УДК 004.4’277

Д.С. Лихачев, И.С. Азаров, А.А. Петровский

ПРИМЕНЕНИЕ МГНОВЕННОГО ГАРМОНИЧЕСКОГО АНАЛИЗА ДЛЯ АНТРОПОМОРФИЧЕСКОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

Рассматривается способ параметрического описания звукового сигнала, основанный на антропоморфической интерпретации его частотных составляющих. Для получения параметров модели предлагается использовать мгновенный гармонический анализ вместо дискретного преобразования Фурье. В работе оценивается точность полученного описания. Приводятся экспериментальные результаты, показывающие, что реконструкция сигнала в большой степени зависит от средств получения частотно-временного описания, причем предложенный способ обеспечивает более высокое качество реконструкции сигнала по сравнению с известными методами.

Введение

Моделирование свойств восприятия звуковой информации человеком является важной задачей современной цифровой обработки речевых сигналов. Подходы к перцептуальному моделированию слуховой системы человека с использованием антропоморфических принципов применяются в задачах автоматического распознавания речи [1], кодирования, конверсии голоса, повышения качества восприятия речи в агрессивной шумовой обстановке [2–6].

Применение антропоморфической обработки сигнала предполагает использование таких устройств и алгоритмов обработки информации, когда вычислительный процесс организовывается по «образу и подобию» человека, т. е. используемые способы и алгоритмы моделируют какие-либо процессы, происходящие в его слуховых и речеобразующих системах. При этом предполагается, что при достаточно точном моделировании используемые системы будут иметь те же полезные свойства, что и их физиологический аналог. В качестве основной задачи антропоморфической обработки ставится выработка такого критерия, который позволял бы отбирать доминирующие составляющие сигнала [7].

Задача моделирования слуховой системы человека сводится к математическому описанию процессов, происходящих в различных частях слухового аппарата человека на физиологическом уровне. В улитке уха речевой сигнал обрабатывается в частотных полосах посредством механических свойств базилярной мембраны [8]. Таким образом, сигнал воспринимается человеком как совокупность отдельных частотных компонент. Адекватным математическим эквивалентом данного процесса может служить частотно-временное преобразование. Выбор наиболее подходящего частотно-временного преобразования представляет особый интерес. При моделировании слуховой системы методами из [9] в качестве частотно-временного преобразования используется дискретное преобразование Фурье (ДПФ), хотя в контексте решаемой задачи данное преобразование имеет ряд недостатков: отсутствие возможности оценки амплитудной модуляции, низкая степень локализации энергии и строгая кратность частоты периодических составляющих. Перечисленные особенности приводят к избыточности описания и появлению слышимых артефактов при реконструкции сигнала [10], что заставляет искать альтернативные пути частотно-временного описания.

Идея, которая лежит в основе настоящей работы, заключается в применении мгновенного гармонического анализа для получения параметров модели. Метод оценки параметров, использованный в данной работе, основан на специальных фильтрах с модулированной импульсной характеристикой [11] и позволяет получить локализованное частотное описание речевого сигнала. Для сопоставления предлагаемого подхода с системой выделения доминирующих частот на основе кратковременного преобразования Фурье сравнивается качество реконструкции речевых сигналов, для чего была реализована система анализа/синтеза.

1. Перцептуальная оценка составляющих речевого сигнала с использованием антропоморфической обработки

Органы слуха человека представляют собой достаточно сложную систему с множеством составных частей, но условно ее можно разделить на две взаимозависимые части: периферическую часть (внешнее, среднее и внутреннее ухо) и слуховой нерв (частично волосковые клетки, непосредственно слуховой нерв и участки головного мозга, отвечающие за обработку импульсов от слухового анализатора).

Периферическая часть слуховой системы человека с каждой стороны представлена внешним ухом (ушная раковина и слуховой канал), средним ухом (барабанная перепонка и слуховые косточки) и внутренним ухом (улитка уха и полукружные каналы). Основным назначением ушной раковины (внешнего уха) является концентрация энергии и согласование импедансов воздушной среды свободного акустического поля и наружного слухового прохода, а основным предназначением системы среднего уха – согласование высокого входного импеданса улитки внутреннего уха, заполненной жидкостью, и сравнительно низкого импеданса воздушной среды в барабанной полости [8]. Таким образом, наружное и среднее ухо выступают в роли специального устройства согласования физических параметров внешней среды с характеристиками внутреннего уха. Наружный слуховой канал закрыт барабанной перепонкой, которая отделяет внешнее ухо от среднего. Барабанная перепонка представляет собой пластинку, имеющую форму эллипса. Звуковые волны воспринимаются внешним ухом и вызывают колебания барабанной перепонки, которые передаются по цепи слуховых косточек среднего уха и достигают внутреннего уха в виде специфических колебательных движений. Во внутреннем ухе эти колебания распространяются по спиральному каналу улитки как волны давления внутриулиточной жидкости. В улитке уха речевой сигнал разлагается на спектральные полосы посредством механических свойств базилярной мембраны.

Базилярная мембрана выполняет роль неоднородной резонансной линии передачи, в которой каждый частотный компонент вызывает резонанс (или вибрацию максимальной амплитуды) в зоне базилярной мембраны, определяемой частотой сигнала. Каждая точка по длине мембраны обладает уникальной полосой пропускания, а отображение множества точек по длине базилярной мембраны на частоту приблизительно логарифмическое.

С физиологической точки зрения спектральное разложение входного речевого сигнала осуществляется с помощью прилегающих к базилярной мембране внутренних волосковых клеток, которые функционируют в зависимости от уровня колебания мембраны [8]. Они преобразуют механическое движение базилярной мембраны в рецепторный потенциал путем высвобождения электрохимической субстанции, которая вызывает возбуждение прилегающих к клеточному телу нервных волокон. Местоположение внутренних волосковых клеток вдоль базилярной мембраны определяет соответствующие частоты возбуждения. Каждая внутренняя волосковая клетка снабжена примерно десятью нервными волокнами, а вся улитка связана более чем с 30 000 волокон, кодирующих до 2000–3000 сигналов от волосковых клеток.

Одной из эффективных и наиболее адекватных реальным физиологическим процессам моделей является представление слуховой системы человека на трех уровнях: в наружном и среднем ухе, в улитке уха человека и на уровне слухового нерва [5, 12] (рис. 1).

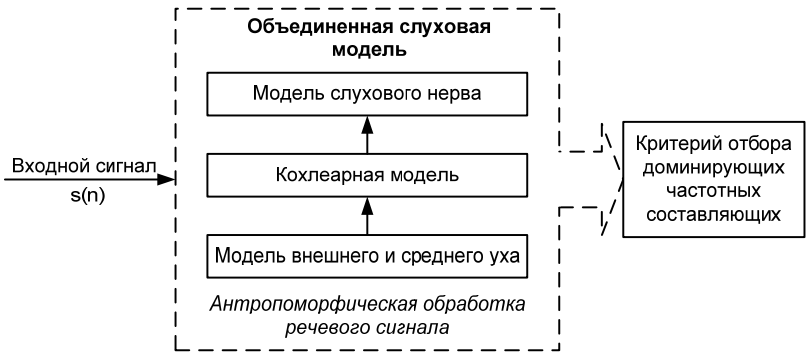


Рис. 1. Обобщенная схема перцептуальной оценки составляющих сигнала с использованием антропоморфической обработки

В качестве основной модели периферической части слуховой системы человека используется модель SDCM (Second Order Difference Cochlea Model) – разностная кохлеарная модель второго порядка [13–15] (рис. 2).

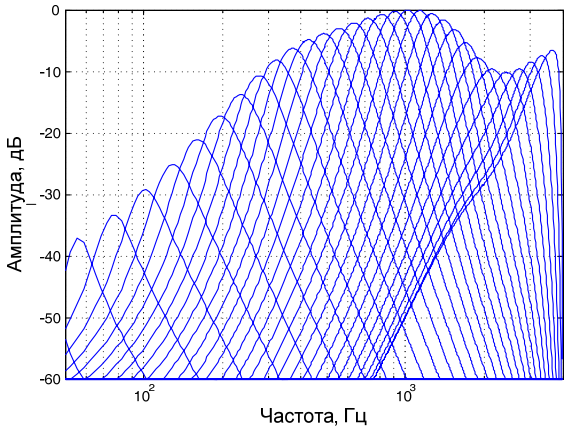


Рис. 2. Амплитудно-частотные характеристики
для 32 модифицированных кохлеарных фильтров

В соответствии с SDCM-моделью функционирование улитки уха описывается работой банка цифровых фильтров второго порядка:

$$y_m(n) + b_{1m} y_m(n-1) + b_{2m} y_m(n-2) = A_m a_{0m} [u_s(n) - u_s(n-2)], \quad (1)$$

где b_{1m}, b_{2m}, A_m и a_{0m} – параметры, которые определяются физическими свойствами базилярной мембраны в позиции x_m и изменяются вдоль базилярной мембраны; m – номер сегмента базилярной мембраны после дискретизации; $y_m(n)$ – перемещение, или так называемая пучность, базилярной мембраны в позиции x_m; $u_s(n)$ – входной синусоидальный сигнал, характеризующий скорость перемещения стремечка.

Горизонтальная координата на рис. 3 определяет изменение центральной частоты кохлеарного фильтра, а вертикальная координата – соответствующую ей полосу пропускания кохлеарного фильтра. В данном случае при вычислении амплитудно-частотных характеристик фильтров минимально возможная полоса пропускания принималась равной 80 Гц.

На рис. 4 изображена кохлеарная карта модели – отношение между частотой возбуждения (центральной частотой) и тем положением на базилярной мембране, которое претерпевает наибольшее смещение.

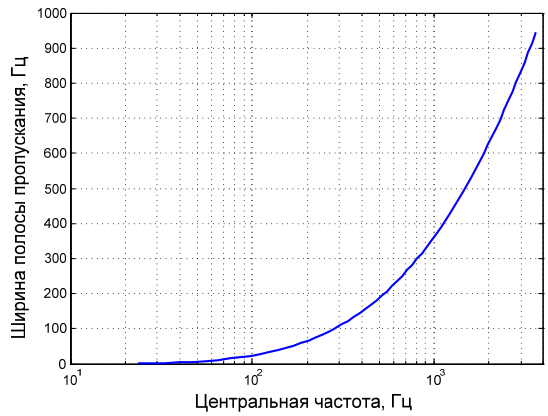


Рис. 3. Характеристика полосы пропускания
по уровню 3 дБ

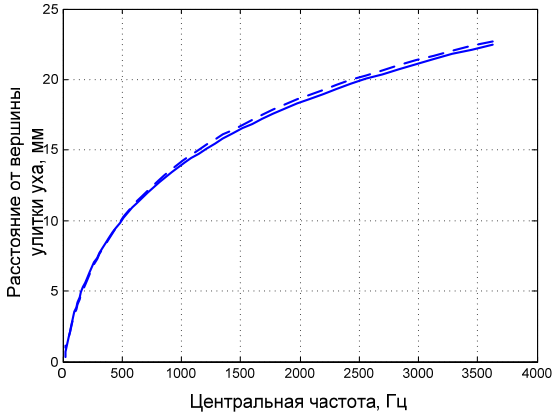


Рис. 4. Кохлеарная карта: штриховая линия –
экспериментальные данные из [16];
непрерывная линия – используемая модель

Хорошей степенью адекватности реальным физиологическим процессам обладает модель слуха человека, представленная в [5, 12, 17]. Результатом работы модели является так называемая слуховая гистограмма $G(f)$, которая позволяет получить представление об акустической информации, «циркулирующей» на уровне слухового нерва. С ее помощью можно дифференцировать частотные составляющие анализируемого речевого сигнала по степени их важности для человеческого слуха. Процесс вычисления модифицированной слуховой гистограммы как дискретной функции частоты $G(k)$ по методике из [10, 18] продемонстрирован на рис. 5.

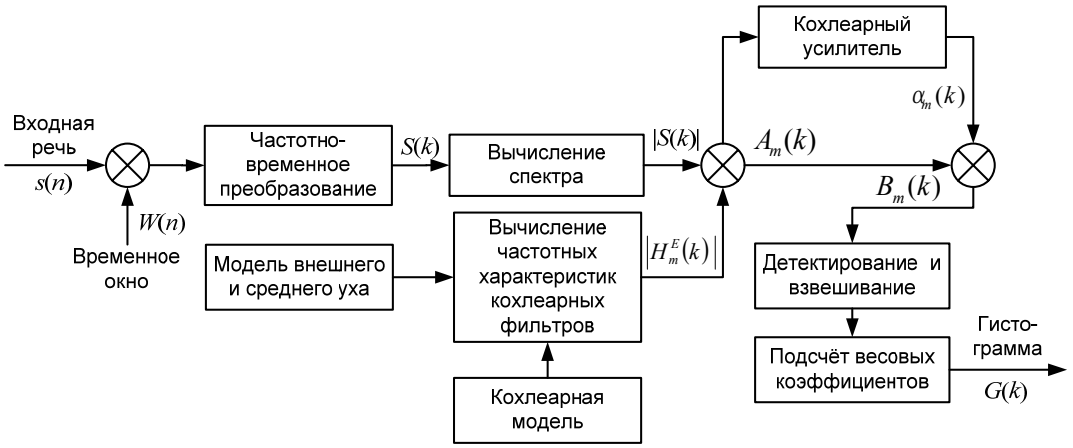


Рис. 5. Схема вычисления модифицированной слуховой гистограммы

В данном случае гистограмма $G(k)$ вычисляется с помощью выражения

$$G(k)=\sum_{m=1}^{M_F} G_m(k),\ \ k=\overline{1,\frac{N_F}{2}}, \tag{2}$$

где m – номер обрабатываемого в текущий момент времени кохлеарного канала; M_F – число кохлеарных фильтров; N_F – формат ДПФ; $G_m(k)$ – k-й элемент гистограммы для m-го кохлеарного фильтра, который может быть вычислен с помощью следующего алгоритма [7]:

Шаг 1. Спектр сигнала $|S(k)|$ взвешивается с помощью весовой функции $|H_m(k)|$, представляющей собой нормированную амплитудно-частотную характеристику m-го кохлеарного фильтра:

$$A_m(k)=|S(k)|\cdot|H_m(k)|. \tag{3}$$

Шаг 2. Определяется коэффициент усиления $\alpha_m(k)$ в зависимости от величины $A_m(k)$ [5]:

$$\alpha_m(k)=40-10\log_{10}[A_m(k)],\ \text{дБ}. \tag{4}$$

Шаг 3. Взвешенный спектр обрабатывается с помощью кохлеарного усилителя:

$$B_m(k)=A_m(k)\cdot 10^{\frac{\alpha_m}{20}}. \tag{5}$$

Шаг 4. Определяются взвешивающие коэффициенты $p_m(k)$ по следующему правилу:

$$p_m(k)=\begin{cases}\beta_l, & \text{если } B_m(k)\ge U_l,\ l=\overline{1,L};\\ 0, & \text{если } B_m(k)<U_l,\end{cases} \tag{6}$$

где l – номер уровня; L – количество уровней, на которое разбивается анализируемый амплитудный диапазон сигнала; U_l – пороговое значение амплитуды для l-го уровня; β_l – постоян-

ная величина, характеризующая степень нервного возбуждения, каждому уровню сопоставляется свое собственное значение ($\beta_l = 2^l$).

Шаг 5. Вычисляются элементы слуховой гистограммы:

$$G_m(k) = B_m(k) \cdot p_m(k). \tag{7}$$

При обработке полученных результатов предполагается, что в каждом кохлеарном канале выбирается одна доминирующая составляющая (рис. 6).

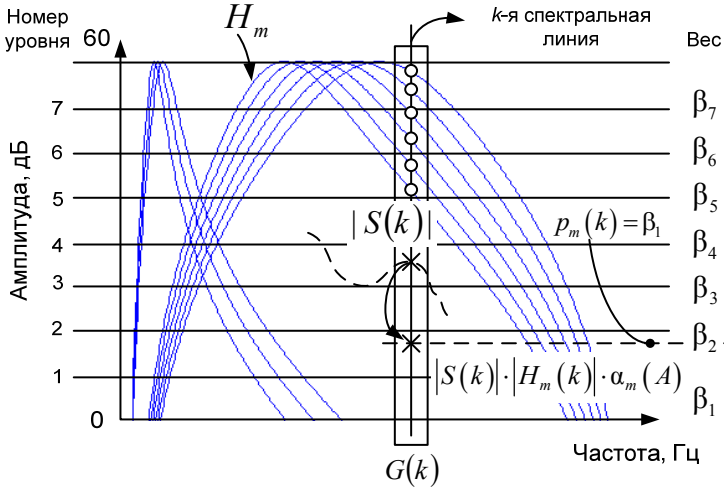


Рис. 6. Процесс вычисления *k*-го элемента гистограммы в частотной области

Привлекательность описанного выше моделирования слуховой системы человека заключается, во-первых, в наличии адекватной интерпретации физических и физиологических процессов и, во-вторых, в практической его применимости для обработки аудио- и речевых сигналов. Возможность выбора перцептуально значимых составляющих позволяет, с одной стороны, практически исследовать свойства восприятия человеком аудиоинформации, с другой – избавиться от избыточности в системах, требующих параметрического представления сигнала.

2. Разделение сигнала на частотные компоненты и их параметрическое описание

Для того чтобы смоделировать трансформацию сигнала слуховой системой человека при помощи вышеизложенного способа антропоморфической обработки, необходимо выполнить предварительное параметрическое описание сигнала в частотно-временной области. При этом сигнал должен быть разделен на отдельные частотные составляющие, параметры которых меняются во времени. Для этого удобно применять синусоидальную модель, изначально предложенную для кодирования речи [19] и впоследствии удачно использованную во многих других приложениях.

Таким образом, сигнал $s(n)$ можно записать в виде соотношения

$$s(n) = \sum_{k=1}^{K} C_k(n) \cos \varphi_k(n) + r(n), \tag{8}$$

где $C_k(n)$ – мгновенная амплитуда *k*-й синусоиды; K – число синусоид; $\varphi_k(n)$ – мгновенная фаза *k*-й синусоиды; $r(n)$ – сигнал-остаток.

Мгновенная фаза $\varphi_k(n)$ и мгновенная частота $f_k(i)$ соотносятся следующим образом:

$$\varphi_k(n) = \sum_{i=1}^{n} \frac{2\pi f_k(i)}{F_S} + \varphi_k(0), \tag{9}$$

где F_S – частота дискретизации; $\varphi_k(0)$ – начальная фаза *k*-й синусоиды.

В настоящее время существует довольно много методов обработки речевых сигналов на основе синусоидальной модели [20], однако цель их одинакова – разложить речевой сигнал на отрезки определенной длины (фреймы), представить их в виде совокупности синусоидальных компонент, выполнить действия по изменению его характеристик и затем каким-либо способом синтезировать (восстановить) речевой сигнал.

Задача параметрического описания сигнала заключается в определении синусоидальных параметров $C_k(n)$, $f_k(n)$ и $\varphi_k(n)$ для заданного момента или интервала времени. Предполагается, что синусоидальные компоненты разделены в частотной области (их можно выделить на всем протяжении анализируемого фрейма фильтрами с неперекрывающимися полосами пропускания), а их амплитуда и частота изменяются медленно, поэтому можно считать, что каждая синусоида может быть ограничена в частотной области узкой частотной полосой. Таким образом, искомые параметры синусоидальной модели $C_k(n)$ и $f_k(n)$ являются гладкими, непрерывными функциями с ограниченным частотным диапазоном.

Оценка синусоидальных параметров является фундаментальной задачей. Точность оценок, как правило, оказывает существенное влияние на качество реконструкции описываемого сигнала.

Основным инструментом для выполнения гармонического анализа служит кратковременное ДПФ [20]. В этом случае предполагается, что анализируемый сигнал является квазистационарным, т. е. на протяжении некоторого периода времени его параметры остаются неизменными. Несмотря на то что с помощью ДПФ были получены достаточно неплохие результаты [7], все же допущение локальной стационарности сигнала не всегда позволяет достичь требуемой точности анализа. Прежде всего, используя ДПФ, сложно получить адекватное параметрическое описание неустойчивых тональных звуков. Другая проблема заключается в сложности анализа сигналов с быстро изменяющимся тоном. Например, ДПФ очень ограниченно применимо к оценке параметров гармоник высокого порядка вокализованной речи из-за свойственного им быстрого изменения частоты. Отдельную проблему при использовании ДПФ в контексте антропоморфической обработки представляют выравнивание частотных компонент по частотной сетке и спектральная утечка, заключающаяся в перераспределении энергии на соседние спектральные составляющие. Это приводит к появлению дополнительных частотных составляющих, которые отдельно оцениваются кохлеарной моделью. На рис. 7 показан пример анализа монокомпонентного периодического сигнала, который вследствие несоответствия его периода и формата преобразования представляется в частотной области множеством смежных частотных компонент. В приведенном примере сигнал представлен пятью компонентами.

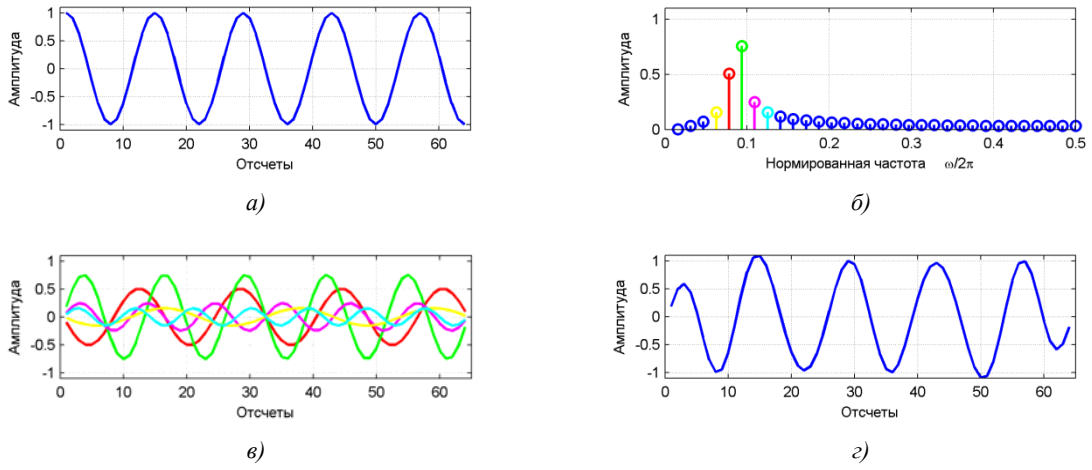


Рис. 7. Определение гармонических параметров при помощи преобразования Фурье: *а)* исходный периодический сигнал; *б)* амплитудный спектр; *в)* набор основных частотных составляющих исходного сигнала; *г)* реконструированный сигнал

В качестве альтернативы ДПФ предлагается использовать технику мгновенного гармонического анализа, основанную на узкополосной фильтрации [21]. Импульсная характеристика фильтра анализа $h(n)$ выражается в следующем аналитическом виде:

$$h(n)=\begin{cases}1, & n=0;\\ \dfrac{F_S}{n\pi}\cos\left(\dfrac{2\pi n}{F_S}F_C\right)\sin\left(\dfrac{2\pi n}{F_S}F_\Delta\right)\\ \hline 2F_\Delta\end{cases}, \quad n\neq 0, \tag{10}$$

где F_C – середина полосы пропускания, Гц; $2F_\Delta$ – ширина полосы пропускания; F_S – частота дискретизации. Выходной сигнал фильтра, представляющий собой свертку входного сигнала $s(n)$ импульсной характеристикой (10), может быть представлен в виде синусоиды с мгновенной амплитудой $C(n)$, фазой $\varphi(n)$ и частотой $f(n)$, определяемыми следующими выражениями:

$$C(n)=\sqrt{A^2(n)+B^2(n)}\,; \tag{11}$$

$$\varphi(n)=\arctan\left(\frac{-B(n)}{A(n)}\right); \tag{12}$$

$$f(n)=\frac{\varphi(n+1)-\varphi(n)}{2\pi}F_S\,, \tag{13}$$

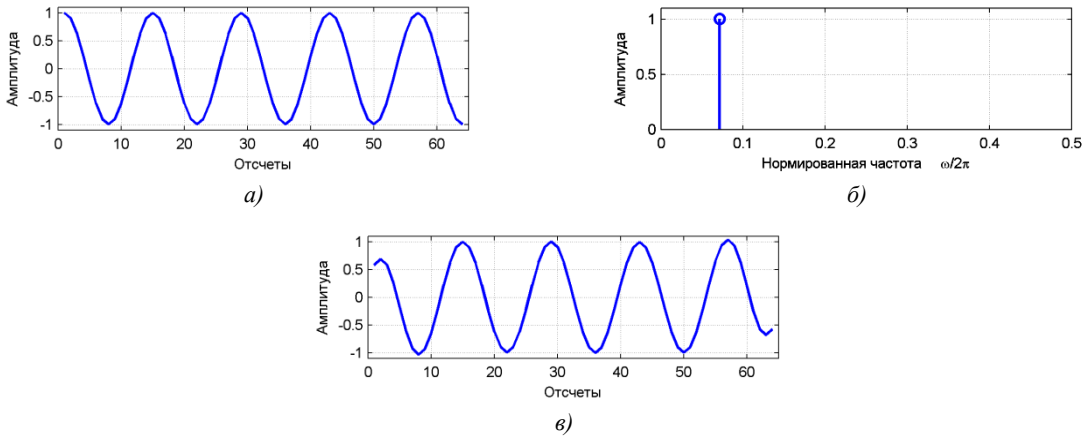
где

$$A(n)=\sum_{i=0}^{N-1}\frac{2s(i)}{\pi(n-i)}\sin\left(\frac{2\pi(n-i)}{F_S}F_\Delta^k\right)\cos\left(\frac{2\pi(n-i)}{F_S}F_C^k\right); \tag{14}$$

$$B(n)=\sum_{i=0}^{N-1}\frac{-2s(i)}{\pi(n-i)}\sin\left(\frac{2\pi(n-i)}{F_S}F_\Delta^k\right)\sin\left(\frac{2\pi(n-i)}{F_S}F_C^k\right). \tag{15}$$

Мгновенные гармонические параметры выходного сигнала могут быть рассчитаны в любой момент времени, принадлежащий анализируемому фрейму сигнала, причем этот момент не ограничивается дискретными отсчетами сигнала, так как выход фильтра записан в виде непрерывных функций. Очевидно, что полоса пропускания фильтра, задаваемая параметрами F_C и F_Δ, должна содержать анализируемый компонент.

На рис. 8 изображено параметрическое представление синусоиды при помощи мгновенного гармонического анализа. В отличие от ДПФ-декомпозиции, представленной на рис. 7, сигнал интерпретируется как один частотный компонент.



а) *б)* *в)*

Рис. 8. Определение мгновенных гармонических параметров:
а) исходный периодический сигнал; *б)* амплитудный спектр; *в)* реконструированный сигнал

На схеме процедуры оценки мгновенных параметров, выполненной на основе фильтров синусоидального анализа (рис. 9), видно, что входной сигнал анализируется и разделяется на периодическую компоненту и остаток, которые вместе с полученными гармоническими параметрами передаются последующим блокам анализа и кодирования.

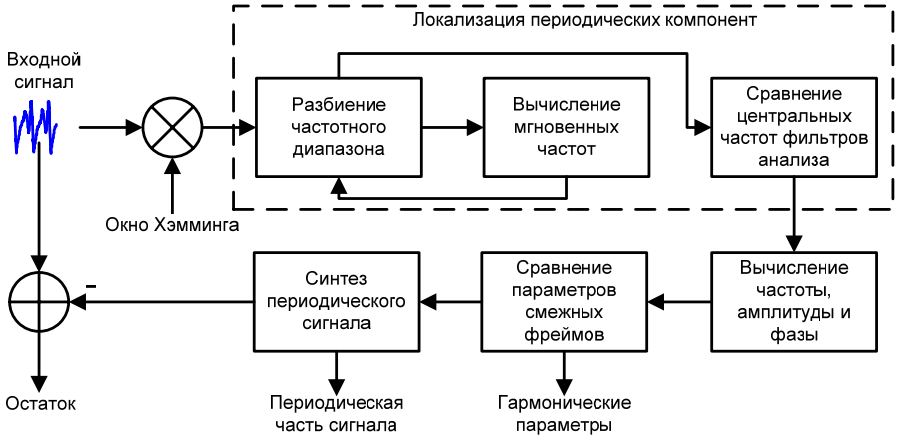


Рис. 9. Схема гармонического анализатора для звукового сигнала

В соответствии с применяемой схемой анализа входной сигнал разбивается на фреймы с перекрытиями, которые анализируются отдельно друг от друга при помощи узкополосной фильтрации. Полученные гармонические параметры смежных фреймов сравниваются при помощи буфера слежения для выявления длинных и стабильных гармонических компонент. С помощью полученных параметров синтезируется синусоидальный сигнал, который вычитается из исходного для получения остатка.

Данная схема анализа была использована в составе гибридного кодера для звука и речи [22]. Благодаря выявлению продолжительных и стабильных компонент гармонический анализатор способен обрабатывать непосредственно входной сигнал, поступающий на вход системы кодирования, без какой-либо предварительной обработки.

3. Результаты экспериментов

Целью проведенных экспериментов являлась оценка качества представления сигналов параметрической моделью на основе антропоморфической обработки при использовании различных способов частотно-временного преобразования. Для этого была реализована система анализа/синтеза, использующая в качестве частотно-временного преобразования ДПФ и метод мгновенного гармонического анализа (рис. 10).

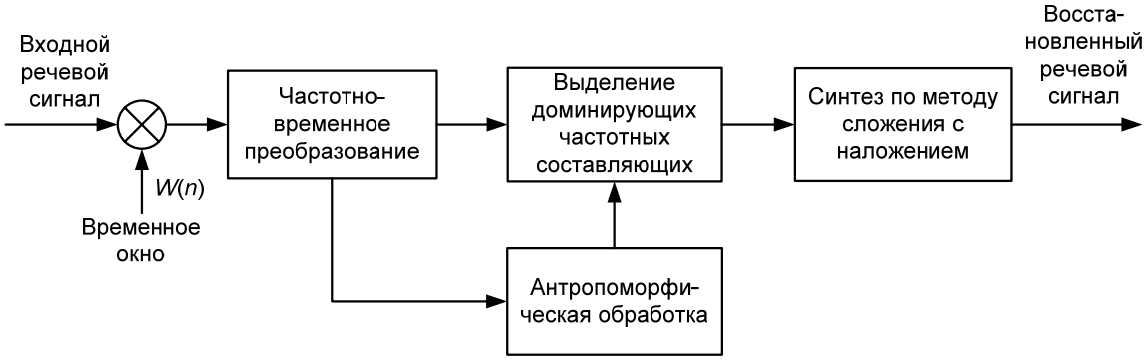


Рис. 10. Система анализа/синтеза

Уровень слышимых искажений, вносимых параметрической моделью определялся путем сравнения реконструированного сигнала с исходным. Для получения объективных оценок ис-

пользовались техники PESQ (Perceptual Evaluation of Speech Quality) и MBSD (Modified Bark Spectral Distortion).

Алгоритм PESQ представляет собой объективную методику определения качества речевых сообщений, передаваемых по каналам связи, которая прогнозирует результаты субъективной оценки качества слушателями-экспертами. Для получения оценки используется специальное программное обеспечение, реализующее методику PESQ [23]. Для определения качества реконструкции речи сравнивается входной (эталонный) сигнал с его обработанной версией. Сигналы представлены в виде файлов формата WAV. Результатом сравнения входного и выходного сигналов является оценка качества речевого сообщения, которая аналогична усредненной субъективной оценке MOS (Mean Opinion Score). Оценка PESQ характеризует субъективную оценку реконструкции, изменяющуюся в пределах от 1 (плохо) до 5 (отлично).

Оценка искажений спектра барков MBSD представляет собой усредненную величину воспринимаемых человеком искажений, присутствующих в восстановленном речевом сигнале (0 соответствует полному отсутствию искажений). Вычисление значений оценок искажений спектра барков MBSD производится с помощью специального пакета программ в среде Matlab с использованием образцов оцифрованных сигналов, представленных в виде файлов формата WAV [24].

Речевая выборка, используемая для экспериментов, содержала образцы женской и мужской речи общей продолжительностью 2 мин. Для проверки качества реконструкции речи в шумовой обстановке к исходным сигналам добавлялся белый шум различной интенсивности. Результаты влияния частотно-временного преобразования на качество реконструкции речи при сохранении моделью всех слышимых компонент приведены в табл. 1 и 2.

Судя по оценке PESQ, качество реконструкции речевого сигнала выше в случае использования мгновенного гармонического анализа и меньше зависит от энергии аддитивного шума по сравнению с вариантом на основе ДПФ.

Оценка MBSD показывает, что схема анализ/синтез на основе мгновенного гармонического анализа является более предпочтительной, поскольку при реконструкции речевого сигнала она вносит значительно меньше искажений спектра барков. Большие значения оценки MBSD при низких значениях SNR (отношения сигнал/шум) связаны с особенностями вычисления оценки перцептуальной значимости составляющих сигнала, которые различны для антропоморфической модели, выбранной в данной работе, и методики MBSD.

Таблица 1

Оценка PESQ реконструированного сигнала с различным уровнем зашумления (ДПФ/мгновенный гармонический анализ)

Голоса	SNR, дБ				
	60	20	10	0	–10
Мужские	2,97/3,49	2,62/3,54	2,64/3,65	2,59/3,61	2,63/3,53
Женские	3,21/3,41	3,30/3,51	3,32/3,62	3,07/3,64	3,13/3,53

Таблица 2

Оценка MBSD реконструированного сигнала с различным уровнем зашумления (ДПФ/мгновенный гармонический анализ)

Голоса	SNR, дБ				
	60	20	10	0	–10
Мужские	0,46/0,18	0,56/0,18	1,05/0,68	3,73/3,22	10,27/10,42
Женские	0,52/0,24	0,89/0,40	1,67/1,40	4,50/4,54	10,95/12,20

Было проведено исследование влияния частотно-временного преобразования на качество реконструкции речевого и аудиосигналов при сохранении фиксированного числа слышимых компонент с наибольшей перцептуальной значимостью. Данный эксперимент, как и предыдущий, проводился для женских и мужских голосов отдельно. Полученные оценки PESQ и MBSD

приведены в табл. 3 и 4. Видно, что близкое качество реконструкции сигнала достигается меньшим количеством компонент в случае использования мгновенного гармонического анализа.

Таблица 3

Оценка PESQ реконструированного сигнала с фиксированным количеством частотных компонент (ДПФ/мгновенный гармонический анализ)

Голоса	Максимальное число компонент			
	5	10	15	20
Мужские	2,21/2,31	2,59/3,11	2,85/3,38	2,97/3,49
Женские	2,46/2,49	3,16/3,23	3,36/3,40	3,21/3,41

Таблица 4

Оценка MBSD реконструированного сигнала с фиксированным количеством частотных компонент (ДПФ/мгновенный гармонический анализ)

Голоса	Максимальное число компонент			
	5	10	15	20
Мужские	1,33/0,76	0,62/0,25	0,48/0,19	0,46/0,18
Женские	1,74/1,05	0,66/0,22	0,51/0,25	0,52/0,24

Результаты эксперимента свидетельствуют о том, что для модели антропоморфического представления сигнала более предпочтительным частотно-временным преобразованием является метод мгновенного гармонического анализа, поскольку он обеспечивает более высокое качество реконструкции сигнала с пониженным уровнем спектральных искажений. Результат объясняется высокой степенью локализации энергии в частотной области.

Заключение

В работе приведены результаты исследования применимости мгновенного гармонического анализа к параметрическому описанию сигналов на основе антропоморфического подхода. В ходе исследования была реализована модель слуховой системы человека, использующая банк кохлеарных фильтров для выделения слышимых компонент. На основе данной модели была выполнена система анализа/синтеза, позволяющая синтезировать волновую форму сигнала, прошедшего через модель слуховой системы, и оценить качество реконструкции. В результате экспериментов были получены объективные оценки, показывающие, что антропоморфическая обработка речи с применением техники мгновенного гармонического анализа вместо преобразования Фурье обеспечивает более высокое качество ее реконструкции. Это обусловлено тем, что антропоморфическая оценка перцепционной значимости спектральных составляющих оказывается более точной. Предложенный метод может быть полезным для антропоморфической обработки зашумленной речи.

Список литературы

1. Morgan, N. Does ASR have a PHD, or is it just piled higher and deeper? / N. Morgan [Electronic resource]. – Mode of access : http://superlectures.com/icassp2011/lecture.php?id=206&lang=en. – Date of access : 21.10.2011.

2. A Perceptual Model for Sinusoidal Audio Coding Based on Spectral Integration / S. van de Par [et. al.] // EURASIP Journal on Applied Signal Processing. – 2005. – Vol. 2005, № 9. – P. 1292–1304.

3. Ravindran, S. A Physiologically Inspired Method for Audio Classification / S. Ravindran, K. Chlemmer, D.V. Anderson // EURASIP Journal on Applied Signal Processing. – 2005. – Vol. 2005, № 9. – P. 1374–1381.

4. Feldbauer, C. Anthropomorphic Coding of Speech and Audio: A Model Inversion Approach / C. Feldbauer, G. Kubin, W.B. Kleijn // EURASIP Journal on Applied Signal Processing. – 2005. – Vol. 2005, № 9. – P. 1334–1349.

5. Ghitza, O. Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition / O. Ghitza // IEEE Transactions on Speech and Audio Processing. – 1994. – Vol. 2, № 1. – P. 115–132.

6. Ivanov, A.V. Analysis of the IHC Adaptation for the Anthropomorphic Speech Processing Systems / A.V. Ivanov, A.A. Petrovsky // EURASIP Journal on Applied Signal Processing. – 2005. – Vol. 2005, № 9. – P. 1323–1333.

7. Лихачев, Д.С. Анализ и синтез устройств кодирования речевого сигнала на основе антропоморфической обработки и синусоидальных моделей / Д.С. Лихачев, А.А. Петровский // Доклады БГУИР. – 2006. – № 3 (15). – С. 35–43.

8. Слуховая система / Я.А. Альтман [и др.] ; под общ. ред. Я.А. Альтмана. – Л. : Наука, 1990. – 620 с.

9. Likhachov, D.S. Improved auditory-based speech coding using psychoacoustic model based on a cochlear filter bank and an average localized synchrony detection / D.S. Likhachov, A.A. Petrovsky // Computer information systems and industrial management applications ; eds. K. Saeed, R. Mosdorf, Z. Sosnowski. – Poland : Bialystok, 2003. – P. 11–19.

10. Лихачев, Д.С. Компрессия речевого сигнала на основе синусоидальной модели с антропоморфической обработкой / Д.С. Лихачев, А.А. Петровский // Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика (с MATLAB-примерами) ; под ред. д.т.н. профессора А.А. Петровского. – Минск : Бестпринт, 2009. – С. 211–233.

11. Азаров, И.С. Вычисление мгновенных гармонических параметров речевого сигнала / И.С. Азаров, А.А. Петровский // Речевые технологии. – 2008. – № 1 (1). – С. 67–77

12. Ghitza, O. Adequacy of auditory models to predict internal human representation of speech sounds / O. Ghitza // J. Acoust. Soc. Am. – 1993. – Vol. 93, № 4. – P. 2160–2171.

13. An anthropomorphic speech processing based on the cochlear model and its application for coding task / A.A. Petrovsky [et al.] // International scientific journal of computing. – 2004. – Vol. 3, № 1. – P. 75–83.

14. Wan, W.G. A two-dimentional non-linear cochlear model for speech processing: response to pure tones / W.G. Wan, A.A. Petrovsky, C.X. Fan // 6th Intern. Fase-Congress. – Zurich, Switzerland, 1992. – P. 233–236.

15. Wan, W.G. A new solution for cochlear macromechanics / W.G. Wan, C.X. Fan // Acustica. – 1991. – Vol. 75. – P. 79–82.

16. Greenwood, D.D. A cochlear frequency-position function for several species-29 years later / D.D. Greenwood // J. Acoust. Soc. Am. – 1990. – Vol. 87, № 6. – P. 2592–2605.

17. Petrovsky, A.A. A digital cochlear model as a base of anthropomorphic speech processing / A.A. Petrovsky, D.S. Likhachov // Neural networks and artificial intelligence : proc. of the 3d Intern. Conf., Belarus, Minsk, November 12–14, 2003. – Minsk, 2003. – P. 126–131.

18. Лихачев, Д.С. Антропоморфический анализ на основе дискретного преобразования Фурье с неравномерной частотной шкалой / Д.С. Лихачев // Известия Белорусской инженерной академии. – 2005. – № 1 (19)/2. – С. 177–180.

19. McAulay, R.J. Low-rate speech coding based on the sinusoidal model / R.J. McAulay, T.F. Quatieri // Advances in Speech Signal Processing ; eds. S. Furui, M.M. Sondhi. – N.Y. : Marcel Dekker, 1992. – P. 165–208.

20. McAulay, R.J. Speech analysis/synthesis based on a sinusoidal representation / R.J. McAulay, T.F. Quatieri // IEEE Trans. on Acoust., Speech and Signal Processing. – 1986. – Vol. ASSP-34. – P. 744–754.

21. Азаров, И.С. Непрерывное и дискретное гармонические преобразования для декомпозиции речевого сигнала на периодическую и шумовую компоненты / И.С. Азаров, А.А. Петровский // Доклады БГУИР. – 2008. – № 4 (34). – С. 92–105.

22. Petrovsky, A. Combining advanced sinusoidal and waveform matching models for parametric audio/speech coding / A. Petrovsky, E. Azarov, A. Petrovsky // EUSIPCO 2009 : proc. of the 17th European Signal Processing Conf. – Glasgow, 2009. – P. 436–440.

23. ITU-T Recommendation P.862, PESQ an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, February 2001.

24. Yang, W. Enhanced Modified Bark Spectral Distortion (EMBSD): an Objective Speech Quality Measure Based on Audible Distortion and Cognition Model (PhD Thesis) / W. Yang [Electronic resource]. – Mode of access : http://www.temple.edu/speech_lab/wonhos_dissertation.pdf. – Date of access : 21.10.2011.

Поступила 04.07.11

Белорусский государственный университет
информатики и радиоэлектроники,
Минск, ул. П. Бровки, 6
e-mail: likhachov@bsuir.by

D.S. Likhachov, E.S. Azarov, A.A. Petrovsky

APPLYING INSTANTANEOUS HARMONIC ANALYSIS TO THE ANTHROPOMORPHIC PROCESSING OF AUDIO AND SPEECH

The paper presents a method of parametric representation of audio signals based on anthropomorphic interpretation of its frequency components. An instantaneous harmonic analysis technique is applied instead of the Fourier transform. The suggested analysis technique provides a more accurate estimation of sinusoidal parameters. It is experimentally proven that objective quality of the reconstructed signals depends on the time-frequency transform and the proposed technique provides higher quality reconstruction as compared to the traditionally used Fourier transform.