

К возможности разделения сложного сигнала на квазигармонические «голоса»

Е.А.Ващенко, М.А.Витушко, В.С.Переверзев-Орлов

Институт проблем передачи информации им. А.А. Харкевича, Российская академия наук, Москва, Россия

Поступила в редколлегию 16.07.2018

Аннотация—Квазигармонические голоса составляют наиболее заметную часть в окружающих нас сигналах. Это дает основания заняться актуальными проблемами их анализа. В статье предлагается метод анализа сложного сигнала и выделения из него отдельного голоса, основывающийся на возможностях гребенчатой фильтрации. Показано, что такой подход позволяет эффективно выделять отдельные квазигармонические голоса из сложных сигналов типа записей оркестровой и оперной музыки и дает тем самым основания для развития “гребенчатой” фильтрации квазигармонических сигналов как альтернативы широко используемому спектральному Фурье-представлению.

КЛЮЧЕВЫЕ СЛОВА: сложный сигнал, выделение голосов, гребенчатая фильтрация.

1. ВВЕДЕНИЕ

В статье рассматривается возможность решения задачи выделения отдельного голоса из сложного многоголосого сигнала, используя для фильтрации гребенчатые фильтры. Сегоднешнее состояние проблемы анализа разноголосого сложного сигнала можно оценить с помощью обзора [1], в котором упоминается и сходная с нашей по постановке задача [2,3]. Существуют и программы для работы со звуковыми сигналами, позволяющие производить различные преобразования этих сигналов, включая выделение и подавление отдельных голосов. Однако, нам не удалось найти ссылки на работы с использованием гребенчатой фильтрации для решения рассматриваемой задачи.

Важно подчеркнуть, что предлагаемый в статье подход принципиально отличается от известного Фурье-анализа полным отказом от спектрального представления акустических сцен при их анализе и разделении композиций квазигармонических голосов во входном сигнале и заменой синусоидальной парадигмы фильтрации гребенчатой. Получаемый при этом метод анализа будем называть “Струнным анализом”, поскольку именно струна генерирует элементарный квазигармонический сигнал с видом спектрального ядра, соответствующим гребенчатым фильтрам.

Пусть $S(t)$ – квазипериодический сигнал, то есть $S(t) \approx S(t + T)$, где T – в каком-то разумном смысле “период” основного тона сигнала $S(t)$ с частотой F . Тогда, как и в случае периодического сигнала, этот сигнал в спектральной области будет представлен всплесками энергии на частотах nF , где n – номер гармоники (основной тон – первая гармоника).

Формально квазипериодичность можно рассматривать как следствие достаточно слабой по интенсивности и низкочастотной по сравнению с F амплитудно-частотной модуляции сигналом $E(t)$ периодического сигнала $S_0(t)$. В этом случае и F , и ее гармоники nF окажутся несущими, синхронно модулированными сигналом $E(t)$.

Следствием такого рода модуляции является уширение спектральных линий, соответствующих в спектре сигнала $S_0(t)$ его основному тону и гармоникам. При слабой модуляции уширение спектральных линий будет незначительным, обеспечивая возможность выделения таких несущих с их модуляционными спектрами с помощью потенциально пассивных фильтров. Специальный интерес в этом смысле представляют “гребенчатые” фильтры [4], позволяющие одномоментно фильтровать сигнал не одной, а множеством спектральных полос [5].

Возможности такого рода фильтрации используются нами для выделения отдельных квазипериодических голосов из их смеси. Решение этой задачи предлагается в виде многофазной процедуры, на каждом этапе которой создаются все более четкие “трековые сцены”, аналогичные спектральным сценам в Фурье-представлении. “Трековые сцены” используются для отслеживания и последующего выделения отдельного голоса.

2. ПЕРВИЧНОЕ ГРЕБЕНЧАТОЕ ОТОБРАЖЕНИЕ СИГНАЛА В СПЕКТРАЛЬНО-ВРЕМЕННУЮ ОБЛАСТЬ – «ТРЕКОВУЮ СЦЕНУ»

Простейшим (базовым) вариантом гребенчатого фильтра является фильтр с суммацией равнозадержанных отсчетов фильтруемого сигнала, алгоритм действия которого определяется следующим выражением:

$$S_{out}(t) = \frac{1}{n+1} \sum_{i=0}^n S_{in}(t - iT).$$

Здесь S_{in} – входной сигнал, S_{out} – выходной сигнал, T – период основного тона, n – число периодов фильтрации, t – время.

Такие фильтры реализуются с помощью n -элементных линий задержки. В приведенной на Рис. 1 частотной характеристике фильтра $n = 8$.

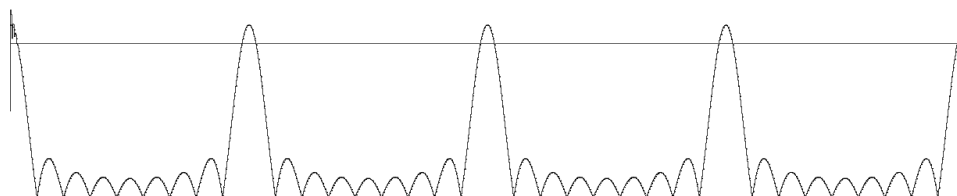


Рис. 1. Фрагмент частотной характеристики базового гребенчатого фильтра, реализуемого 8-элементной линией задержки. Максимальные всплески селективности – на частотах гармоник основного тона фильтра F , равного $1/T$, где T – время задержки сигнала одним элементом.

Видны интенсивные боковые лепестки. Максимальные всплески селективности – на частотах гармоник основного тона фильтра F , равного $1/T$, где T – время задержки сигнала одним элементом. Вид частотной характеристики сильно зависит от параметров фильтра. Чтобы уменьшить интенсивность паразитных лепестков – дополнительных полос пропускания, сигнал на выходе которых еще достаточно велик – были найдены такие модификации базового фильтра, которые обладали этим недостатком в существенно меньшей мере.

Блок-схема гребенчатого фильтра и частотные характеристики для базовых фильтров с различными параметрами приведены в *Приложении А*. Там же детально описаны две модификации базового фильтра, использующиеся в процедуре формирования “Трековой сцены”, и представлены их частотные характеристики.

Сканируя сигнал фильтрами такого рода с монотонно меняющейся величиной F в диапазоне от нижней до верхней частот спектра анализируемого сигнала, можно получить первичное

отображение сигнала в “Трековую сцену”. Вариант такого рода сцены показан на Рис. 2. Использовался 16-ти элементный гребенчатый фильтр для тестового сигнала в виде композиции двух полусинусоид с одинаковыми амплитудами и частотами 200 и 300 Гц, где второй сигнал имеет еще и небольшую частотную модуляцию низкочастотным синусоидальным сигналом. Полусинусоидой здесь считается сигнал, получаемый после замены отрицательных значений синусоиды нулем. Это было сделано для того, чтобы у компонентов теста возникли богатые собственные спектры.

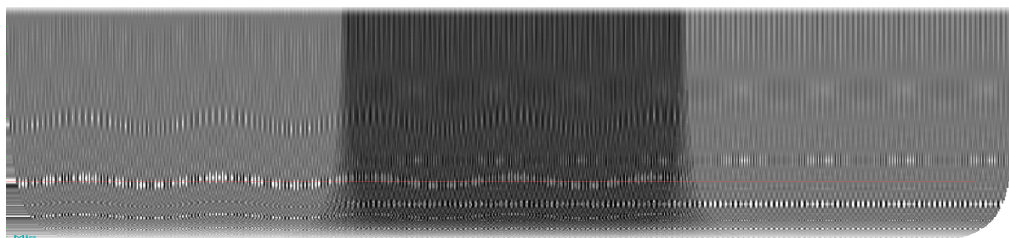


Рис. 2. Вид фрагмента первичной “Трековой сцены”. В левой нижней части – трехсотгерцовый сигнал с частотной модуляцией, а непосредственно под ним со сдвигом по времени на 1/3 экрана – трек второго компонента теста с постоянной частотой 200 Гц. Выше и ниже этих двух располагаются треки гармоник и субтонов компонентов теста.

Обычная спектрограмма этого же тестового сигнала, полученная при наилучших по спектральному разрешению параметрах с помощью программного инструмента Cool Edit, показана на Рис. 3.

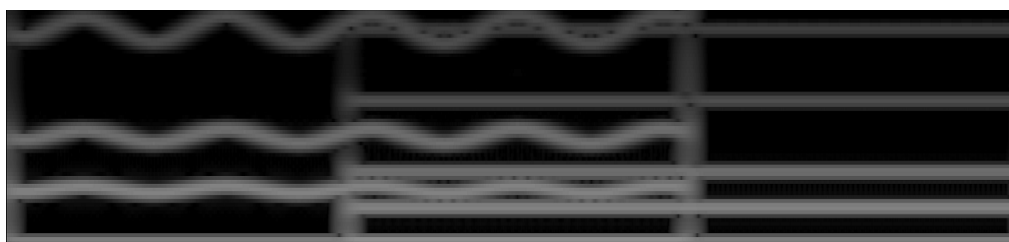


Рис. 3. Фрагмент стандартной спектрограммы с наилучшими параметрами разложения. Cool Edit Pro, FFT – 512 отсчетов.

Из сопоставления первичной “Трековой сцены” и спектрального представления того же сигнала следует, что первичная сцена несет в себе лишь намек на спектральное представление использованного тестового сигнала. Цель, однако, состоит в выделении треков голосов, а не спектральный анализ, и потому, прежде чем переходить к новым фазам выделения треков в тестовом сигнале, попробуем разобраться в том, что получается на уровне первичной трековой сцены.

Рассмотрим следующее рассуждение.

Как несложно видеть, гребенчатые фильтры (типа рассмотренного выше) откликаются не только на квазипериодические сигналы с частотой основного тона F , но также и на сигналы с частотами основных тонов, образующими по отношению к F “гармонический ряд” следующего вида:

$$F_k = kF, \quad k = (i \text{ XOR } 1/i), \quad i \in 1, \dots, N.$$

Здесь частоты F_k для $k \leq 1$ соответствуют субтонам сигнала с основным тоном F , а для $k \geq 1$ соответствуют гармоникам этого сигнала.

Таким образом, спектральный отклик гребенчатого фильтра для частоты основного тона F будет иметь регулярную структуру, всплески которой соответствуют частотам, находящимся в гармоническом отношении с F .

При этом, если не учитывать ограничения по частоте, энергия всплеска, соответствующего F будет максимальной, а энергии более низкочастотных и высокочастотных всплесков будут монотонно убывать по мере их удаления от основного.

Природа этого эффекта заключается в том, что при отсутствии ограничений на число зубьев гребенчатого фильтра, но при ограничении на ширину спектра анализируемого сигнала все гармоники сигналов, совпадающих по частоте основного тона с основным (F) для фильтра, обязательно будут попадать на все соответствующие им зубья гребенчатого фильтра, обеспечивая на выходе неискаженное воспроизведение этих сигналов, включая и энергию. В то же время для тех сигналов, частота основного тона которых совпадает с субтонами или гармониками F , ситуация оказывается несколько иной.

Действительно, если взять сигнал с половинной относительно основной частотой $F/2$, то все четные гармоники такого сигнала обязательно совпадут с гармониками основного сигнала и гребнями фильтра, однако сам основной тон этого субтона и его нечетные гармоники через фильтр не пройдут, и тем самым исказится как вид сигнала, так и его энергия. Это же будет справедливо и для всех других субтонов, но там еще и станут выпадать их гармоники из числа тех, чьи частоты окажутся меньше F (а у них, как правило, максимальная энергия). Поэтому, полагая в первом приближении, что энергия сигналов равномерно распределена по их гармоникам, получим, что для субтонов, начиная со второго (первый субтон, как и первая гармоника совпадают с основной частотой фильтра F), их энергия на выходе гребенчатого фильтра будет ослабляться, по крайней мере, в k раз, где $k > 1$ – номер соответствующего субтона. В реальности же подавление субтонов будет еще более эффективным, поскольку обычно максимум энергии связан с основным тоном сигнала и его первыми гармониками.

Получается, что спектральным образом квазипериодического сигнала с основной частотой, меньшей F , при сканировании его гребенчатыми фильтрами с различными основными частотами будет структура в виде семейства параллельных и вытянутых вдоль оси времени всплесков энергии (хребтов). Причем энергии всплесков, соответствующих субтонам, будут локально максимальны, убывая обратно пропорционально их порядковым номерам, отсчитываемым от основного тона фильтра.

Аналогичная картина наблюдается и для сигналов с частотами основных тонов, совпадающими с гармониками основного тона фильтра, при наличии ограничений сигнала или фильтра по спектру.

Действительно, пусть суммарная полоса K -зубого фильтра превышает ширину спектра фильтруемого сигнала. В этом случае, если основной тон фильтруемого сигнала совпадает со второй гармоникой фильтра, на выход фильтра пройдет лишь целая часть $K/2$ гармоник исходного сигнала с соответствующей потерей энергии. В общем же случае, когда основной тон фильтруемого сигнала соответствует k -ой гармонике фильтра, на выходе фильтра останется лишь та часть сигнала, которая будет определяться его гармониками с номерами не более K/k . Таким образом, и в этом случае наблюдается эффект гиперболической по частоте фильтрации.

Следует заметить, что ограничение числом зубьев фильтра усиливает и эффект фильтрации субтонных сигналов в дополнение к рассмотренному выше.

Таким образом, можно считать, что K -зубый гребенчатый фильтр имеет частотную характеристику, убывающую примерно по гиперболическому закону, начиная со второй гармоники – для возрастающих частот, и начиная со второго субтона – для убывающих частот.

Важно подчеркнуть еще раз, что фильтрация сигналов, соответствующих гармоникам и субтонам, в этой схеме осуществляется несколько по-разному. Фильтрация субтоновых сигна-

лов в основном осуществляется подавлением их основного тона и “прореживанием” гармоник, в случае же сигналов, соответствующих гармоникам фильтра, осуществляется просто низкочастотная их фильтрация, соответствующая пропусканию их основного тона и первых гармоник и подавлению высокочастотных гармоник. Но главным в обоих случаях остается гиперболичность спектрального подавления сигнала.

3. ВТОРИЧНОЕ ГРЕБЕНЧАТОЕ ПРЕОБРАЗОВАНИЕ ПРИ ФОРМИРОВАНИИ «ТРЕКОВОЙ СЦЕНЫ»

Итак, в первой фазе происходит формирование двумерной спектрально-временной сцены со сравнительно слабым выделением треков голосов фильтрацией с гиперболическими по частоте частотными характеристиками.

Гребенчатая фильтрация согласованна с частотной структурой отображения исходного сигнала в сцену первичной фазы. Применяя ее ко всем частотам f_i в спектральной полосе сцены для всех моментов времени, можно добиться дополнительного подчеркивания всплесков энергии (хребтов), соответствующих основному тону сигнала.

Селективные свойства такого рода фильтрации на этом этапе еще достаточно скромны, так как в соответствии со сказанным выше ее частотные характеристики тоже будут гиперболическими по номерам субтонов и гармоник, что в итоге обеспечит квадратично гиперболические частотные характеристики по основному тону и его гармоникам и субтонам.

Для музыки, в частности, где энергия достаточно близких по частоте голосов может многократно отличаться, такая селективность оказывается действительно невысокой. Однако принципиально важным достижением этой фазы следует считать получение эффекта фильтрации, подчеркивающего именно требуемый трек основного тона сигнала.

Кроме того, наличие этого эффекта позволяет добиться и большей селективности, например, простым повторением фильтраций, приведших к этому результату, или усложнением схемы гребенчатых фильтров. Сказанное на примере все того же двухкомпонентного тестового сигнала иллюстрируется Рис. 4.

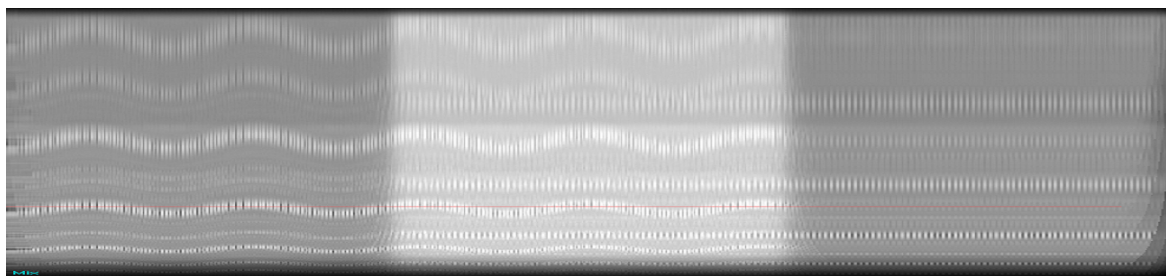


Рис. 4. Результат применения гребенчатой фильтрации к частотным срезам первичной сцены, представленной на Рис. 2. Видно, что она дает больше информации для выделения треков присутствующих в тесте полусинусоидальных “голосов”.

4. ЧАСТОТНО-ВРЕМЕННАЯ ГРЕБЕНЧАТАЯ ФИЛЬТРАЦИЯ

Сформированную в результате применения двух гребенчатых преобразований сцену можно рассматривать как двумерное изображение, сводя задачу выделения треков основного тона голосов к поиску и выделению неких образов, как-то себя ведущих (проявляющих) в этом изображении.

Например, мы можем считать, что спектрально-временные треки это – некие линии произвольной формы и топологии, как-то связанные с координатами t и f .

В этом случае учет природы этих линий может лишь накладывать на них какие-то упрощающие ограничения типа того, что каждая из них оказывается однозначной функцией горизонтальной координаты t и что частная производная такой линии по горизонтальной координате вне точек ее разрыва ограничена по величине.

Но все это не исключает наличия пересечений и соприкосновений различных треков-линий, а означает просто, что общий линейчатый рисунок сцены может быть описан множеством такого рода моделей.

С учетом сказанного мы можем продолжать использовать обозначения t и f и называть их временем и частотой, но не стеснять себя ограничениями сверх минимально необходимого. В частности, с учетом сказанного мы можем теперь рассматривать упомянутую выше процедуру вторичной фильтрации сцены, полученной на первом этапе, всего лишь как частный случай двумерной фильтрации изображения, когда используемое ядро свертки, реализующей фильтрацию, является одномерной функцией только лишь вертикальной координаты f .

В терминах двумерной сцены проблему выделения голосов можно интерпретировать как выделение и прослеживание треков голосов.

Несложно предположить, что использование двумерных ядер, потенциально предоставляющих существенно большие возможности для реализации двумерных фильтров, с точки зрения решаемой задачи может предоставить нам и дополнительные возможности выделения и отслеживания треков.

С этой целью гармонический фильтр, исходно рассмотренный в первичной фазе, во вторичной фазе был усилен для каждой из его частотных составляющих учетом коррелированности динамики рельефа сцены вдоль горизонтальной координаты.

Пусть

$$C_{corr1,k}(t, F, N) = \frac{\sum_{i=0}^M (D(t+i\delta, F) \cdot D(t+i\delta, kF))}{\sqrt{\sum_{i=0}^M D(t+i\delta, F)^2 \cdot \sum_{i=0}^M D(t+i\delta, kF)^2}}, \quad \text{где } M = \frac{N \cdot T}{\delta}$$

– коэффициент корреляции по времени фрагмента трека с основной частотой F с параллельным ему фрагментом трека с частотой kF , соответствующим либо k -ой гармонике, если $k > 1$ $k \in 1, \dots, N$, либо k -ому субтону, если $k = 1/K$, $K \in 1, \dots, N$. $M = \frac{N \cdot T}{\delta}$ – шаг дискретизации по времени, N – число периодов T частоты основного тона F фильтруемого сигнала. Тогда результат корреляционной фильтрации сцены $D(\cdot)$ в каждой ее точке можно определить выражением:

$$D(t, n, F, N) = \sum_{k=2}^n C_{corr1,k}(t, F, N) + \sum_{k=2}^n C_{corr1,1/k}(t, F, N),$$

а отфильтрованная таким образом сцена принимает вид, показанный на Рис. 5.

Фильтрация истинных треков – треков, соответствующих основным тонам голосов, получаемая таким образом, оказывается еще достаточно далекой от идеала, когда в сцене оставались бы лишь истинные треки. Однако избыточность сцены в виде ложных треков (или ложных гипотез о треках) на этом этапе оказывается уже не слишком большой. Анализ многочисленных тестов и примеров реальной музыки показывает, что, как правило, наряду с треками основных тонов присутствуют еще 1-2 явно различимых трека, соответствующих наиболее мощным гармоникам голосов. При этом достаточно часто оказывается, что одна из ложных гипотез оказывается более мощной, чем истинная.

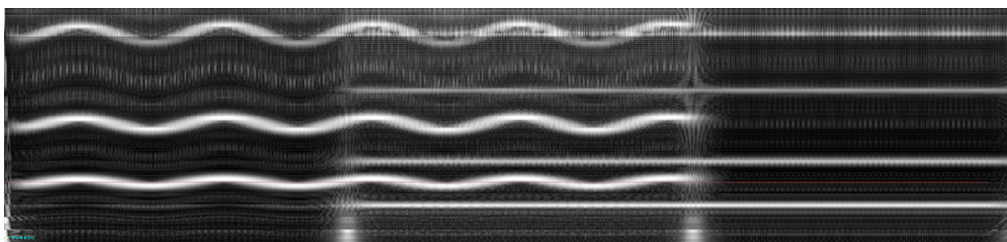


Рис. 5. Вид сцены, усиленной учетом взаимной коррелированности треков.

С решением проблем более качественного рафинирования “Трековых сцен”, получаемых на выходе корреляционной фазы “Струнного анализа” связаны его более поздние фазы.

5. ПРОСЛЕЖИВАНИЕ ГОЛОСОВ В ТРЕКОВОЙ СЦЕНЕ

Если треки не меняются по высоте, то выделить их довольно просто. Если же треки меняются по высоте, как это имеет место в случае чирпов и вибрато, то для их выделения представляется целесообразным использовать сравнительно изощренные процедуры прослеживания линий, для которых безразлично по сути, как вообще выглядит трек в сцене. Процедуры такого рода называются “слеящими развертками” и активно исследовались нами в прошлом [6].

С учетом имевшегося опыта разработки таких систем в данном случае было решено использовать алгоритм прослеживания с совмещенными апертурами и зеркально симметричными весовыми функциями в них, согласованными с характерными сечениями трека в непосредственной близости от него. Реализованная в таком виде процедура прослеживания треков голосов оказалась и простой и эффективной. Результаты ее применения показаны в серии иллюстраций сквозной обработки сигнала на Рис. 5. А ее подробное описание приведено в *Приложении В*.

6. РЕАЛЬНЫЙ СИГНАЛ

На основе предварительных результатов анализа тестовых квазигармонических сигналов с помощью исследуемого метода “Струнного анализа”, мы получили возможность перейти от анализа упрощенных тестовых сигналов к анализу многоголосых естественных сигналов, ради чего все это, собственно, и затевалось. В качестве источника такого рода сигналов была выбрана музыка, в которой можно без труда найти аналоги если не всех, то большинства, природных сигналов.

Далее, на Рис. 6а–8d, иллюстрируются некоторые из результатов, полученных после применения гребенчато-корреляционной процедуры (методов “Струнного анализа”) к различным образчикам музыкальных записей.

Пример фрагмента из арии Далилы в опере Сен-Санса “Самсон и Далила”, приведенный на Рис. 8а–8d, интересен тем, что по виду трека основного тона певицы он практически не отличается от рассмотренного выше фрагмента Арии Любаши из оперы Римского-Корсикова “Царская невеста”, но здесь – в отличие от того – ария исполняется в сопровождении оркестра, позволяя увидеть проявление неких важных свойств реальных смесей голосов.

Как следует из приведенных результатов, при представлении спектров сложных сигналов типа оркестровой музыки предлагаемый метод “Струнного анализа” не уступает классическому спектральному анализу. Но если иметь в виду не просто общую картину спектра, а разложение сложной смеси на составляющие ее квазигармонические “голоса”, то “Струнный

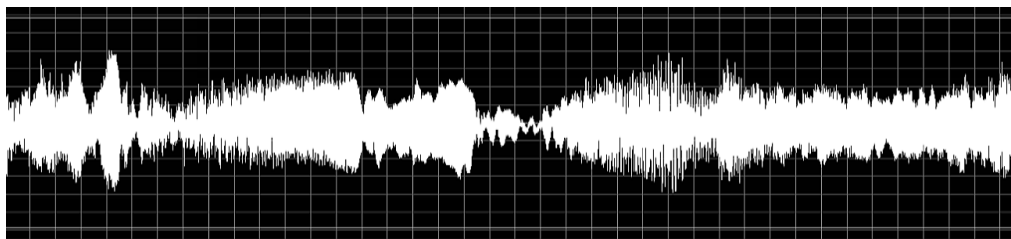


Рис. 6а. Четырехсекундный фрагмент Арии из Мессы #23 И.С.Баха, колоратурное сопрано. Спектр ограничен сверху частотой 5.5 кГц.

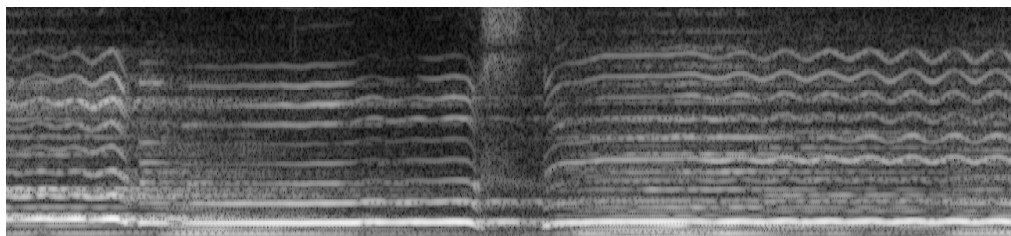


Рис. 6б. Фурье-спектр четырехсекундного фрагмента Арии из Мессы #23 И.С.Баха, колоратурное сопрано со струнным оркестром. FFT 512, Спектр ограничен сверху частотой 5.5 кГц.

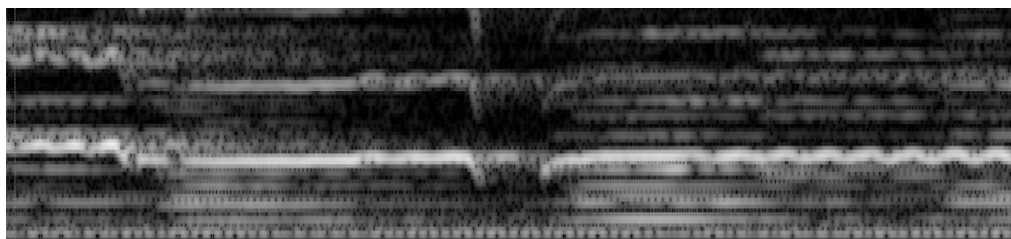


Рис. 6с. Фурье-спектр четырехсекундного фрагмента Арии из Мессы #23 И.С.Баха, колоратурное сопрано со струнным оркестром. FFT 512. Спектр ограничен сверху частотой 1.5 кГц, для более детального понимания происходящего в нижней его части.

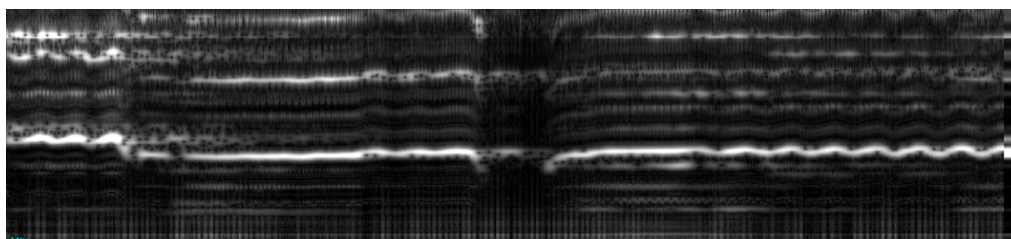


Рис. 6д. Трековая сцена того же четырехсекундного фрагмента Арии из Мессы #23 И.С.Баха, колоратурное сопрано. Спектр ограничен сверху частотой 1.5 кГц.

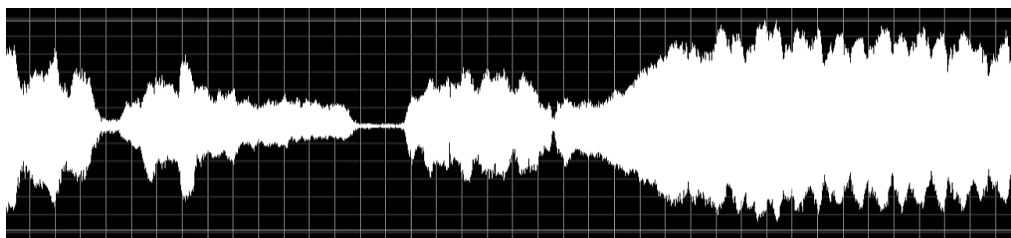


Рис. 7а. Четырехсекундный фрагмент Арии Любаши из оперы Римского-Корсакова "Царская невеста"

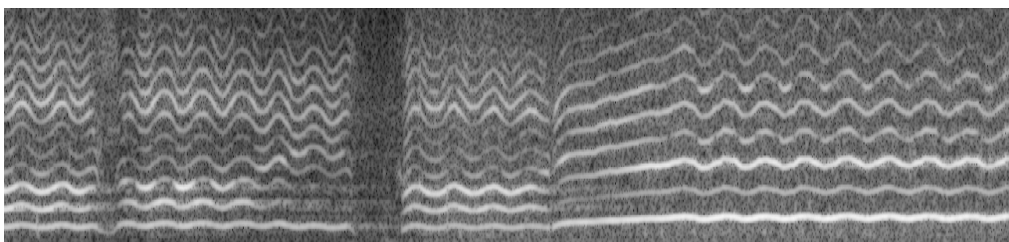


Рис. 7b. Фурье-спектр четырехсекундного фрагмента Арии Любаши из оперы Римского-Корсакова “Царская невеста”. Колоратурное сопрано. FFT 512, спектр ограничен сверху частотой 5.5 кгц.

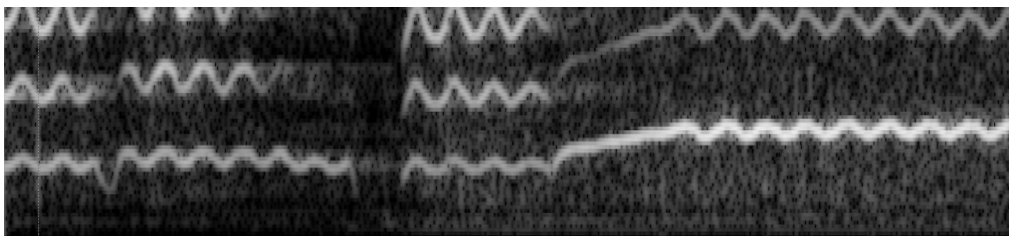


Рис. 7с. Фурье-спектр четырехсекундного фрагмента Арии Любаши из оперы Римского-Корсакова “Царская невеста”. Колоратурное сопрано. FFT 512, спектр ограничен сверху частотой 1.5 кгц, для более детального понимания происходящего в нижней его части.

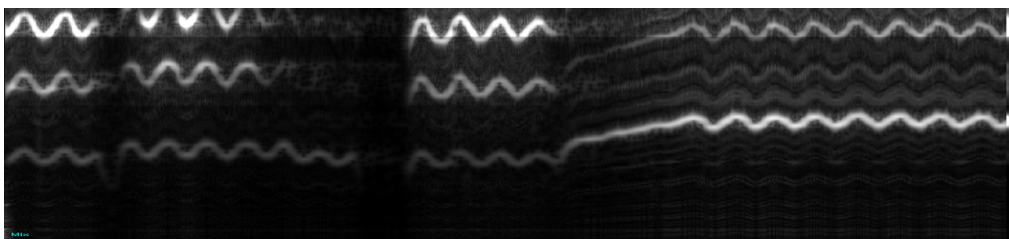


Рис. 7d. Трековая сцена того же четырехсекундного фрагмента Арии Любаши из оперы Римского-Корсакова “Царская невеста”.

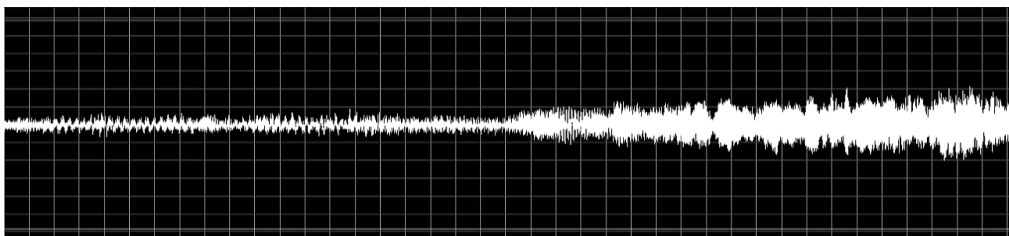


Рис. 8a. Двухсекундный фрагмент арии Далилы из оперы Сен-Санса “Самсон и Далила”, колоратурное сопрано, струнный оркестр.

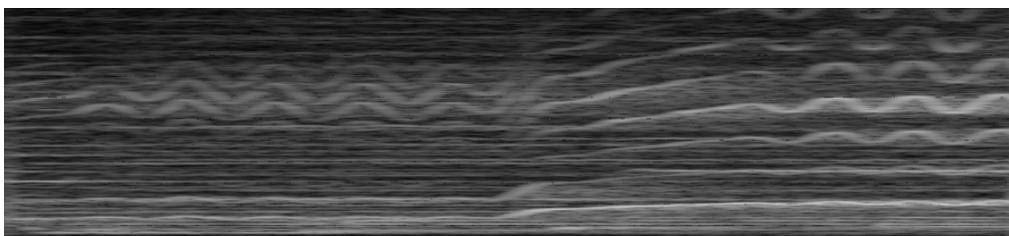


Рис. 8b. Фурье-спектр двухсекундного фрагмента арии Далилы из оперы Сен-Санса “Самсон и Далила”, колоратурное сопрано, струнный оркестр. FFT 512. Спектр ограничен сверху частотой 5.5 кгц.

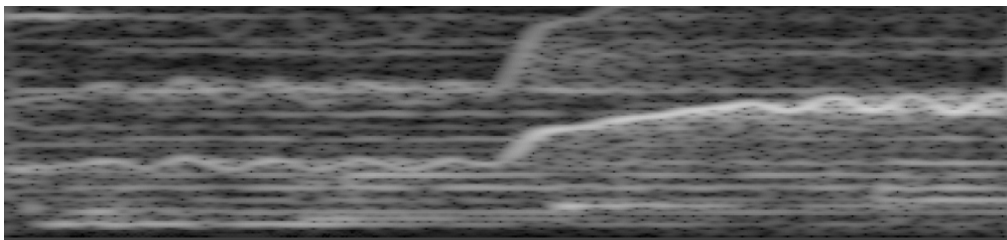


Рис. 8с. Фурье-спектр двухсекундного фрагмента арии Далилы из оперы Сен-Санса “Самсон и Далила”, колоратурное сопрано, струнный оркестр. Спектр ограничен сверху частотой 1.5 кГц. для более детального понимания происходящего в нижней его части. Видны совпадения голосов оркестра и их гармоник с голосом певицы.

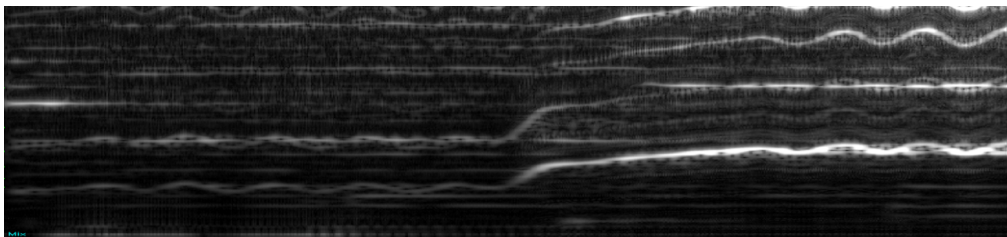


Рис. 8д. Трековая сцена двухсекундного фрагмента арии Далилы из оперы Сен-Санса “Самсон и Далила”, колоратурное сопрано, струнный оркестр. Спектр ограничен сверху частотой 3 кГц. для более детального понимания происходящего в нижней его части. Видны совпадения голосов оркестра и их гармоник с голосом певицы.

анализ” должен иметь преимущество в силу того, что в нем параллельно осуществляется выделение всех гармоник голосов вместо отдельных гармонических компонентов при обычном Фурье-анализе. И проблема заключается лишь в эффективной реализации этого потенциального преимущества.

В прикрепленном файле приведена запись двухсекундного фрагмента арии Далилы из оперы Сен-Санса “Самсон и Далила”, колоратурное сопрано, струнный оркестр.

7. НЕЛИНЕЙНОЕ ВРЕМЯ: КОНВЕЙЕР ПРЕОБРАЗОВАНИЙ

Принципиальная важность частотно непостоянных голосов и разнообразие динамических преобразований, которые потенциально необходимо с ними производить в связи с решаемыми проблемами выделения чистых голосов из сложных их смесей привели к решению перейти к новому, более удобному их представлению, когда анализируемый трек голоса превращается в частотно-постоянный с априорно заданной частотой основного тона независимо от того, каким он на самом деле является.

Это оказывается возможным благодаря тому, что путем подходящего преобразования времени любой трек в сцене можно превратить в любой другой, лишь бы при этом сохранялась их монотонность во времени. В частности, если в сцене со временем t и частотой трека f_0 произвести замену времени на kt , то частота трека в сцене с новым временем превратится в f_0/k . Это – фундаментальное соотношение. Принципиально важно при этом, что это преобразование обратимо, что позволяет восстанавливать в исходном времени результаты манипуляций с преобразуемым так сигналом.

Кроме того, такое преобразование с помощью “конвейера резинового времени” может быть итеративным, позволяя на каждой из таких итераций повышать точность прослеживания треков. В наших экспериментах, однако, оказалось достаточно всего лишь одной итерации.

8. ПРИМЕР СКВОЗНОГО ПРИМЕНЕНИЯ ТЕХНОЛОГИИ СТРУННОГО АНАЛИЗА

Ниже (Рис. 9а–9л) иллюстрируются результаты применения последовательных этапов обработки сложных сигналов в технологии Струнного анализа к фрагменту арии Далилы, рассмотренному выше на Рис. 8а–8д, включая и выделение голоса певицы.

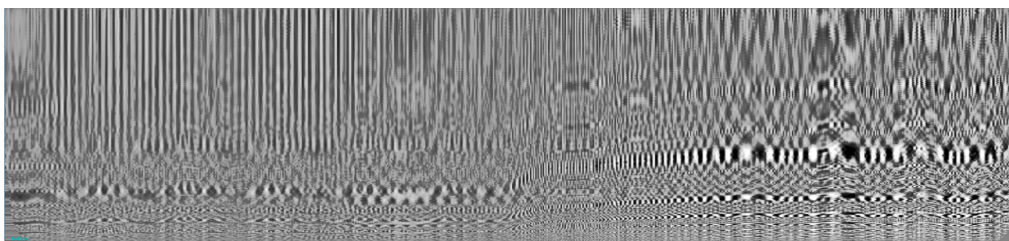


Рис. 9а. Результат гребенчатой фильтрации фрагмента арии Далилы по времени.

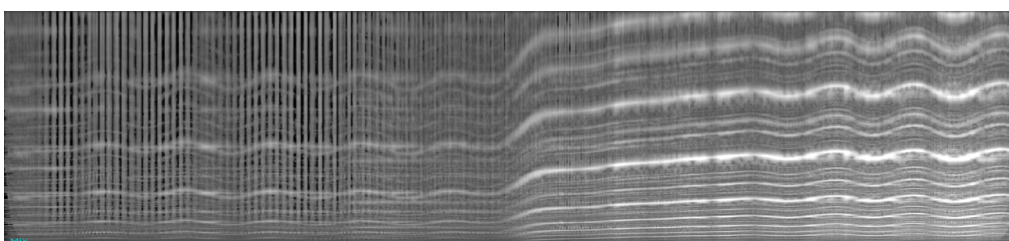


Рис. 9б. Результат гребенчатой фильтрации фрагмента арии Далилы по частоте.

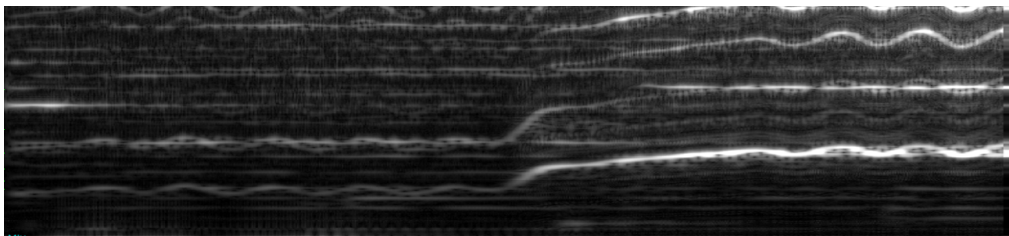


Рис. 9с. Результат гребенчатой фильтрации фрагмента арии Далилы по частоте и времени.

Прослушивание восстановленного голоса певицы в дополнение к проблемам точного автоматического прослеживания его трека выявляет на акустическом уровне некие искажения, обусловленные несовершенством процедуры экстракции голоса. Сильнее всего это проявляется в некоем искажении тембра за счет усиления высокочастотных составляющих его спектра и неких призвуках, обусловленных, по-видимому, добавлением к голосу консонантных гармоник струнных инструментов, сопровождающих пение и не воспринимаемых селектором голоса как чужие для него. И тембр, таким образом, становится центральным объектом дальнейших исследований возможностей повышения качества выделения отдельных квазигармонических голосов из их смесей.

В прикрепленном файле приведена запись восстановленного сигнала голоса певицы из двухсекундного фрагмента арии Далилы из оперы Сен-Санса “Самсон и Далила”.

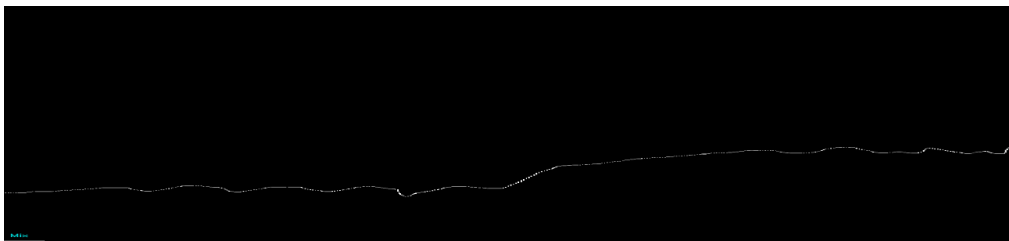


Рис. 9d. Результат автоматического прослеживания трека голоса певицы с установкой начала трека вручную.

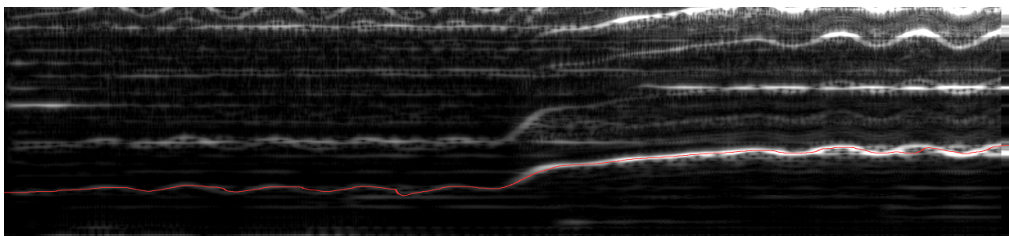


Рис. 9е. Результат наложения выделенного трека на истинный трек голоса певицы в частотно-временной сцене. Видна пара срывов прослеживания на консонантные голоса оркестра.

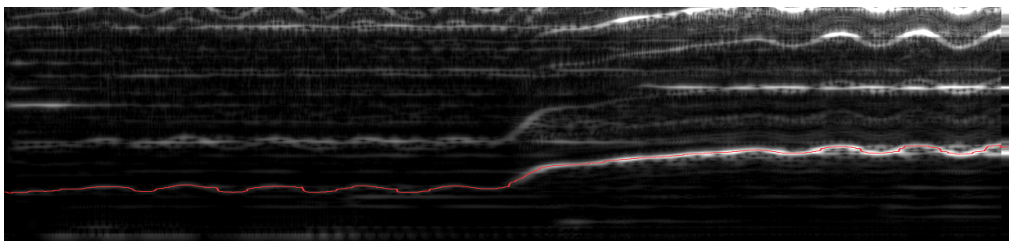


Рис. 9f. Результат автоматической коррекции трека “подтягиванием” к спектральному максимуму после ручной коррекции срывов.



Рис. 9g. Результат заглаживания скорректированного трека певицы.

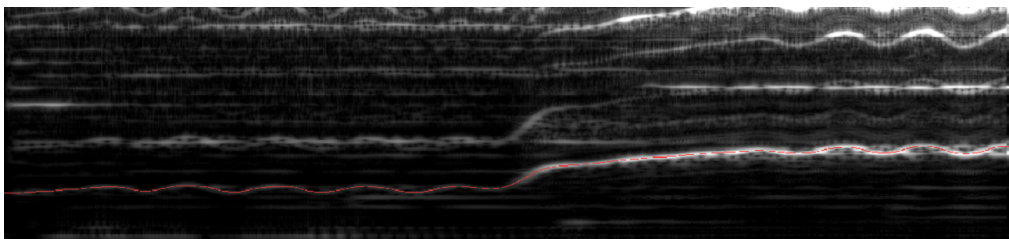


Рис. 9h. Результат наложения сглаженного и скорректированного трека на сцену с голосом певицы.

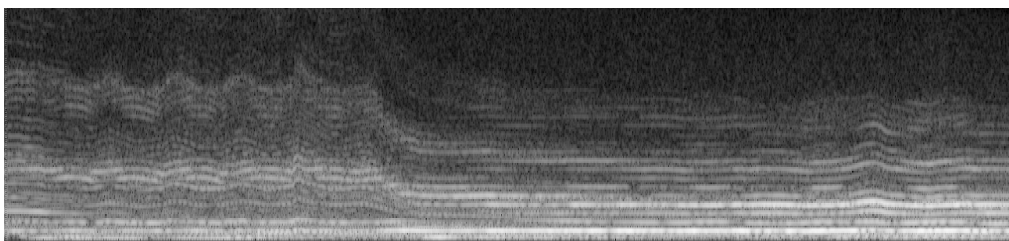


Рис. 9i. Спектрограмма исследуемого фрагмента записи голоса певицы после преобразования “резинового времени” с отображением полученного трека основного тона певицы на частоту 50 Гц. Наблюдаемые глиссандо и вибрато здесь уже принадлежат голосам инструментов оркестра. Они в противофазе треку голоса певицы.



Рис. 9j. Результат выделения голоса певицы в конвейере резинового времени фильтрацией гребенчатым селектором с основной частотой 50 Гц.

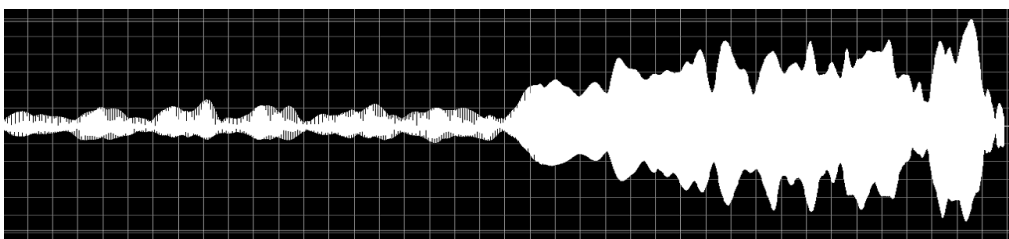


Рис. 9k. Результат восстановления сигнала голоса певицы после его фильтрации в конвейере резинового времени.

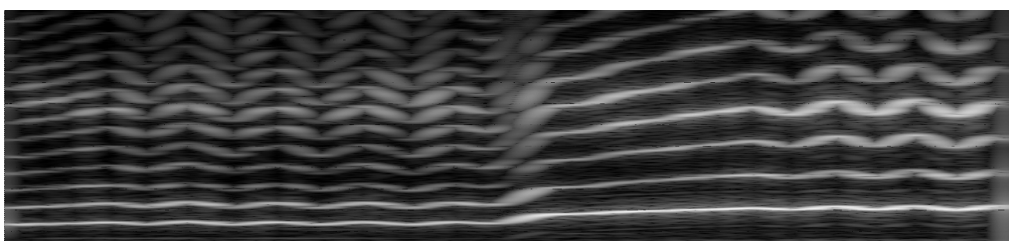


Рис. 9l. Спектр восстановленного сигнала голоса певицы после его фильтрации в конвейере резинового времени.

9. ЗАКЛЮЧЕНИЕ

В данной работе в качестве модельной рассматривалась задача разделения сложного сигнала – типа записи оркестровой музыки с человеческими голосами – на составляющие этот сигнал “голоса” отдельных инструментов и певцов. Класс такого рода задач весьма обширен и представляет большой интерес, в частности, в связи с проблемами диагностики состояния различного рода сложных технических систем.

Исследовавшийся нами подход основывался на допущении, что “голоса” в такого рода смесях, возникающие и прекращающиеся в какие-то моменты времени, являются преимущественно квазигармоническими сигналами с непостоянными во времени частотами и интенсивностями. При этом энергии отдельных голосов могут быть и как достаточно велики по сравнению с суммарной энергией остальных голосов, так и с существенно меньшей общей воспринимаемой энергии сигнала. Это в сочетании с тем что отдельные голоса могут сильно различаться и спектральными областями, с характерным для этих голосов расположением основной энергии, приводит к тому, что и спектр суммарного многоголосого сигнала и его динамический диапазон могут быть существенно более широкими, чем спектры и диапазоны составляющих смесь голосов. Несложно видеть, что в музыкальных сигналах можно обнаружить примеры, иллюстрирующие все это. При этом голоса с порождаемыми ими шумовыми и импульсными сигналами мы пока оставили за рамками рассмотрения. В основу решения выделения голосов нами было положено использование адаптивных гребенчатых фильтров, которые можно считать согласованными с природой квазигармонических сигналов. Эта часть работы была достаточно продвинута и показала целесообразность такого рода подхода и важность его самостоятельного развития, включая и теоретическое обоснование.

Следующее упрощающее допущение было связано с проблемами разделения выделяемых прослеживанием треков голосов в точках их пересечения или на интервалах совпадения. Достаточно быстро стало ясно, что в этой фазе принципиальное значение приобретают “тембр” и знания еще более высокого порядка, связанные с общим функционированием источника смеси голосов. Первый из этих уровней – тембровый – нам пока удалось проработать лишь на самом общем концептуальном уровне, о чем кое-что говорится в Приложениях. А к “семантическому” уровню анализа мы даже и не приступили, ограничившись лишь общим пониманием того, что уже существуют достаточно продвинутые подходы к решению такого рода задач на основании использования разного рода идей обучения распознаванию образов, в частности, и в связи с решением задач искусственного интеллекта.

Выбранный метод выделения голосов, условно называемый нами “Струнным анализом”, исследовался с помощью двух разработанных для этого программ “Golos” и “String”. Первая из них преобразует сложный сигнал в двумерную. “Сцену”, особая фильтрация которой и позволяет выделять и отслеживать треки основных частот голосов, решая первичную задачу их селекции. А вторая является специальным редактором, позволяющим вручную компенсировать еще имеющиеся недоработки в автоматическом выделении голосов, позволяя тем самым выявлять и уточнять проблемы, приближающие нас к полностью автоматическому анализу смесей разложением их на квазигармонические голоса.

Практически важным результатом проведенного исследования является то, что выделяемые голоса обладают при прослушивании достаточно высоким качеством и близостью к тому, что слышит ухо в смеси. Обнаруживаемые при этом искажения являются тембровыми по сути и поэтому на следующем этапе должны быть объектом основного внимания.

10. БЛАГОДАРНОСТИ

Авторы признательны В.Н.Сорокину за внимательное прочтение работы, сделанные замечания

А. ДЕТАЛИЗАЦИЯ ОПИСАНИЙ ДОПОЛНИТЕЛЬНЫХ ВАРИАНТОВ ГРЕБЕНЧАТЫХ ФИЛЬТРОВ

Простейшим (базовым) вариантом гребенчатого фильтра является фильтр, блок-схема которого приведена на Рис. 10.

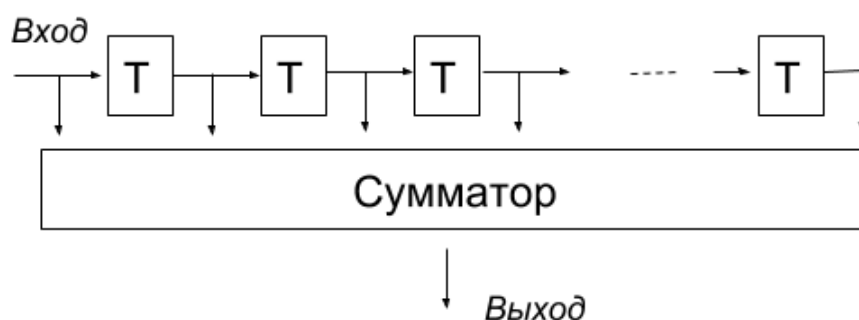


Рис. 10. Блок-схема базового гребенчатого фильтра на n -элементной линии задержки. Здесь T – элементы, обеспечивающие задержку входного сигнала на время T .

Такие фильтры реализуются с помощью n -элементных линий задержки и имеют частотные характеристики типа приведенной на Рис. 11, где показана частотная характеристика для восьми-элементного фильтра.

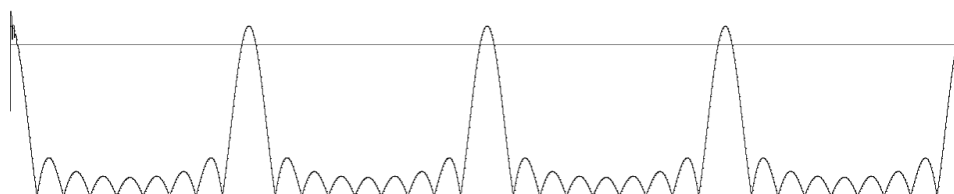


Рис. 11. Фрагмент частотной характеристики базового гребенчатого фильтра, соответствующего 8-элементной линии задержки. Видны интенсивные боковые лепестки. Максимальные всплески селективности – на частотах гармоник основного тона фильтра F , равного $1/T$, где T – время задержки сигнала одним элементом.

Недостатком базового варианта n -элементного гребенчатого фильтра для сигнала является наличие паразитных лепестков – дополнительных полос пропускания, сигнал на выходе которых еще достаточно велик – коэффициент пропускания d в них порядка и менее 0.066 от величины основного. Поэтому представлялось интересным найти такие модификации базового фильтра, которые обладали бы этим недостатком в существенно меньшей мере.

С самого начала было понятно, что последовательное включение k одинаковых базовых фильтров позволит получить суммарный коэффициент пропускания порядка d^k . В частности, самый сильный побочный лепесток окажется на уровне примерно $2.2 \cdot 10^{-2}$ при $k = 2$, $3.4 \cdot 10^{-3}$ при $k = 3$, $5 \cdot 10^{-4}$ при $k = 4$.

Ширина главного лепестка, взятая между двумя ближайшими к нему нулями, составит $2F/n$, где F и n – соответственно собственная частота (частота основного тона) и число элементов задержки. Если же полосу пропускания определять на уровне 0.7 от максимума, то для одного базового фильтра полоса пропускания составит примерно $0.1F$, для последовательности из двух базовых – $0.07F$, для трех – $0.06F$, для четырех – $0.05F$. При этом длина суммарного фильтра определится как $1n, 2n, 3n, 4n$. А полная длительность его переходного процесса составит – $2Tn, 4Tn, 6Tn, 8Tn$.

Частотные характеристики рассмотренных четырех вариантов 8-элементного базового гребенчатого фильтра и его усилений повторениями приведены на Рис. 12а–12д.

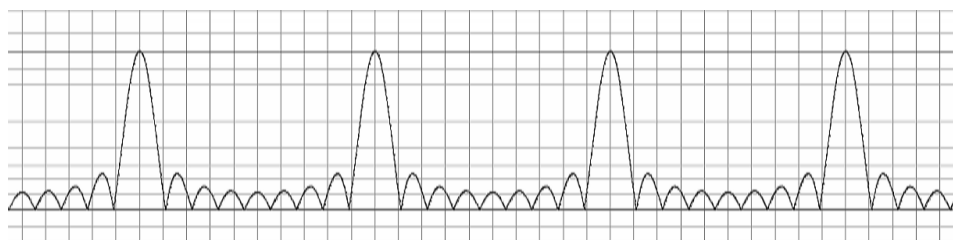


Рис. 12а. Фрагмент частотной характеристики одиночного 8-элементного гребенчатого фильтра.

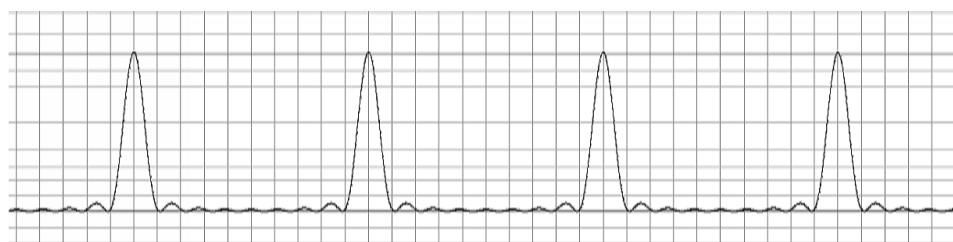


Рис. 12б. Фрагмент частотной характеристики сдвоенного 8-элементного гребенчатого фильтра.

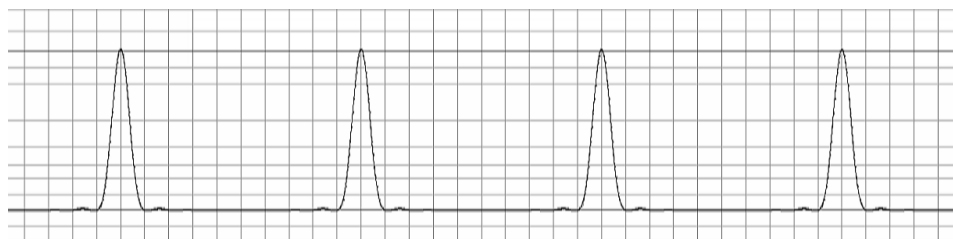


Рис. 12с. Фрагмент частотной характеристики строенного 8-элементного гребенчатого фильтра.

Видно, что каскадное использование таких фильтров позволяет довольно быстро избавляться от побочных лепестков, правда, ценой кратного увеличения длительности переходных процессов. В связи с этим представлялось интересным посмотреть, нельзя ли путем неких модификаций базового фильтра улучшить качество его фильтрации, не выходя, конечно, за пределы принципиальных ограничений, устанавливаемых взаимосвязью спектральных и временных описаний сигналов [7]. Для этого были разработаны две модификации базового фильтра, рассматриваемые далее.

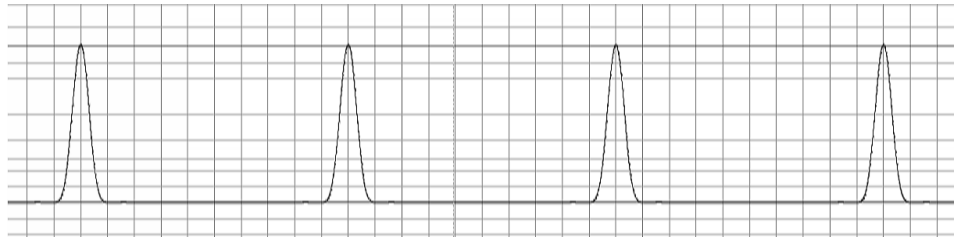


Рис. 12d. Фрагмент частотной характеристики счетверенного 8-элементного гребенчатого фильтра.

А.1. Первая модификация базового фильтра

В базовой модели все суммируемые значения фильтруемого сигнала брались с одинаковым весом $1/(n+1)$, где n – число элементов задержки сигнала на период частоты основного тона $T = 1/F$. В первой рассматриваемой нами модификации гребенчатого фильтра веса суммируемых значений запаздывающего сигнала определялись “Треугольником Паскаля”, начало которого демонстрируется ниже:

$$\begin{array}{ccccccc} & & & & & & 1 \\ & & & & & & 1 & 1 \\ & & & & & 1 & 2 & 1 \\ & & & & 1 & 3 & 3 & 1 \\ & & 1 & 4 & 6 & 4 & 1 \\ & 1 & 5 & 10 & 10 & 5 & 1 \end{array}$$

Алгоритм определения членов ряда этого треугольника ясно виден из приведенного рисунка: i -ый член ряда равен сумме стоящего над ним и ему предшествующего. При этом пустым позициям приписывается значение 0. Как видно, k -ая строка Треугольника Паскаля содержит k элементов. Таким образом, на основе этого треугольника, используемого в качестве весового вектора, и базового гребенчатого фильтра можно создавать гребенчатые фильтры Паскаля любой требуемой длины.

Пусть $P_{i,k}$ – i -ый элемент k -ой строки Треугольника Паскаля. Тогда гребенчатый фильтр Паскаля длины $n = k - 1$ можно описать следующим образом:

$$S_{out}(t) = \frac{\sum_{i=0}^{k-1} P_{i,k} S_{in}(t - iT)}{\sum_{i=0}^{k-1} P_{i,k}},$$

где S_{out} – выходной сигнал фильтра, $S_{in}(t - iT)$ – значение входного сигнала в момент, запаздывающий относительно текущего на i периодов $T = 1/F$ частоты основного тона F .

Частотные характеристики такого рода фильтров для различных n приведены на Рис. 13а–13д.

Важно отметить, что при той же длине, что и у базового фильтра, фильтр Паскаля имеет заметно большую (примерно в 2 раза) полосу пропускания. Поэтому при желании получить с его помощью фильтр с такой же полосой пропускания, что и у базового, число его элементов должно соответственно быть больше в те же 2 раза. За счет этого несколько возрастает и время переходного процесса. Но зато у фильтра Паскаля полностью отсутствуют побочные лепестки.

Если полосу пропускания фильтра Паскаля определять на уровне 0.7 от максимума, то для одиночного фильтра полоса пропускания составит примерно $0.19F$, для двух базовых – $0.14F$,

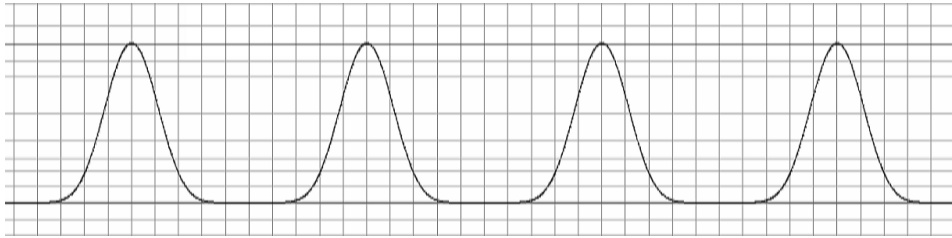


Рис. 13а. Одиночный 8-миэлементный гребенчатый фильтр Паскаля.

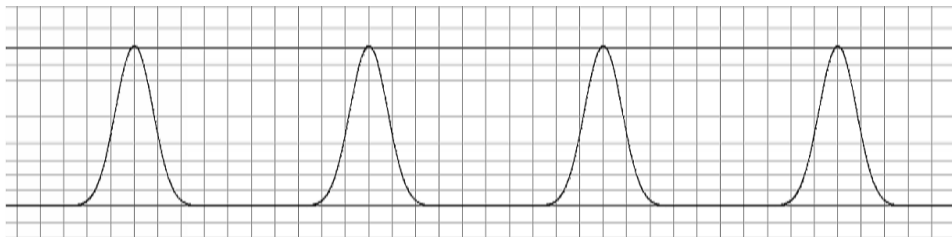


Рис. 13б. Сдвоенный 8-миэлементный гребенчатый фильтр Паскаля.



Рис. 13в. Строенный 8-миэлементный гребенчатый фильтр Паскаля.

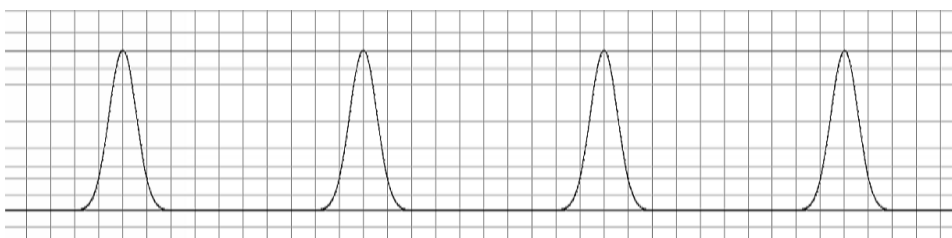


Рис. 13д. Счетверенный 8-миэлементный гребенчатый фильтр Паскаля.

для трех – $0.11F$, для четырех – $0.096F$. При этом длина суммарного фильтра определится как $1n, 2n, 3n, 4n$. А полная длительность его переходного процесса составит примерно $2Tn, 4Tn, 6Tn, 8Tn$.

А.2. Вторая модификация базового фильтра

При заданном числе элементов базовый вариант фильтра с одинаковыми весами равнопаздывающих отсчетов сигнала имеет более узкую полосу, чем фильтр с паскалевским взвешиванием, и достаточно заметные побочные лепестки, а фильтр Паскаля имеет более широкую полосу при практически полном отсутствии побочных лепестков. Поэтому возникает естественное желание создать такой вариант гребенчатого фильтра, который сочетал бы в себе достоинства как базового, так и паскалевского вариантов.

Оказалось, что это желание почти идеально может быть исполнено в следующем варианте фильтра, который мы называем “компенсационным”. Основная идея компенсационного фильтра связана со следующим.

В частотной характеристике базового фильтра из n элементов частоты нулей определяются соотношением: $F_k = Fk/n$, $k \in 1, \dots, n-1$. А побочные лепестки размещаются в интервалах между этими точками. Если изменить n , то и нули и побочные лепестки сместятся в новые положения.

Поэтому, если последовательно включить два базовых фильтра с некрратными длинами n_1 и n_2 , то их нули не совпадут и, следовательно, нули одного придутся на ненулевые значения другого и наоборот, что с неизбежностью приведет к подавлению боковых лепестков.

Очевидно, что можно в связи с этим решать оптимизационную задачу по наилучшему соотношению длин фильтров n_1 и n_2 . Мы, однако, пока не видим в этом необходимости, тем более что в такой схеме заведомо можно использовать больше двух фильтров, и вместо этого исследовали несколько вариантов такого рода цепочек фильтров, синтезируемых по принципу $n_i = n_1 - i$, $i \in 0, 1, 2, 3$.

Результаты таких экспериментов для $n_1 = 8$ и числе цепочек, меняющемся от 2 до 4, приведены парами на Рис. 14а–16б.

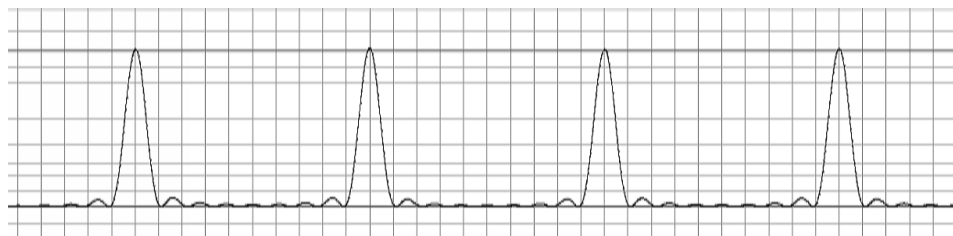


Рис. 14а. Общий вид частотной характеристики фрагмента двухэлементного компенсационного фильтра с 8 и 7 элементами задержки.

Видно, что, уже начиная с трехэлементного компенсационного фильтра “8–7–6”, побочными лепестками практически можно пренебречь.

А.3. Режекторные фильтры

Наряду с фильтрами, селективирующими интересующие сигналы, в анализе возникает нужда и в альтернативном виде фильтров – режекторах, являющихся своеобразным “отрицанием” такого рода селекторов, смысл которых заключается в подавлении определенных видов сигналов. Вообще говоря, всякий фильтр является одновременно и селектором и режектором,

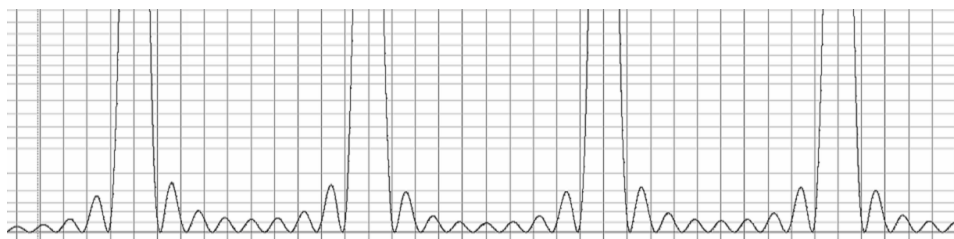


Рис. 14b. Общий вид нижней зоны частотной характеристики фрагмента двухэлементного компенсационного фильтра с 8 и 7 элементами задержки. Относительная величина максимального побочного лепестка составляет 0.047 от основного лепестка.



Рис. 15a. Общий вид частотной характеристики фрагмента трехэлементного компенсационного фильтра с 8, 7 и 6 элементами задержки.

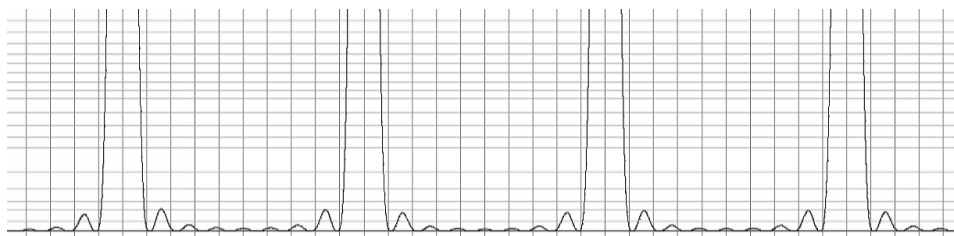


Рис. 15b. Общий вид нижней зоны частотной характеристики фрагмента трехэлементного компенсационного фильтра с 8, 7 и 6 элементами задержки. Относительная величина максимального побочного лепестка составляет 0.0084 от основного лепестка.

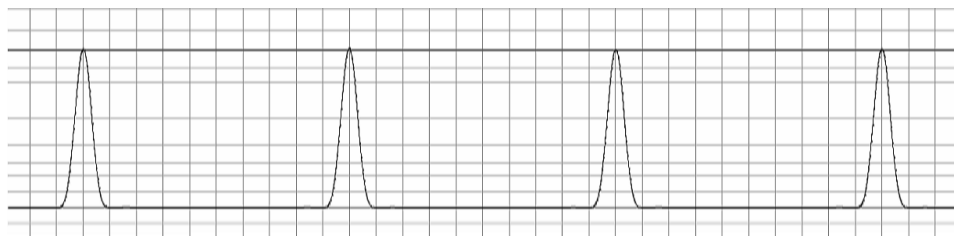


Рис. 16a. Общий вид частотной характеристики фрагмента четырехэлементного компенсационного фильтра с 8, 7, 6 и 5 элементами задержки.

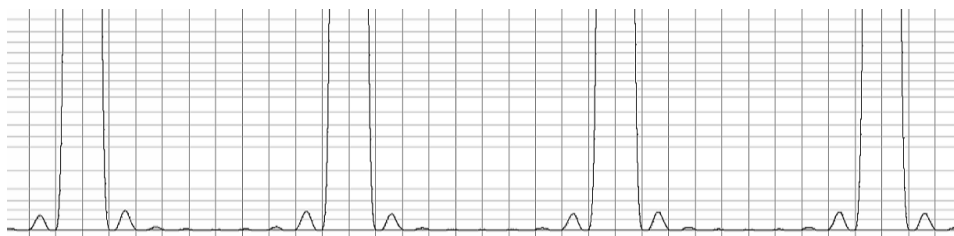


Рис. 16b. Общий вид нижней зоны частотной характеристики фрагмента четырехэлементного компенсационного фильтра с 8, 7, 6 и 5 элементами задержки. Относительная величина максимального побочного лепестка составляет 0.00094 от основного лепестка.

но в обычной ситуации режектируется “отрицание” селектируемого сигнала, а не сам сигнал. Впрочем, все это – вид некой технологической игры, в которой “сигнал” и его “отрицание”, хотя и являются сущностями одной природы, но обычно неявно предполагается, что про “сигнал” априорно известно много больше, чем про его “отрицание”, обладающее, как правило, существенно большей сложностью. В частности, и в нашем случае под сигналом рассматривается нечто квазипериодическое, а его дополнением может быть сложная смесь других квазипериодичностей, шумовых и импульсных процессов. И в этом случае вполне возможны ситуации, когда выделение и подавление сильного сигнала может существенно упростить последующий анализ остающейся смеси. Поэтому и в нашем исследовании, ориентированном в отличие от обычных методов на анализ квазигармонических сигналов, нам пришлось подумать и о режекторах квазигармонических сигналов тоже.

В качестве режекторных исследовались варианты гребенчатых фильтров, полученные на основе базового селектирующего фильтра и двух его модификаций, описанных выше, по следующей схеме, представленной на Рис. 17.

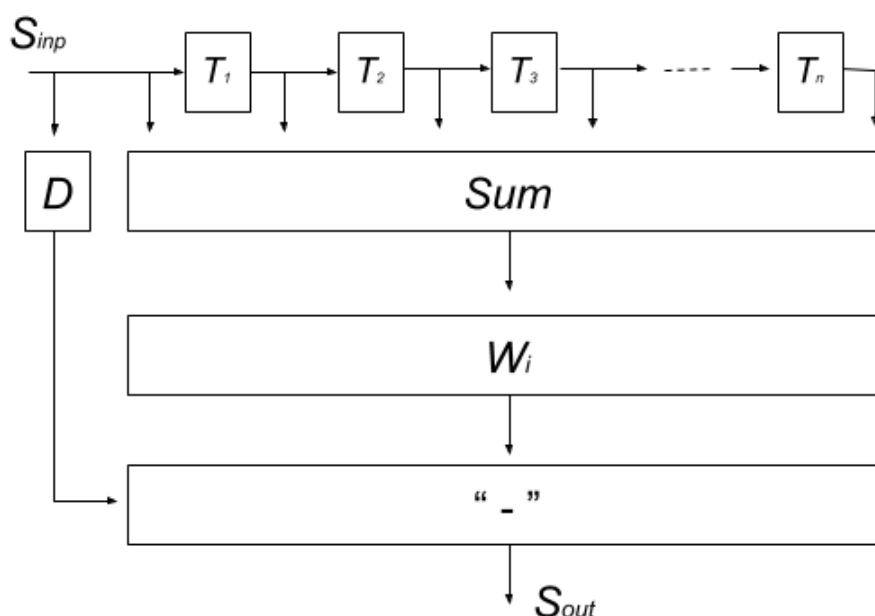


Рис. 17. Схема гребенчатого режектора.

В соответствии с этой схемой S_{inp} – анализируемый сигнал, S_{out} – выходной сигнал гребенчатого режектирующего фильтра, n – общая длина селектирующего фильтра в периодах T основной частоты F фильтруемого сигнала, W_i – веса, D – задержка входного сигнала на целую часть $n/2$, что важно в нашем случае только лишь для фильтров Паскаля.

Полученные результаты представлены серией иллюстраций для частотных характеристик таких фильтров на Рис. 18а–20с. Они дают достаточное представление о возможностях получаемых так гребенчатых режекторов и позволяют думать об их адекватном использовании и развитии с учетом возникающих пожеланий.

Как видно, начиная с тройного компенсационного фильтра с формулой “8+7+6”, этот тип фильтра оказывается наилучшим среди рассмотренных вариантов режекторов.

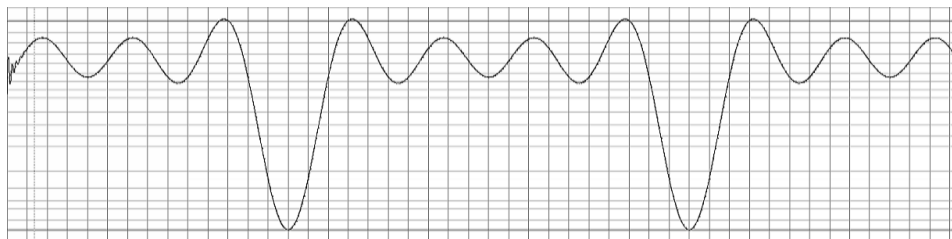


Рис. 18а. Общий вид фрагмента частотной характеристики режектора на основе одиночного восьми-элементного базового фильтра. Диапазон пульсаций в полосе пропускания составляет 3 Дб. Полоса подавления по уровню 20 Дб составляет $0.06F$.

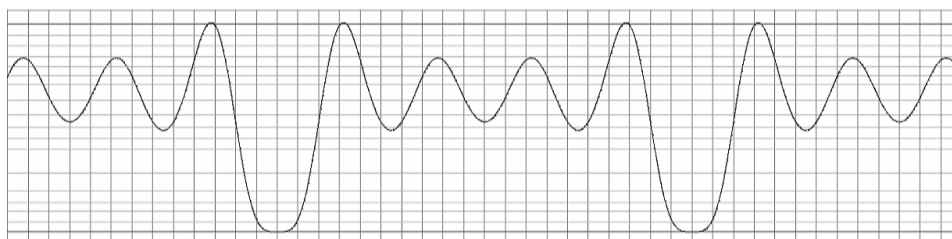


Рис. 18b. Общий вид фрагмента частотной характеристики режектора на основе двоянного восьми-элементного базового фильтра. Диапазон пульсаций в полосе пропускания составляет 6 Дб. Полоса подавления по уровню 20 Дб составляет $0.12F$, $0.07F$ по уровню 40 Дб. и $0.04F$ по уровню 60 Дб.

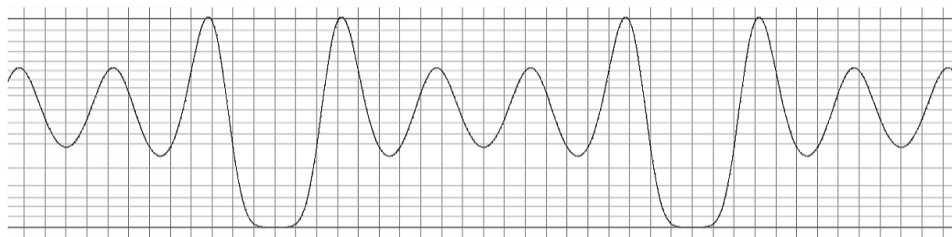


Рис. 18с. Общий вид фрагмента частотной характеристики режектора на основе строенного восьми-элементного базового фильтра. Диапазон пульсаций в полосе пропускания составляет 9 Дб. Полоса подавления по уровню 20 Дб составляет $0.15F$, по уровню 40 Дб. и $0.06F$ по уровню 60 Дб.

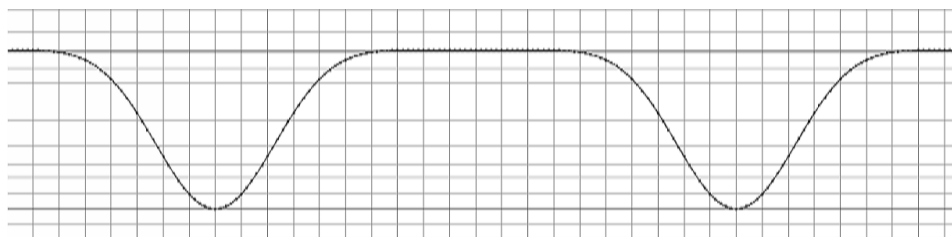


Рис. 19а. Общий вид фрагмента частотной характеристики режектора на основе одиночного восьми-элементного фильтра Паскаля. Пульсаций в полосе пропускания нет, ширина полосы пропускания по уровню 3 Дб $0.83F$. Полоса подавления по уровню 3Дб составляет $0.17F$, $0.051F$ по уровню 20 Дб, $0.016F$ по уровню 40 Дб. и $0.004F$ по уровню 60 Дб.

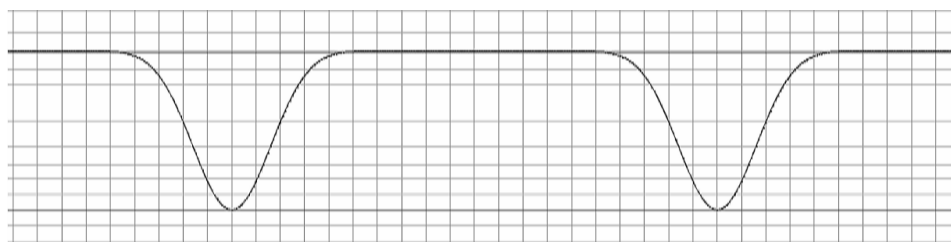


Рис. 19b. Общий вид фрагмента частотной характеристики режектора на основе сдвоенного восьми-элементного фильтра Паскаля. Пульсаций в полосе пропускания нет, ширина полосы пропускания по уровню 3 Дб $0.88F$. Полоса подавления по уровню 3Дб составляет $0.12F$, $0.036F$ по уровню 20 Дб, $0.011F$ по уровню 40 Дб. и порядка $0.0005F$ по уровню 60 Дб.

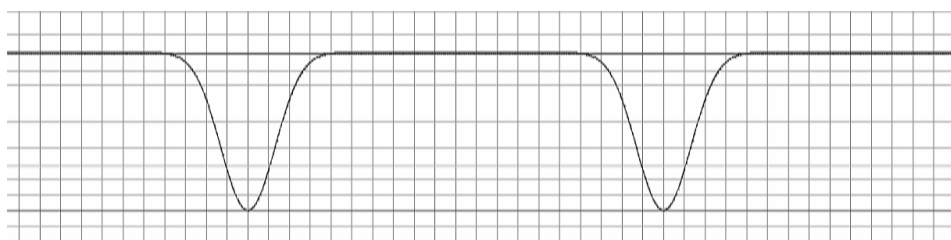


Рис. 19с. Общий вид фрагмента частотной характеристики режектора на основе строенного восьми-элементного фильтра Паскаля. Пульсаций в полосе пропускания нет, ширина полосы пропускания по уровню 3 Дб $0.90F$. Полоса подавления по уровню 3Дб составляет $0.10F$, $0.03F$ по уровню 20 Дб, $0.0085F$ по уровню 40 Дб. и порядка $0.00025F$ по уровню 60 Дб.

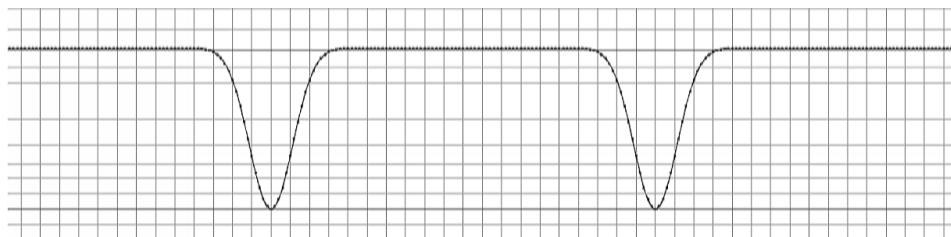


Рис. 19d. Общий вид фрагмента частотной характеристики режектора на основе счетверенного восьми-элементного фильтра Паскаля. Пульсаций в полосе пропускания нет, ширина полосы пропускания по уровню 3 Дб $0.914F$. Полоса подавления по уровню 3Дб составляет $0.086F$, $0.026F$ по уровню 20 Дб, $0.0080F$ по уровню 40 Дб. и порядка $0.00015F$ по уровню 60 Дб.

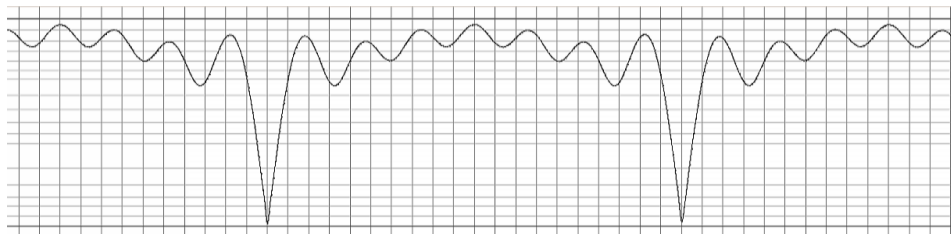


Рис. 20a. Общий вид фрагмента частотной характеристики режектора на основе одиночного компенсационного фильтра с элементами "8–7". Диапазон пульсаций 3 Дб в полосе пропускания $0.91F$. Полоса подавления по уровню 20 Дб составляет $0.015F$, $0.002F$ по уровню 40 Дб. И порядка $0.0003F$ по уровню 60 Дб.

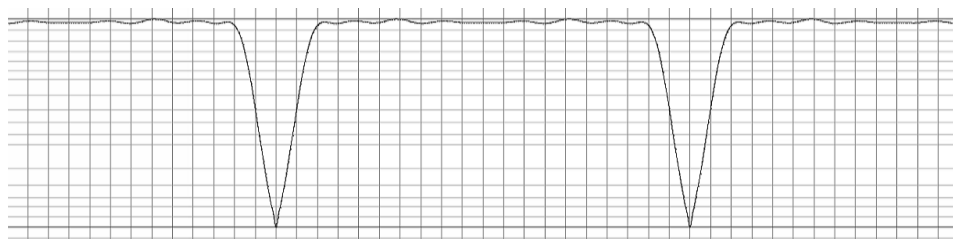


Рис. 20b. Общий вид фрагмента частотной характеристики режектора на основе одиночного компенсационного фильтра с элементами “8–7–6”. Пульсаций практически нет. Диапазон в полосе пропускания по уровню 3 Дб $0.86F$. Полоса подавления по уровню 20 Дб составляет $0.022F$, $0.004F$ по уровню 40 Дб. и порядка $0.001F$ по уровню 60 Дб.

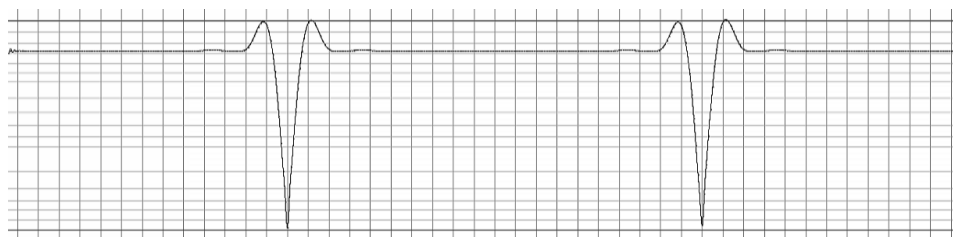


Рис. 20c. Общий вид фрагмента частотной характеристики режектора на основе одиночного компенсационного фильтра с элементами “8–7–6–5”. Диапазон пульсаций 3 Дб в полосе пропускания $0.955F$. Полоса подавления по уровню 20 Дб составляет $0.008F$, $0.002F$ по уровню 40 Дб. и $0.0005F$ по уровню 60 Дб.

А.4. Перемежающиеся гребенчатые фильтры

Исследования, связанные с поиском гребенчатых режекторов, требуемых для преодоления эффектов консонантности, привели к неожиданному побочному результату, показавшемуся при более внимательном рассмотрении настолько важным, что это позволило сформулировать еще одну проблему в отношении поиска гребенчатых фильтров, обладающих специфическими свойствами.

Оказалось, что при использовании знакопеременных весов для отсчетов сигнала на входе гребенчатого режектора он превращается в гребенчатый селектор, выделяющий не все подряд гармоники основного тона, а только лишь нечетные.

Подобрать подходящее название такого рода фильтрам оказалось достаточно сложно. Условно мы их пока называем “перемежающимися”, а особый интерес этот тип фильтров привлек потому, что с его помощью стало возможным решать задачу разделения голосов с консонантностью типа $2/3$, когда основные тона консонирующих источников относятся как $2/3$, без использования режекторов. Обнаружив этот вид гребенчатых фильтров, мы решили, что он является частным случаем гребенчатых фильтров с прореживанием зубцов, когда исключаемыми в фильтре оказываются не только четные его зубцы, а в общем случае исключаются группы из n подряд идущих зубцов, перемежаемые одиночными селектирующими зубцами. Тогда исходные гребенчатые фильтры можно было бы считать перемежающимися порядка 0 (так как n в этом случае оказывается равным 0), найденный нами фильтр оказывается перемежающимся порядка 1, а требуемые для работы с консонантностями более высоких уровней, чем $2/3$, перемежающиеся фильтры будут иметь порядки 2 и более. Однако, нам пока неизвестен общий принцип синтеза таких гребенчатых фильтров.

На Рис. 21a–21b ниже приведены частотные характеристики для сопряженной пары паскалевских перемежающихся гребенчатых фильтров – Рис. 21a для нечетных гармоник, а Рис. 21b – для четных.

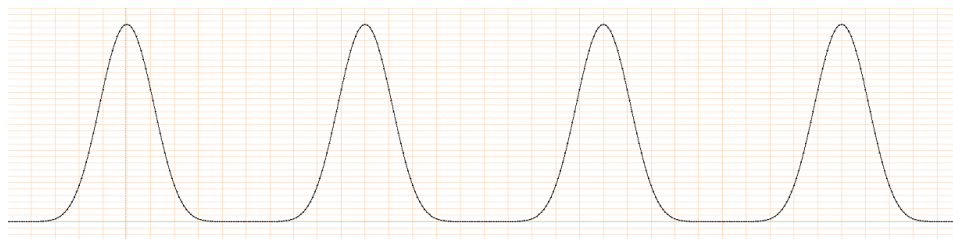


Рис. 21а. Перемежающийся гребенчатый паскалевский фильтр нечетных гармоник: выделяются частоты 50, 150, 250, 350 и т.д. Гц. 50 Гц – основной тон.

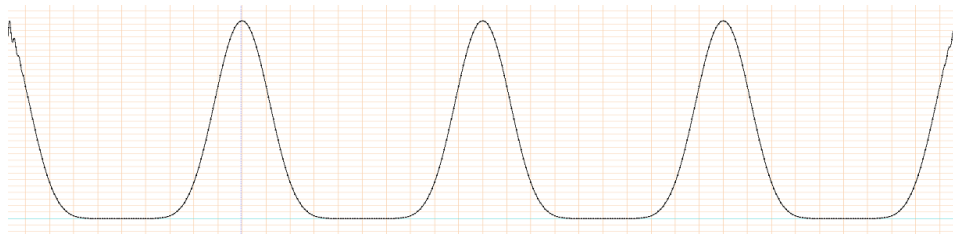


Рис. 21б. Перемежающийся гребенчатый паскалевский фильтр четных гармоник: выделяются частоты 100, 200, 300 и т.д. Гц. 50 Гц – основной тон. Частотная характеристика в этом случае совпадает с характеристикой селекторного фильтра с удвоенной основной частотой.

В. БАЗОВЫЙ АЛГОРИТМ ПРОСЛЕЖИВАНИЯ ГОЛОСОВ В ТРЕКОВОЙ СЦЕНЕ

Спектральная характеристика одиночного трека в непосредственной близости от главного максимума этой функции (в зоне между ближайшими к нему нулями) имеет вид функции $\text{Sin}(x)/x$, и с достаточной точностью можно считать, что эта функция имеет вид равнобедренного треугольника.

Пусть в соответствии с Рис. 22а основание AC треугольника ABC имеет ширину « $2a$ ». Рассмотрим пару симметричных весовых функций, ось симметрии которых проходит через точки D – центр основания и B – вершину треугольника ABC , а сами функции – тоже треугольного вида – EFG и $E'FG$ соответственно. Т.е. первая весовая функция (EFG) имеет линейно убывающий вид, а вторая ($E'FG$) – линейно и с той же скоростью нарастающий.

Если пара таких весовых функций (апертура) расположена симметрично относительно треугольника ABC , то взвешивание этого треугольника каждой из этих функций даст одинаковый результат $W1 = W2$, а их разность $\Delta W = W1 - W2$ будет равна нулю.

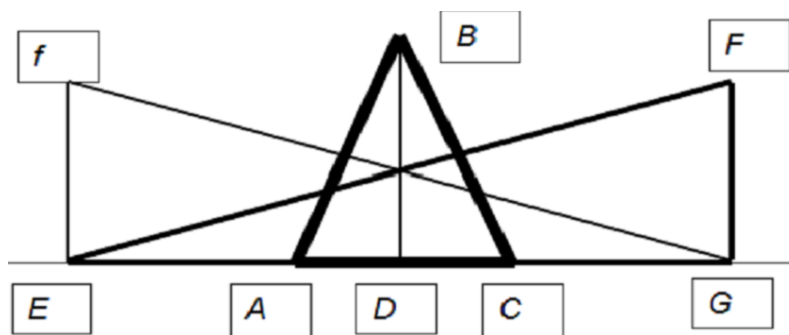


Рис. 22а. Иллюстрация метода отслеживания треков в сцене. Треугольник ABC – поперечное сечение трека. Треугольники EFG и $E'FG$ – две симметричные весовые функции, заданные на апертуре EG . Поскольку на этом рисунке весовые функции расположены симметрично относительно сечения трека, то сечения, взвешенные этими функциями, дают одинаковые величины.

Если теперь весовые функции сдвинуть вправо, как показано на Рис. 22b, то вес ABC , получаемый с помощью весовой функции EfG возрастет, а получаемый с помощью EFG – уменьшится. В результате этого их разность ΔW , став положительной, возрастет. Если сдвиг весовых функций вправо продолжить, то разность весов ΔW будет увеличиваться.

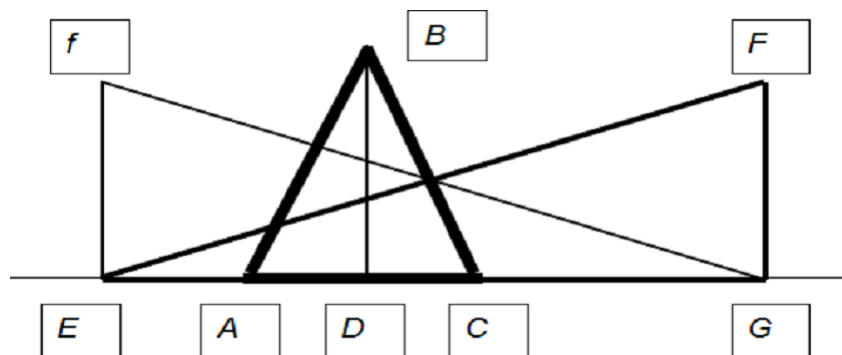


Рис. 22b. Иллюстрация метода прослеживания треков в сцене. Весовые функции сдвинулись относительно сечения трека вправо. Теперь величины, получаемые в результате взвешивания сечения трека функциями EfG и EFG различны: первая более чем вдвое превышает вторую. Поэтому их разность, пропорциональная величине сдвига, будет положительной.

Если же весовые функции из равновесного (нейтрального) положения будут сдвигаться влево, разность ΔW станет отрицательной, но ее абсолютная величина будет увеличиваться тем больше, чем большим будет становиться сдвиг влево. Все это справедливо, конечно, пока апертура и треугольник ABC имеют пересечение по оси x .

Таким образом, использование симметричной пары весовых функций над общим для них основанием позволяет выработать сигнал ΔW , величина и знак которого меняются в зависимости от величины и направления рассогласования в положении центров треугольника ABC и апертуры в соответствии со следующим правилом: если ΔW положительна, то апертуру нужно сдвигать влево, если ΔW отрицательна, то апертуру нужно сдвигать вправо, причем величина сдвигов должна быть равна или пропорциональна производной ΔW по сдвигу. Если коэффициент пропорциональности $k > 1$, то управление оказывается в области перерегулирования и может быть потенциально неустойчивым. При $k \leq 1$ управление будет устойчивым.

Рассмотренный алгоритм обеспечивает устойчивость апертуры на треке в сцене, но не обеспечивает смещения этой апертуры вдоль трека. Понятно, что добиться этого несложно, если после стабилизации апертуры на треке в поперечном его направлении заставить апертуру переместиться в перпендикулярном направлении на некоторую величину – шаг прослеживания.

В новом положении апертуры нужно будет снова повторить цикл ее стабилизации относительно нового положения сечения трека, после чего можно будет сделать следующий шаг прослеживания и т.д., пока в соответствии с какими-то условиями не будет считаться, что прослеживание трека завершено.

Итак, весовые функции в виде проникающих друг в друга “односкатных крыш” обеспечивают возможность устойчивого прослеживания хребтов рельефа сцены – треков. Однако в этой процедуре есть недостаток, становящийся ясным, если представить, что в какой-то ситуации гребень апертурной крыши станет близким к направлению трека. Легко видеть, что в этом случае апертура перестает видеть хребет и, следовательно, не сможет больше осуществлять его прослеживание.

С этой проблемой несложно справиться, допустив, что одновременно используется набор таких “крыш”, гребни которых, например, вытянуты не только вдоль горизонтали, но и по

вертикали и вдоль диагоналей. Число направлений можно и увеличивать, но в реальности и этого абсолютно достаточно для прослеживания треков произвольной сложности, например – в виде замкнутых петель. Алгоритм прослеживания в этом случае усложняется еще одной процедурой, когда управление отдается той паре весовых функций, которой соответствует максимальная в этой точке управляющая величина ΔW . В реальности же, когда для всех реальных треков существует вынужденное направление вдоль горизонтали, для реализации прослеживания достаточно всего лишь трех пар весовых функций – горизонтальной и двух диагональных.

Принципиально важными параметрами этого алгоритма прослеживания треков являются параметр k , управляющий устойчивостью апертуры на треке, шаг прослеживания – St , размер или диаметр самой апертуры – $2R$ (EG на Рис. 22а и 22а) и эффективная ширина трека – $Wt \leq AC$, от учета которых зависят устойчивость, точность и скорость прослеживания:

- Размер апертуры должен быть сопоставим или больше поперечного размера (ширины) типового трека, т.е. $2R > Wt$. В принципе, можно его адаптировать к этому размеру. Это желательно, поскольку от этого зависят величина управляющего сигнала, помехоустойчивость и качество прослеживания.
- Шаг прослеживания не должен превосходить размера апертуры, т.е. $St < 2R$, так как в противном случае прослеживание становится потенциально неустойчивым.
- Величина k , как это было сказано раньше должна быть порядка 1, т.е. $k < \approx 1$, что обеспечивает максимально быструю стабилизацию центра апертуры на треке.

С. ОЧЕВИДНЫЕ ПРОБЛЕМЫ РАЗДЕЛЕНИЯ ГОЛОСОВ

Применение методов Струнного анализа к реальным многоголосым сигналам, включающим наряду с музыкальными инструментами и вокал, выявило следующие проблемы :

1. Разработанные методы выделения и прослеживания треков голосов, достаточно легко теряют прослеживаемые треки, перескакивая на другие, не принадлежащие выделяемому голосу в случае, если эти треки соответствуют вибрато или чирпам, резко меняют свою энергию, соседствуют с другими треками, энергия которых оказывается существенно большей, пересекаются или соприкасаются с другими треками.
2. Частотные характеристики гребенчатых селекторов часто намного превосходят необходимую ширину спектра анализируемого сигнала для выделения голоса, что приводит к подмешиванию спектральных компонент других голосов к спектру выделяемого голоса и к искажению его тембра. То же относится и к проблеме отделения выделяемого голоса от присутствующих в смеси сигналов консонантных голосов, пересекающихся с выделяемым голосом в каких-то подмножествах гармоник.

Главный вывод, при этом, заключается в понимании глубоко органичной связи этих проблем с ролью тембра голоса в анализе и восприятии акустических сцен. Знание тембра позволяет совершить качественный прорыв в выделении и прослеживании треков голосов, переводя на второстепенные роли критерии непрерывности энергии и частоты, использовавшиеся нами для этой цели ранее. Тембр позволяет ограничить требуемым образом спектральную полосу выделяемого голоса, а также дает возможность окончательно восстанавливать искаженные консонантностью голоса. С другой стороны, точность выделения трека голоса позволяет оценивать его тембр и только лишь понимание того, что и как искажено и не искажено консонантными голосами, позволяет определять тембр соответствующего голоса.

Но так как формализация тембра голоса представляет собой довольно трудную задачу, было решено временно остановиться на интерактивном выделении треков, позволяя человеку вручную корректировать формально выделяемые треки.

Программа визуализации акустических сцен “String” позволяет как автоматически корректировать выделенные треки, так и вручную рисовать треки по сцене. Тонкая корректировка также может осуществляться как вручную, так и полуавтоматически путем поочередного смещения точек выделяемого трека в сторону ближайших к ним энергетических всплесков и заглаживания полученного скользящим окном.

Д. ПРОБЛЕМЫ КОНСОНАНТНОСТИ

Консонантностью (созвучностью) в музыке называются ситуации, когда соотношения высот двух голосов таковы, что какие-то их гармоники совпадают. В противном случае говорят о диссонантности голосов. Из сказанного можно сделать несколько важных выводов:

1. Имея лишь фрагмент сигнала с унисонным звучанием голосов, когда совпадают их основные тоны, и ничего не зная сверх этого, разделить эти голоса нельзя. И человек в такого рода ситуации будет воспринимать такую пару голосов как голос с новым тембром. Задача тем не менее становится потенциально разрешимой, если известен тембр хотя бы одного из таких голосов.
2. При октавном сдвиге, когда основной тон одного из голосов сдвинут относительно другого на некоторое количество октав, ситуация оказывается заметно более простой, поскольку хотя в этом случае более высокий голос оказывается полностью искаженным консонантностью, зато второй голос распадается на два набора гармоник, один из которых, включая и основной тон, оказывается не искаженным консонантностью, а второй – искаженным. Это позволяет надеяться, что с эффектами консонантности можно было бы справиться на основании той информации, что содержится в участках сигнала, не искаженного октавным сдвигом. Необходимой дополнительной информацией в этом случае оказывается информация о тембре хотя бы одного из консонирующих голосов.
3. Случай обычной консонантности, когда подмножества совпадающих гармоник двух голосов не включают основные тоны этих голосов, представляется наиболее простым. В этом случае консонантные искажения присутствуют в минимальной степени, при этом в каждом из консонирующих голосов есть свои два подмножества “чистых” и искаженных консонантностью гармоник. Это позволяет при минимальных дополнительных предположениях (типа гладкости спектров голосов по частоте) пытаться с реальными шансами на успех восстанавливать консонирующие голоса только лишь по информации о них, присутствующей в интервалах времени, где они консонируют.

Есть несколько подходов к решению проблем консонантности. Условно мы разделили их на два типа:

- первый назвали “протезирование”.
- второй – “демодуляционным”.

Идея решения в вариантах с протезированием заключается в том, что в голосе, искаженном консонантностью, вырезаются все искаженные гармоники, после чего вместо них подставляются искусственно смоделированные протезы этих гармоник, соответствующие тому, что должно было бы быть, но тут возникают технические проблемы.

Первая проблема связана с тем, что эксперименты по применению режекторных гребенок для подавления в выделяемом голосе консонантных гармоник показали, что полное удаление этих гармоник не происходит, и в исходном спектре сигнала продолжают наблюдаться, хоть и ослабленные, но явные следы удаляемого.

Анализ показал, что это связано с тем, что в реальных сигналах с заметной амплитудно-частотной модуляцией голоса ширина спектральной полосы каждой из его гармоник достаточ-

но велика. В соответствии с теорией модуляции ширина этой полосы минимум вдвое превышает спектр модулирующих колебаний, обычно составляющий несколько герц. Режекторные же фильтры, используемые нами, имеют заметно меньшую эффективную полосу подавления. Поэтому после подавления центральных частей треков по бокам от них остаются “валики” с заметной энергией.

Естественным решением этой проблемы является использование наборов полосовых фильтров, моделирующих гребенчатые с нужными свойствами.

Вторая возникшая на этом этапе проблема оказалась связанной с точностью выделения и отслеживания трека голоса.

Дело в том, что особенностью гребенчатых фильтров является то, что его частотная характеристика в районе основного тона просто тиражируется на все гармоники, обеспечивая по мере роста номера гармоники пропорциональное увеличение селективности.

Из этого следует, что если спектр голоса описывается n гармониками и допустимое отклонение n -ого зубца от n -ой гармоники не должно превышать Δf , то точность соответствия первого зубца основному тону голоса должна при этом составлять не менее $\Delta f/n$. И это резко повышает требования к точности определения основного тона.

Третья проблема связана с определением реальной ширины спектра голоса, выделяемого с помощью гребенчатого селектора, имеющего по сути неограниченную общую частотную характеристику, определяемую лишь теоремой отсчетов, т.е. для сигналов типа музыки – примерно 20 КГц, и все, что попадает в любой из зубцов такого фильтра автоматически относится им к голосу с основным тоном, совпадающим с основным тоном этого фильтра.

Таким образом, чтобы избежать попадания в спектр выделяемого голоса нежелательных гармоник других голосов, нужно в первую очередь научиться определять число эффективных гармоник выделяемого голоса, для чего необходимо распознавать его тембр.

Однако, не все так сложно с обозначенными проблемами, так как в случае использования технологии “резинового времени” их решение кардинально упрощается возможностью использования синтеза гребенчатых фильтров из соответствующих полосовых фильтров в силу того, что операции с сигналами в этом случае осуществляются на фиксированной частотной решетке, задаваемой высотой того основного тона, к которому приводятся эти сигналы.

Е. К ОПРЕДЕЛЕНИЮ ЭФФЕКТИВНОЙ ШИРИНЫ СПЕКТРА ГОЛОСА

Когда голос один, нет никакой проблемы в определении ширины его спектра или присутствующего в нем числа гармоник. Однако, ситуация драматически изменяется, когда голос выделяется гребенчатым фильтром из сложной смеси сходных по природе голосов. Об этом говорилось в связи с третьей проблемой в *Приложении D*.

На примере музыки это можно пояснить следующим образом:

1. Гребенчатый фильтр по определению имеет “бесконечное” число зубцов и простирается по спектру до бесконечности. Фактически это означает, что для “цифрового звука” гребенчатый фильтр будет вырезать спектральные полосы, соответствующие основному тону голоса, вплоть до 20 КГц – максимум половины частоты семплирования при записи сигнала с CD-качеством.
2. Реальные голоса практически никогда не занимают всю спектральную полосу от 20 Гц до 20 КГц.
3. Это значит, что если анализируемый голос имеет высокий основной тон (скрипка, флейта, например), то в сигнале найдутся низкочастотные голоса, верхние гармоники которых попадут в спектр этого голоса, создавая в нем консонантные искажения.

4. Если спектральная ширина зубцов гребенчатого фильтра превосходит критические полосы слуха на соответствующих частотах, то искажения за счет проникновения в выделяемый спектр голоса инородных голосов могут быть и диссонансными тоже.
5. Если же анализируемый голос имеет низкий основной тон (басы, альты), то даже при большом числе его реальных гармоник спектр такого голоса может простирается до 5–8 КГц, но гребенчатый фильтр может добавить к этому голосу гармоники более высокочастотных голосов, попадающие в зоны пропускания такого фильтра за пределами спектра выделяемого голоса.
6. Если анализируемый голос имеет среднюю высоту, скажем – порядка 300–400 Гц, то искажать такой голос будут своими гармониками как более низкочастотные, так и более высокочастотные другие голоса.
7. В силу динамической и частично – частотной несогласованности попадающих за счет этих причин в спектр выделяемого голоса гармоник других голосов тембр этого голоса заметно искажается появлением в нем совершенно атипичных для него призвуков. И во всем этом в конце концов необходимо найти ширину реального спектра выделяемого голоса.

Решение этой проблемы мы связываем с достаточно сильной гипотезой, заключающейся в том, что в одиночном голосе динамика всех его гармоник согласована, поскольку энергетически голос возбуждается как некая целостность.

Следовательно, огибающие гармоник голоса, не искаженные другими тонами, должны быть сильно коррелированы, а искаженные гармоники этого голоса или просто гармоники других голосов могут быть сильно коррелированы между собой, но их корреляция с анализируемым голосом должна быть малой.

Проведенное на тестовых сигналах предварительное исследование степени динамической коррелированности “гармоник”, соответствующих отдельным зубцам гребенчатого фильтра для выделяемого голоса подтвердило эти предположения.

Г. СПЕКТРАЛЬНЫЙ ГРЕБЕНЧАТЫЙ АНАЛИЗ

Развиваемый нами Струнный анализ на основе использования гребенчатых фильтров может рассматриваться как альтернатива или более адекватный анализу акустических сцен метод, чем используемый для этих целей Фурье-анализ. При этом можно говорить о достоинствах и недостатках этих методов в их сопоставлении такого рода, но рациональное зерно заключается в том, чтобы понять, в каких условиях более целесообразно применять тот или иной метод.

Конечно, сегодня в вычислительном отношении Струнный анализ (СА) находится примерно в том положении, в каком был анализ Фурье до изобретения БПФ. Очевидно, что это положение временно, и по этому показателю качественно сравнивать эти методы не стоит.

СА разрабатывался для формирования акустических сцен с целью выделения и анализа в них квазигармонических голосов. С этой задачей СА успешно справляется, и на основе предварительных исследований можно сказать, что он в этом смысле превосходит Фурье анализ, используемый для той же цели.

Исходя из соображений некоторой внутренней однородности вычислительных процессов, для практической реализации исследуемых подходов может оказаться важным, чтобы и спектральный анализ в смысле Фурье-анализа было возможным осуществлять, не выходя за рамки СА. С точки зрения определения тембров голосов, выделяемых с помощью гребенчатых фильтров, это вполне возможно по крайней мере для определения энергетических спектров этих голосов.

Идея подхода к определению энергетического спектра голоса на основе гребенчатой фильтрации заключается в последовательном разложении голоса на подголоса с только лишь четными и нечетными гармониками исходного голоса.

Пусть голос, спектр которого анализируется, определяется n гармониками и имеет общую энергию E_{Σ}^0 . Разделим с помощью перемежающихся гребенчатых селекторных фильтров этот голос на два канала первого порядка C_o^1 и C_e^1 , соответственно – для нечетных и четных гармоник с энергиями E_o^1 и E_e^1 , причем $E_o^1 + E_e^1 = E_{\Sigma}^0$. Применим теперь эту же процедуру к каналам C_o^1 и C_e^1 . В результате получим уже четыре канала второго порядка C_{oo}^2 и C_{oe}^2 для канала C_o^1 и C_{eo}^2 и C_{ee}^2 для канала C_e^1 . И здесь тоже суммы энергий в ветвях будут равны энергиям в узлах, из которых эти ветви исходят: $E_{oo}^2 + E_{oe}^2 = E_o^1$ и $E_{eo}^2 + E_{ee}^2 = E_e^1$. Этот процесс должен быть продолжен, пока на k -ой его итерации общее число порождаемых так уравнений 2^k для энергий отдельных гармоник, входящих в получаемые каналы-ветви бинарного дерева, не окажется достаточным для определения n неизвестных энергий для каждой из гармоник исследуемого голоса.

Проблема фазового спектра при этом остается открытой.

Г. ПЕРСПЕКТИВЫ АНАЛИЗА ТЕМБРОВ

Общепринятое представление о тембре, не считая того, что это – окраска голоса, связано с его определением как статического и динамического спектров голоса. И чтобы лучше разобраться в этом вопросе, есть смысл рассмотреть тембр с нескольких точек зрения.

Со спектральной точки зрения тембр это – статическая и динамическая картины спектра голоса. Естественно, они могут быть скорректированы с учетом известного из психофизики о свойствах слуха. В частности, лет шестьдесят назад в связи с исследованиями по проблеме вокодеров А.А.Пироговым была введена “Фонетическая функция” в виде производной спектра по времени и частоте [8].

Чтобы понять, что в такого рода функциях становится смысловым инвариантом окраски голоса, требуется, как всегда, поставить и решить соответствующую задачу обучения распознаванию, которая позволила бы смоделировать человеческие оценки тембра на обучающих и контрольных образцах музыкальных сигналов. Работа такого рода вполне возможна, но технически достаточно сложна.

С сигнальной точки зрения тембр это – специфический оттенок (окраска) звука, порождаемого некоторым источником, в котором есть возбудитель звука, система параметрических резонаторов, излучатель и среда, в которой живут и постепенно затухают звуковые колебания. Если рассматривать сигнал и его тембр с этой точки зрения, то интегральный звуковой эффект, воспринимаемый ухом и анализируемый при спектральном подходе, может рассматриваться как результат произведения четырех факторов-преобразований, знание которых может позволить целенаправленно решать некоторые обратные задачи по реконструкции управления процессом синтеза голосов на основе результирующего сигнала.

В частности, известные особенности возбудителя позволяют говорить об общих спектральных характеристиках голоса, знание свойств и особенностей управления параметрическими резонаторами источника голоса позволяет выделить, например, области “формант”, характер излучателя может быть связан со скоростью затухания порождаемых источником колебаний, а свойства среды определяют характер реверберационных эффектов. Все это – относительно ортогональные эффекты, что в принципе позволяет существенно упростить решение возникающих обратных задач [9].

Очевидно, что это – достаточно сложная проблема. Есть много свидетельств того, что по крайней мере в отношении восприятия голосов людей мозг такого рода реконструкции про-

изводит. Представляется, что что-то из таких реконструкций могло бы иметь отношение к собственно тембру. Однако, и здесь ожидать простых и быстрых результатов, скорее всего, не приходится.

С точки зрения восприятия тембр это – та самая трудно объяснимая сущность голосов, которая позволяет нам их идентифицировать как самостоятельные объекты восприятия. Кое-что, связанное с этим, считается известным.

В частности, одно время была очень популярна так называемая моторная теория восприятия, в соответствии с которой в мозгу порождается акустический образ источника звука, подстраиваемый с учетом возможностей механизмов звуковоспроизведения таким образом, чтобы слышимый и внутренний голоса минимально различались [10]. Тогда параметры управления синтезом голоса становятся именно тем, что и распознается в конце концов как слышимое.

С помощью этой модели изящно объясняются все известные эффекты слухового восприятия голосов в окружении других голосов, но хорошие формальные модели, требуемые для этого, пока так и не были построены. Однако, современные идеи, связанные с активно развиваемыми ныне методами “глубокого обучения”, дают для этого вполне реальную основу [11].

Исходно мы рассматривали подход к выделению и распознаванию голосов на основе Струнного анализа лежащим именно в такого рода парадигме. Именно поэтому нами последовательно предпринимались возможные попытки привнесения в анализ голосов всего по возможности известного относительно как их природы, так и тех закономерностей, которые сопровождают жизнь таких голосов во времени в сложном взаимодействии с другими голосами мира и музыкальных произведений. И пока у нас нет основания считать, что такого рода установка не будет эффективной, хотя до окончательных результатов исследований пока еще далеко.

Н. ДИНАМИЧЕСКИЙ ТЕМБР ГОЛОСА

В связи с разделением треков голосов следует заметить, что в отличие от таких сравнительно легко поддающихся формализации свойств голосов, как высота и инерционность, формализация тембра голоса представляет собой существенно более трудную задачу, поскольку тембр – это определенного вида ощущение, возникающее у слушателя и характеризующее “окраску” звука.

Допуская, что вся необходимая информация для определения тембра голоса содержится в сигнале, соответствующем ему, можно предположить, что динамический спектр голоса может рассматриваться в качестве аналога или прообраза тембра. Проблема при этом лишь в том, какую информацию из спектра нужно извлекать и как ее преобразовывать.

Чтобы в точке пересечения треков выбрать правильное продолжение, важно лишь, чтобы тембр верного продолжения совпадал или мало отличался от тембра предшествующего участка.

Что конкретно значит “мало отличаться”, является предметом специального исследования, но интуитивно ясно, что определение близости такого рода является задачей, существенно более простой, чем идентификация тембров. Все это имеет смысл, если тембр голоса сравнительно медленно меняется, то есть обладает инерционностью, что, по-видимому, является достаточно естественным допущением.

Следующая по сложности задача разделения треков возникает в том случае, когда два трека, соответствующие разным голосам, на каком-то участке начинают совпадать, что соответствует наступлению унисонного звучания голосов. В этой ситуации можно выделить три фазы:

- отдельные треки (голоса) до момента наступления унисонного звучания
- слитые треки (голоса) на интервале унисонного их звучания

– разделившиеся треки (голоса) после завершения их унисонного звучания.

Если интервал второй фазы имеет нулевую длину, то это соответствует рассмотренному выше случаю пересечения треков (голосов). Если же его длиной нельзя пренебречь, то в дополнение к усложненной перерывом во времени задаче узнавания правильных продолжений треков по инерционности их тембров возникает еще одна и к тому же весьма интересная задача узнавания исходно разделенных голосов в интервале их унисонного совпадения.

Особую пикантность этой задаче придает то обстоятельство, что то, что исходно представляется как интервал унисонного звучания, на самом деле может и не быть таким интервалом, а вместо этого на этом интервале (так как в музыке, если ее рассматривать в качестве универсального примера, почти все возможно) может звучать нечто, естественно соответствующее новому голосу.

В психоакустике бытует мнение, что для восприятия тембра обычно достаточно первых пяти гармоник голоса (основной тон при этом считается первой гармоникой) [12, 13]. Если пренебречь зависимостью тембра от громкости и считать, что тембр с достаточной для наших целей точностью определяется формой своей огибающей, то отнормировав спектр по амплитуде или энергии основного тона, мы получим 4 параметра – относительные энергии выбранных пяти гармоник, начиная со второй, которые полностью описывают спектр-тембр. Понятно, что число рассматриваемых так гармоник и способ нормировки могут быть иными, но это по сути мало что меняет. Таким образом, можно считать, что каждый из отслеживаемых в сцене треков в каждый момент времени характеризуется четырехмерным вектором тембра. Теперь, располагая этой информацией, мы можем решать поставленную задачу.

СПИСОК ЛИТЕРАТУРЫ

1. Harshada P. Burute, Madhuri Patil, Kirtimalini Chaudhari, Dr. Pradeep B. Mane. Comparative Study of Filter Performance for Separation of Singing Voice from Music Accompaniment. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, April 2015, vol. 3, no. 4, pp. 58–61.
2. M.Gainza, B.Lawlor, E.Coyle. Harmonic Sound Source Separation using FIR Comb Filters. *117th. AES (Audio Engineering Society) Convention*. October 28–31, San Francisco, 2004.
3. Yipeng Li, DeLiang Wang. Separation of Singing Voice From Music Accompaniment for Monoural Recordings. *IEEE Transactions on Audio, Speech, and Language Processing*, May 2007, vol. 15, No. 4, pp 1475–1487.
4. Вайнштейн Л.А., Зубаков В.Д. *Выделение сигналов на фоне случайных помех*. М.: Советское радио, 1960.
5. В.К. Слока. *Вопросы обработки радиолокационных сигналов*. М.: Советское радио, 1970.
6. В.Г.Поляков, В.С.Переверзев-Орлов. *Электронные системы следящей развертки*. М.: Энергия, 1968.
7. А.А.Харкевич. *Спектры и Анализ*. Москва; Ленинград: Государственное издательство теоретико-технической литературы, 1952.
8. А.А.Пирогов. Основы фонетической теории речи. Фонетическая функция как универсальный природный инструмент кодирования (декодирования) речевой информации любого происхождения. *ЖРФМ*, 2001, №1–12, стр. 15–28.
9. В.Н. Сорокин. *Речевые процессы*. М.: Народное образование, 2012.
10. Л.А. Чистович, А.В. Венцов, М.П. Гранстрем. *Физиология речи. Восприятие речи человеком*. Л.: Наука, 1976.

11. В. Перервенко. *Третье поколение нейросетей: “Глубокие нейросети”*. URL: <https://www.mql5.com/ru/articles/1103>
12. Тембр. URL: <https://ru.wikipedia.org/wiki/тембр>
13. И. Алдошина. *Основы психоакустики*. URL: <https://docplayer.ru/38385771-Osnovy-psihoakustiki-chast-1-irina-aldoshina.html>

The possibility of separating a complex signal into quasi-harmonic voices

E.A. Vashchenko, M.A. Vitushko, V.S. Pereverzev-Orlov

Quasi-harmonic voices are the most prominent part in the signals around us. The article proposes a method for analyzing a complex signal and extracting a separate voice from it, based on the possibilities of comb filtering. It is shown that this approach allows one to efficiently extract individual quasi-harmonic voices from complex signals such as those of orchestral and opera music. This gives grounds for the development of the comb filtering method of quasiharmonic signals as an alternative to the widely used Fourier spectral representation.

KEYWORDS: complex signal, voice extraction, comb filtering.