы слуховой системы человека и особенностей восприятия им звуков. В данной работе производится обзор основных методов биоинспирированного и психоакустически мотивированного анализа звуковых сигналов. Дается краткое пояснения какие именно свойства и особенности слуховой системы описывает каждый метод. Среди биоинспирированных методов анализа описываются: гамматон и гаммачирп банки фильтров, roex-фильтры, а также модель слуховых фильтров Кейтса. Среди психоакустически мотивированных способов анализа приводятся метод ДПФ с суммированием энергий в субполосах, банки фильтров на основе вейвлет преобразования, а также банки фильтров на основе фазового преобразования.

ВВЕДЕНИЕ

Частотно-временные методы анализа звуковых сигналов, основанные на принципах психоакустики, являются базой для построения моделей слухового восприятия. Задача моделирования слухового восприятия возникает во многих приложениях цифровой обработки сигналов. В частности, она возникла в связи с повсеместным распространением цифровых мультимедийных систем. Так, например, перцентуальное (т.е. основанное на модели слухового восприятия) аудиокодирование используется для уменьшения скорости передачи данных в таких алгоритмах, как МРЗ и ААС [1]. Перцентуальное аудиокодирование основано на использовании феномена маскирования, который можно рассчитать исходя из того, как сигнал-возбуждение представляется в улитке уха. Системы распознавания речи также используют модель слуха для начального анализа сигнала. Этим эмулируется предварительная обработка сигнала до того этапа, на котором мозг выполняет собственные процедуры распознавания речи. Ценность таких методов анализа звука для систем распознавания речи заключается в их устойчивости к шумам акустической среды [2].

Используемые в настоящее время модели слухового восприятия значительно отличаются по числу моделируемых ими феноменов, присутствующих реальной слуховой системе. Это объясняется тем, что в каждой области применения к модели слуха предъявляются собственные требования. В данной работе дается краткий обзор существующих методов анализа звуковых сигналов, используемых в моделях слухового восприятия.

Методы анализа сигналов, используемые на первом этапе моделирования слухового восприятия, можно условно разделить на *биоинспирированные* и *психоакустически мотивированные*. В биоинспирированных методах ставится задача моделирования работы органов слуховой системы. В частности, при помощи банка цифровых фильтров моделируется механическое движение базилярной мембраны улитки уха. Такие фильтры называют кохлеарными (англ. *cochlear* – улитковый) или слуховыми. Психоакустический мотивированный анализ основывается на закономерностях, выявленных в процессе изучения восприятия слушателем различных акустических стимулов [3].

Существуют несколько биоинспирированных моделей слуховых фильтров:

- roex-фильтры [4];
- гамматон банк фильтров [5];
- модель Кейтса [6];
- гаммачирп банк фильтров [7];

Модели слуховых фильтров получают путем подгонки параметрических выражений под имеющиеся экспериментальные данные. Простое описание слухового фильтра в частотной области можно получить с использованием скругленной экспоненциальной функции (roex-функции):

$$|H(f)|^2 = (1 - r)(1 - pg)e^{-pg} + r,$$

где g – нормализованное отклонение частоты f от центральной частоты фильтра f_c :

$$g = |f - f_c|/f_c.$$

Параметр p характеризует добротность фильтра и должен выбираться, как $p = 4f_c/\text{ERB}(f_c)$, где

$$\text{ERB}(f_c) = 0.1079f_c + 24.7.$$

Второй параметр r определяет пологость формы фильтра за пределом полосы пропускания. Недо-

статком данной модели является то, что слуховые фильтры определяются только в частотной области, а временные параметры фильтров в явном виде не определены.

Гамматон функции [5], используются для описания во временной области импульсных характеристик слуховых фильтров:

$$h(t) = t^{(l-1)} e^{-2\pi b \text{ERB}(f_c)t} \cos(2\pi f_c t), \quad t > 0, \quad (1)$$

где b – параметр, определяющий ширину полосы, а l – порядок. На практике используют значения $l = 4$, $b = 1,019$. Фильтры получили свое название от того, что их огибающая формируется как произведение степенной и экспоненциальной функций, которое является функцией гамма-распределения, а косинусная несущая является тоном, когда её частота находится в слышимом диапазоне.

На практике гамматон фильтры моделируются в виде цифровых КИХ-фильтров. При этом, однако, возникает затруднение от того, что получаемые импульсные характеристики фильтров имеют большую длительность. На рис. 1 приведена импульсная характеристика гамматон фильтра для центральной частоты 500 Гц, где также показана его огибающая. Для сравнения на рис. 1 также приведена огибающая гамматон фильтра с центральной частотой 80 Гц. Можно заметить, что для центральной частоты 500 Гц длина импульсной характеристики фильтра будет составлять порядка 240 отсчетов (при частоте дискретизации 8000 Гц). Для более низких центральных частот, длина импульсной характеристики возрастает ещё больше (например для $f_c = 80$ Гц, она будет содержать 600 отсчетов). Таким образом, получающиеся КИХ-фильтры в вычислительном плане являются довольно затратными.

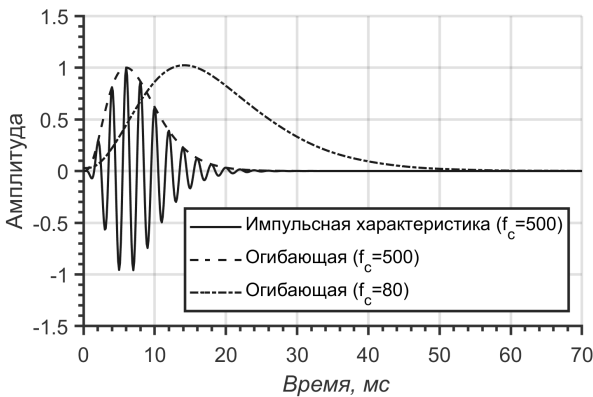


Рис. 1 – Импульсная характеристика и огибающие импульсных характеристик гамматон фильтров для различных центральных частот

В работе [7] была уточнена модель слухового гамматон фильтра, путем добавления

частотно-модулирующего члена:

$$g_c(t) = at^{(l-1)} e^{-2\pi b \text{ERB}(f_r)t} \cos(2\pi f_r t + c \ln t + \phi), \quad t > 0, \quad (2)$$

Единственная разница между выражениями (2) и (1) это наличие члена $c \ln t$, где c – дополнительный параметр. Фильтр имеет монотонно возрастающую частотно-модулированную несущую (чирп-сигнал) с огибающей, которая является функцией гамма-распределения, и отсюда и название «gammachirp». В (2) используется обозначение f_r вместо f_c для параметра частоты, потому что пиковая частота АЧХ более зависит от c и, в меньшей степени, от b и l . Эквивалентная прямоугольная полоса пропускания фильтра зависит от уровня стимула. Для удобства параметр b связывается с уровнем стимула, так что основная формула для ширины фильтра, такая же, как в выражении (1).

Несмотря на то, что гаммачирп фильтры лучше описывают имеющиеся физиологические данные в плане сложности реализации они имеют те же недостатки, что и гамматон фильтры. Более того, гаммачирп фильтры зависят от уровня подаваемого сигнала (в частности изменяется ширина полосы пропускания), что также является препятствием для получения их эффективной реализации.

Алтернативным подходом к построению слуховых фильтров является модель Кейтса, реализующая фильтры бегущей волны [6]. Блок-схема слуховых фильтров Кейтса приведена на рис. 2.

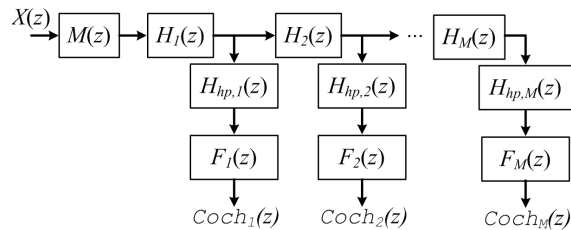


Рис. 2 – Модель слуховых фильтров Кейтса

Важной особенностью этой модели является то, что она представляет собой единую вычислительную структуру. Тем не менее, поскольку вся обработка выполняется во временной области, реализация данной модели также требует значительных вычислительных затрат.

II. ПСИХОАКУСТИЧЕСКИ МОТИВИРОВАННЫЙ АНАЛИЗ

Психоакустика – научное направление, задачей которого является изучение структуры и функции слухового пути с помощью анализа взаимосвязи акустических стимулов, возбуждающих слуховую систему, и психологических (моторных) реакций слушателя на эти системы [3]. В методах психоакустически мотивированного

анализа не ставится задача подогнать какую-либо параметрическую модель под имеющиеся физиологические данные. Целью такого анализа является смоделировать особенности частотной селективности уха. Известно, что слуховая система разбивает весь слышимый частотный диапазон на *критические полосы*. При восприятии звуков со сложным спектром, громкость частотных компонент, попадающие в одну критическую полосу, суммируется. Таким образом, компоненты с различной частотой, в том случае, когда они попадают в одну критическую полосу, воспринимаются, как единый сигнал.

Можно выделить следующие три группы психоакустически мотивированных методов:

- ДПФ с суммированием энергий в субполосах;
- банки фильтров на основе вейвлет преобразования [8];
- банки фильтров на основе фазового преобразования [9];

Согласно первому из перечисленных методов, неравномерное спектральное разрешение слуховой системы учитывается путем суммирования энергий соответствующего числа соседних частотных субполос ДПФ. Следовательно, фазовое соотношение между спектральными компонентами различных субполос в пределах полосы слухового фильтра не принимается во внимание. Поскольку фронты слухового фильтра менее крутые, чем крутизна субполос, они должны быть аппроксимированы путем распределения энергий субполос по нескольким критическим полосам. Данный способ отображения равных субполос в полосы слуховых фильтра дает огибающие выходного сигнала, которые отличаются от тех, которые измеряются на выходе слуховой улитки. Временное разрешение такого спектрального разложения определяется размером преобразования (т.е. длиной ДПФ), и, таким образом, является постоянным для всех центральных частот. Для высокочастотных полос это приводит к значительно более низкому временному разрешению по сравнению с соответствующими слуховыми фильтрами. Все описанные несоответствия способствуют неточному моделированию частотного анализа, производимого ухом.

Банки фильтров, получающиеся на основе вейвлет преобразования имеют большое значение для моделирования частотной селективности уха. В частности, при использовании вейвлет банков фильтров достигается разрешение с постоянной добротностью (англ. *constant-Q resolution*) [10], при котором временное разрешение возрастает, а частотное уменьшается по мере увеличения частоты. В отличие от предыдущего метода анализа, банки фильтров на основе вейвлетов имеют неравномерное временное разрешение, что отражает свойство реальной слуховой

системы. Тем не менее, сложность вейвлет банков фильтров заключается в том, что производимая ими частотно-временная декомпозиция сигнала напрямую зависит от древовидной структуры банка фильтров. Также ширина критических полос не достаточно точно моделируется при помощи таких банков фильтров [11-12].

Банки фильтров, основанные на применении фазового преобразования, являются альтернативным методом психоакустически мотивированного анализа по отношению к вейвлет банкам фильтров [13]. Особенностью данного вида банков фильтров является то, что они образуются из обычных равнополосных структур анализа путем замены элементов задержки на фазовые звенья:

$$z^{-1} \rightarrow A(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}, \quad A(e^{j\omega}) = e^{j\theta(\omega)}, \quad (3)$$

где α – параметр отвечающий за степень растяжения частотной оси, в результате которой банк фильтров получает свойство неравнополосности. Выражение (3) носит название *фазового преобразования* [14]. В результате применения (3) к линейной дискретной системе происходит трансформация частотной оси, описываемая выражением:

$$\theta^{-1} : \Omega \rightarrow \omega, \quad (4)$$

где Ω – деформированная частотная ось, ω – исходная частотная ось, а θ – функция, отображающая ω в Ω :

$$\theta(\omega) = \omega + \arctg \left(\frac{\alpha \sin \omega}{1 - \alpha \cos \omega} \right). \quad (5)$$

Таким образом, при использовании фазового преобразования исходный банк фильтров является равнополосным относительно деформированной частотной оси Ω . Однако, относительно частоты ω на которой производится обработка сигнала в силу (4) банк фильтров является неравнополосным. Описанный процесс иллюстрируется рис. 3, где показано получение 8-канального неравнополосного ДПФ-модулированного банка фильтров. Более подробно механизм фазового преобразования описан в [13].

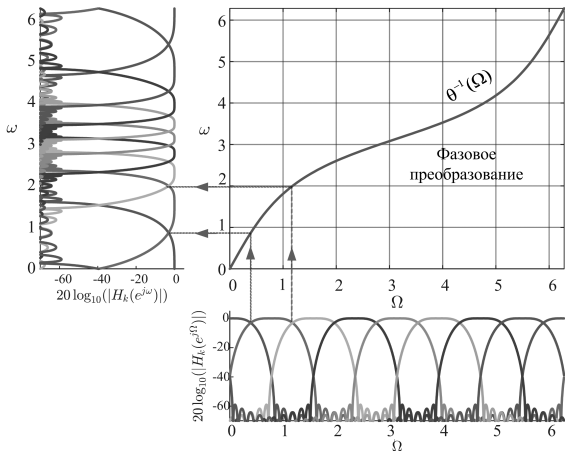


Рис. 3 – Процесс получения неравнополосного банка фильтров с использованием фазового преобразования, коэффициент определяющий степень деформации частотной оси $\alpha = 0,4$ [13]

Существенным отличием банков фильтров на основе фазового преобразования является то, что выполняемая ими частотно-временная декомпозиция сигнала не имеет жесткой привязки к структуре банка фильтров (как это имеет место для вейвлет банков фильтров). Также использование фазового преобразования позволяет более точно аппроксимировать психоакустическую шкалу барков [11, 15], чем банк фильтров на основе вейвлетов. К недостаткам банков фильтров на основе фазового преобразования можно отнести то, что их структура включает цепочку фазовых звеньев, которая может содержать более 100 элементов [9,11]. Вычисления производимые в цепочке фазовых звеньев трудно распараллелить, и как следствие при её программной реализации тратится значительное процессорное время. При аппаратной реализации цепочки фазовых звеньев такой проблемы не возникает [16].

III. Выводы

В докладе рассмотрены основные методы биоинспирированного и психоакустически мотивированного анализа звуковых сигналов, используемых в современных информационных системах и технологиях. Изначально применение описанных методов анализа было обусловлено бурным развитием систем кодирования речи и аудио в 90-х года прошлого столетия [1], далее их стали активно применять в системах распознавания речи [2]. На настоящем этапе биоинспирированные и психоакустически мотивированные методы анализа звуков применяются в различных информационных системах, осуществляющих шумоподавление [11-12], конверсию голосового сигнала, повышение разборчивости речи. Особо можно выделить медицинские системы, в которых также используются указанные методы, это слуховые аппараты [16], акустические анализаторы голоса, программные комплексы для выявления патологии по голосу [17].

СПИСОК ЛИТЕРАТУРЫ

1. Ковалгин, Ю. А. Аудиотехника / Ю. А. Ковалгин, Володин Э. И. – М.: Горячая линия–Телеком, 2013. – 742 с.
2. Kim, D. S. Auditory processing of speech signals for robust speech recognition in real-world noisy environments / D. S. Kim, S. Y. Lee, R. M. Kil // IEEE Transactions on speech and audio processing. – 1999. – vol. 7, №. 1. – P. 55–69.
3. Слуховая система / ред. Я. А. Альтман. – Л.: Наука, 1990. – 620 с.
4. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold / R. D. Patterson, *et al.* // The Journal of the Acoustical Society of America. – 1982. – vol. 72, №. 6. – P. 1788–1803.
5. An efficient auditory filterbank based on the gammatone function / R. D. Patterson, *et al.* // Proceeding of the IOC Speech Group on Auditory Modelling at RSRE. – 1987. – vol. 2, №. 7. – P. 1-34.
6. Kates, J. M. A time-domain digital cochlear model // IEEE Transactions on Signal Processing. – 1991. – vol. 39, №. 12. – P. 2573–2592.
7. Irino, T. A time-domain, level-dependent auditory filter: The gammachirp / T. Irino, R. D. Patterson // The Journal of the Acoustical Society of America. – 1997. – vol. 101, №. 1. – P. 412-419.
8. Петровский, Ал. А. Масштабируемые аудиоречевые кодеры на основе адаптивного частотно-временного анализа звуковых сигналов / Ал. А. Петровский, Петровский А. А. // Труды СПИИРАН. – 2016. – Т. 1, №. 50. – С. 55–92.
9. Вашкевич, М. И. Проектирование передискретизированного неравнополосного косинусно-модулированного банка фильтров / М. И. Вашкевич, А. А. Петровский // Информатика. – 2011. – №. 2. – С. 21–39.
10. Quatieri, T.F. Discrete-time speech signal processing: principles and practice / Prentice Hall Signal Processing Series. — Prentice Hall PTR, 2002. — P. 781.
11. Gulzow, T. Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement / T. Gulzow, A. Engelsberg, U. Heute // Signal processing. – 1998. – Vol. 64, no. 1. – pp. 5–19.
12. Carnero, B. Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms / B. Carnero, A. Drygajlo // IEEE Transactions on Signal Processing. – 1999. – vol. 47, №. 6. – P. 1622–1635.
13. Вашкевич, М.И. Сравнение частотно-временных преобразований: Фурье-анализ, вейвлеты и банки фильтров на основе фазового преобразования / М. И. Вашкевич, И. С. Азаров // Цифровая обработка сигналов – 2020. – №2. – С. 13–26.
14. Vary, P. Digital filter banks with unequal resolution // Short Communication Digest of European Signal Processing Conference (EUSIPCO), 1980. – pp. 41–42.
15. Smith, J. O. Bark and ERB bilinear transforms / J. O. Smith, J. S. Abel // IEEE Transactions on speech and Audio Processing. – 1999. – vol. 7, №. 6. – P. 697–708.
16. Вашкевич, М. И. Неравнополосные банки фильтров для слуховых аппаратов: анализ алгоритмов, автоматизация проектирования / М. И. Вашкевич, А. А. Петровский // Автоматизация проектирования дискретных систем: материалы 7-й междунар. конф. Беларусь, Минск, 16–17 ноября. — 2010. – С. 53-60.
17. Вашкевич М. И., Азаров И. С. Определение патологии голосового аппарата на основе анализа модуляционного спектра речи в критических полосах / М. И. Вашкевич, И. С. Азаров // Труды СПИИРАН. – 2020. – Т. 19. – №. 2. – С. 249–276.