РОССИЙСКАЯ АКАДЕМИЯ НАУК
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР

СООБЩЕНИЯ ПО ПРОГРАММНОМУ ОБЕСПЕЧЕНИЮ ЭВМ

Чучупал В.Я., Чичагов А.С.Маковкин К.А.

ЦИФРОВАЯ ФИЛЬТРАЦИЯ ЗАШУМЛЕННЫХ РЕЧЕВЫХ СИГНАЛОВ

Вычислительный центр РАН
Москва, 1998

УДК 621.391;518.5

Ответственный редактор
академик Журавлев Ю. И.

В работе приводится описание и основные принципы функциониро-вания программного комплекса для цифровой обработки речевых сигналов, содержащих аддитивные шумы и импульсные помехи. Рассмотрены основные методы фильтрации речевых сигналов, основанные на фильтрах Винера, вычитании амплитудных спектров, цифровых фильтрах и скрытой марковской моделью.

Рецензенты: к.ф.м.н. К.В.Огарышев
к.ф.н. В.Б.Кузнецов

Научное издание

1. Классификация цифровых методов повышения качества и разборчивости речи[1]

Речевые сигналы, с которыми приходится иметь дело на практике, всегда в той или иной степени зашумлены. В тех случаях, когда шум имеет значительную интенсивность, его наличие может существенно исказить результаты обработки, анализа или распознавания речи. В целом ряде других случаев, например, при анализе зашумленных записей в криминалистических целях или восстановении аудиозаписей в архивах, задача очистки сигнала от шума носит самостоятельный характер и является единственной целью работы. Поэтому разработка методов очистки сигнала от шума является весьма актуальным направлением исследований.

К настоящему времени разработано очень большое количество различных методов цифровой обработки зашумленных речевых сигналов. В связи с этим следует отметить, что в настоящий обзор не включены специализированные методы предобработки речевого сигнала (которые тоже можно трактовать как очистку от акустического шума), обеспечивающие повышение робастности систем распознавания речи при работе в шумных условиях и нелинейных искажениях в канале передачи, такие как RASTA [Hermansky, 1994], удаление смещения сигнала [Rahim, 1996], вычитание кепстрального среднего[Rahim, 1996] и афинные преобразования кепстра [Mammone, 1996]. Это сделано по той причине, что эти методы не носят самостоятельного харакетра и относительно них не выполнялось каких-либо измерений, которые позволили бы оценить качество обработки в терминах изменения разборчивости речевого сигнала или качества речи (выраженного отношением сигнал/шум).

Основным типом шумов, для методов, представленных в обзоре, является аддитивный шум.

В целях упорядочения рассмотрения методов очистки сигнала от шума целесообразно произвести их классификацию. Основным признаком, по которому будут классифицироваться алгоритмы, является характер или тип тех закономерностей, которые служат основой для выделения речевого сигнала из смеси с шумом. В качестве вспомогательного признака будет использоваться классификация по типу того математического или алгоритмического аппарата, который использован для фильтрации. Подобная классификация, конечно, весьма условна, так как многие из рассматриваемых методов нельзя безоговорочно отнести к какой-либо одной категории. Как правило, одни и те же методы используют одновременно различные принципы, и в этом случае можно говорить лишь о преимущественном влиянии какой-либо концепции.

С учетом сделанного замечания можно выделить следующие группы методов цифровой обработки зашумленных речевых сигналов:

---

[1] Работа проводилась при поддержке гранта РФФИ 96-0101402

методы адаптивной компенсации помех;

методы, основанные на использовании математических моделей речевых сигналов во временной области (например, авторегресионная модель речевого сигнала и рекуррентные алгоритмы оценки параметров и речевого сигнала);

методы, основанные на использовании математических моделей речевых сигналов в частотной области (оценивание минимальной среднеквадратической ошибки, марковские модели сигнала и шума);

методы, основанные на использовании спектральных характеристик шума (вычитание амплитудных спектров, винеровская фильтрация);

методы, основанные на использовании моделей искусственных нейронных сетей;

методы, основанные на моделях восприятия речи человеком;

Перейдем к рассмотрению конкретных методов цифровой обработки зашумленных речевых сигналов.

1.2. Адаптивные компенсаторы помех

Этот класс методов цифровой обработки зашумленных сигналов основан на использовании, помимо собственно зашумленного сигнала, который подлежит очистке, также одного или нескольких опорных сигналов - сигналов, которые коррелированы с шумовым сигналом и некоррелированы (или слабо коррелированы) с полезным сигналом, подлежащим выделению. С помощью опорных сигналов формируется сигнал, который является оценкой помехи. Этот сигнал затем вычитается из зашумленного сигнала и результат этой операции рассматривается как оценка незашумленного сигнала. На рисунке 2.1 представлена схема адаптивного компенсатора помех, который использует один опорный сигнал [Widrow, 1975].

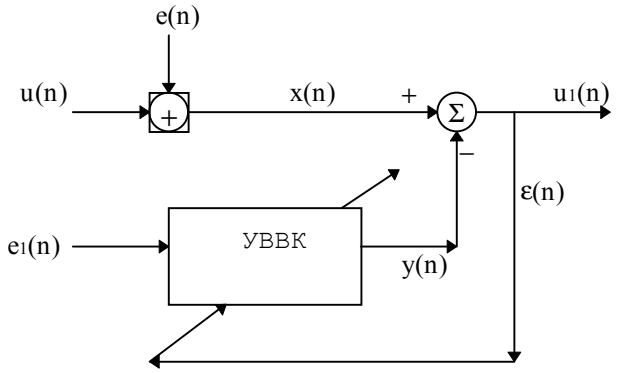


Рис 2.1. Схема адаптивного компенсатора помех. u(n) - дискретный отсчет полезного сигнала в момент времени n, n=0,1,2… e(n) - шумовой сигнал, e1(n)- опорный сигнал, ε(n) - сигнал ошибки, u1(n) - выходной сигнал компенсатора, УУВК - устройство управоения весовыми коэффициентами.

Наиболее важной частью адаптивного компенсатра помех является устройство управления весовыми коэффициентами - линейный фильтр, через который пропускается опорный сигнал e1(n). Задача адаптивной компенсации помехи e(n) сводится к подбору коэффициентов фильтра таким образом, чтобы минимизировать энергию сигнала на выходе компенсатора u1(n). В этом случае будет максимизировано выходное отношение сигнал/шум. Минимизация энергии обычно осуществляется на основе градиентных методов поиска экстремума функций многих переменных[Моисеев, 1978].

Известно, что адаптивные компенсаторы помех позволяют значительно улучшить качество зашумленных сигналов - на несколько десятков децибел [McWhirer, 1982], но требование наличия опорного сигнала существенно сужает их область применения. Во многих приложениях цифровой обработки речевых сигналов (например, при реставрации архивных записей или в криминалистике), опорного сигнала, по крайней мере, в явном виде, не имеется. Поэтому для применения методов адаптивной компенсации помех опорный сигнал в таких случаях приходится получать на основе косвенных соображений, связанных с осбенностями речевого сигнала, а сам адаптивный компенсатор в этом случае будет являться одной из составных частей более сложного алгоритма выделения речевого сигнала.

1.3. Методы, основанные на использовании статистических моделей речевых сигналов во временной области.

Класс методов цифровой обработки зашумленных речевых сигналов, который основан на построении математических моделей речевых сигналов и обработке речевых сигналов с использованием этих моделей быстро развивается и в настоящее время эти методы привдят к самым успешным результатам. Задача выделения речевого сигнала из смеси с шумом в случае использования достаточно адекватной модели сводится к оценке каким-либо образом параметров этой модели и последующим синтезом или фильтрации речевого сигнала фильтром, построенным на основе или с помощью оцененных параметров.

Одними из наиболее перспективных методов в этом классе являются методы статистической фильтрации во временной области, которые развивались в работах [Прохоров 1977, 1980, Гурьев, 1982]. Фильтрация речевого сигнала, моделируемого авторегрессией, осуществляется при этом методами теории оптимального оценивания, напрмер, с помощью построения оптимального линейного фильтра (фильтра Калмана [Сейдж, 1976]).

Предположим, что некоторая линейная система с переменными параметрами возбуждается шумовым сигналом w(k), где k - индекс, соответствующий дискретному времени. Соотношение между выходным сигналом системы x(k) (вектором состояния) и сигналом возбуждения w(k) в момент времени к=1 будет иметь вид

$$x(k) = F(k+1, k)\, x(k) + G(k)w(k) \qquad (3.1)$$

В (3.1) предполагается, что сигналы x и w - векторные величины, компоненты которых являются случайными величинами. Матрицы F(k+1,k) и G(k) характеризуют состояние системы в соответствующие моменты времени.

Допустим далее, что вектор состояния неизвестен и требуется произвести его оценку по наблюдаемым (до момента времени k включительно) величинам z(k) (наблюдениям), которые связаны с вектором состояния x(k) соотношением:

z(k) = H(k)x(k)+v(k) (3.2.),

где v(k) - шум наблюдения, который нужно отфильтровать.

Если заданы матрицы F(i,i+1), G(i),H(i), 0<=i<=k, определены статистические свойства шумов w,v и указаны подходящие начальные условиях в нулевой момент времени: x(0), то оптимальная, по критерию минимума дисперсии ошибки, линейная оценка вектора состояния x(i) по наблюдениям ,z(1),z(2),…,z(i), для процесса, описываемого соотношениями (3.1), (3.2), дается в рекуррентном виде алгоритмом фильтрации Калмана [Сейдж, 1976]:

x(i) = F(i+1,i)*x(i-1)+K(i)[z(i)-H(i)f(i,i-1)x(i-1)] (3.3)

Одной из наиболее распространенных моделей речевых сигналов является модель авторегрессии, либо ее эквиваленты [Прохоров, 1977].

В соответствии с этой моделью речевой сигнал {x(n)}, n=..-1,0,1,. ..описывается уравнением авторегрессии:

$s(n) = \sum_{k=1}^{p} a(k)s(n-k) + be(n-1)$ (3.4)

где e(n)- последоваельность некоррелированных случайных величин, таких, что E(e(n)), $E(E_2(n))=1$, n=…-1,0,1,…; a(k)- параметры модели, b - постоянный коэффициент. Величина p называется порядком модели.

В соответствии с гипотезой о локальном постоянстве параметров, параметры модели авторегрессии обычно считаются постоянными в течение малых промежутков времени (10-20 мс.), либо каким-то образом задается закон их изменения.

Пусть речевой сигнал s(n) зашумлен аддитивным шумом v(n). Оозначим наблюдемую последовательность через {z(n)}, n=..-1,0,1,…:

z(n) = s(n) + v(n) (3.5)

Соотношения (3.4) и (3.5) являются частными случаями соотношений (3.1) и (3.2), когда:

x(n) = [s(n), s(n-1), s(n-2), …, s(n-p+1)]t

w(n) = [e(n), 0, 0, …, , 0]t

$$F(n+1, n) = \begin{pmatrix} a(1) & a(2) & \dots & a(p) \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

G(k) = diag ||b||,

z(n) = [z(n), z(n-1), z(n-2), …, z(n-p+1)]t

H(n) = E, единичная матрица размерности p

v(n) = [v(n), v(n-1), v(n-2), …, v(n-p+1)]t

Одно из отличий рассматриваемой задачи от задачи (3.1),(3.2) заключается в том, что помимо фильтрации речевого сигнала s(n), требуется производить также оценку параметров a(k). Такая оценка может быть произведена разными способами. Например, считая параметры a(k) изменяющимися со временем и задавая закон их изменения в виде:

$$a(n) = \Lambda a(n-1) + \Lambda_0 + be(n-1) \qquad (1.7)$$

(где a(n) - вектор параметров, Λ - матрица, описывающая связи между параметрами, Λ0 - диагональная матрица с постоянными значениями, b - вектор постоянных значений, e(n) - случайная последовательность типа белого шума с нулевым средним и единичной дисперсией) можно оценить параметры модели рекуррентно с помощью оптимального линейного фильтра (3.3)[Назаров, 1982].

Дальнейшее развитие методы статистической фильтрации получили в работах [Санников, 1982, Гурьев, 1982]. В этих работах рассматриваются новые нелинейные модели речевых сигналов, которые более точно описывают такие особенности речеобразования как квазипериодичность звонких звуков. Рассмотрены вопросы накопления априорных данных о речи диктора - создания банка данных, в котором хранились бы оценки параметров, аналогичных матрицам в модели (3.6), в целях использования такого рода сведений для обработки зашумленного речевого синала. Для фильтрации речевого сигнала и оценки параметров рассматривается применение нелинейного оптимального оценивания на основе метода инвариантного погружения [Гурьев, 1982].

Экспериментальные исследования алгоритмов фильтрации, построенных на таких принципах показали. что они могут дать заметный выигрыш в качестве речевого сигнала: + 8 - +10 дб при первоначальном отношении сигнал/шум около 0 дб.

Модель (3.4) является линейной моделью, передаточная функция которой содержит только полюса. Для описания некоторых звуков речи, например, назальных, “м”. “н”, более адекватной является линейная модель авторегрессии со скользящим средним:

$$s(n) = \sum_{k=1}^{p} a(k)s(n-k) + \sum_{m=0}^{q} b(m)e(n-m) \qquad (3.7)$$

передаточная функция этой модели содержит как полюса так и нули. Поэтому для улучшения качества речевых сигналов часто выгоднее использовать именно эту модель. Сравнение нескольких методов моделирования зашумленного речевого сигнала на основе модели (3.7) показало. что выигрыш в отношении сигнал/шум в этом случае примерно на 5 дб превышает аналогичный выигрыш при использовании тех же методов фильтрации, но для авторегрессионной модели речевого сигнала. (Первоначальное отношение сигнал/шум в эксперименте составляло 0 - 5 дб).

Вычислительно эффективная (но с менее удачным результатом обработки) реализация алгоритма фильтрации речевого сигнала, моделируемого

авторегрессионной моделью с параметрами, связанными в марковскую цепь, предложена в [Lee, 1996]. Совместная оценка сигнала и параметров марковской цепи вычисляются рекуррентным способом с помощью алгоритма максимизации математического ожидания (expectation maximization - EM), причем для вычисления условного ожидания (expectation step) сигнала относительно наблюдений использован фильтра Калмана-Бьюси.

Экспериментальные испытания на речевом сигнале в смеси с некоррелированным аддитивным белом шумом с отношениями сигнал/шум 0, 10 и 20 дб показали увеличение отношения сигнал/шум в среднем на 4 дб.

Авторегрессионная модель речевого сигнала, как показывает практика, не имеет такого выраженного дефекта как музыкальные тона, однако, артифакты обработки также имеют место. В целом фильтрация высоко-энергетических частотных диапазонов звуков выполняется точнее, в то время как моделирование нулей и минимумов спектра затруднено[Yoo, 1996]. Этот эффект непосредственно связан с порядком p используемой модели авторегрессии (3.4).

Высокие (предварительные) результаты продемонстрировал комбинированный метод анализа и фильтрации речевого сигнала, основанный на частотной декомпозиции сигнала с последующим раздельным моделированием сигнала в каждой частотной полосе моделью авторегрессии [Yoo, 1996].

Речевой сигнал декомпозируется в несколько частотных полос, в каждой из которых оценивается отношение сигнал/шум, на основании каторого выбирается порядок p модели сигнала в этой частотной полосе. В общем случае правило выбора веричины p тиково, что сигнал в частотном диапазоне с высоким отношением сигнал/шум (как правило это соответствует формантным максимумам речевого сигнала) моделируется авторегрессией большего порядка, а сигнал для частотных полос с низким отношением сигнал/шум (нули спектра сигнала) моделируется авторегрессией более низкого порядка.

Собственно фильтрация осуществляется модифицированным фильтром Винера в частотной области. Предварительные измерения (смесь речи с белым шумом) показали значительное увеличение отношения сигнал/шум: на +15 дб при начальном отношении -5 дб (соответственно, при начальном SNR +5 дб улучшение составило 11 дб).

1.4. Методы, основанные на обработке речевого сигнала с использованием аппарата скрытых марковских моделей.

Другим классом методов обработки зашумленных речевых сигналов основанных на использовании статистических моделей речевого сигнала являются методы, в которых речевой сигнал моделируется скрытой марковской цепью. То есть для моделирования речевого сигнала использован наиболее эффективный для распознавания речи подход.

Известно, что традиционно используемые методы фильтрации (вычитание спектров или фильтр Винера) не используют фонетическую информацию,

переносимую речевым сигналом. Недавние исследования [Hansen, Pellom, 1997] показали, что знание и применение в процессе обработки фонетической структуры сигнала приводит к улучшению качества фильтрации. Поэтому вполне естественным является применение в процессе очистки речевого сигнала от шумов его статистической модели в виде скрытой марковской цепи, которая связана с фонетической структурой сигнала.

Идея реализации такого подхода заключается в том. что первоначально, по записям незашумленного речевого сигнала строятся статистические модели единиц речевого потока (фонов либо более широких классов звуков). После того, как статистическая модель для множества состояний сигнала построена, по ней можно рассчитать оптимальный фильтр Винера.

При обработке зашумленного речевого сигнала сначала оценивается (по отфильтрованному на предыдущем шаге сигналу) текущее состояние марковской модели, в соответствии с которым выбирается оптимальный фильтр, который затем используется для фильтрации сигнала и получения очередной оценки.

Алгоритм фильтрации выглядит следующим образом [Sheikhzadeh, 94]:

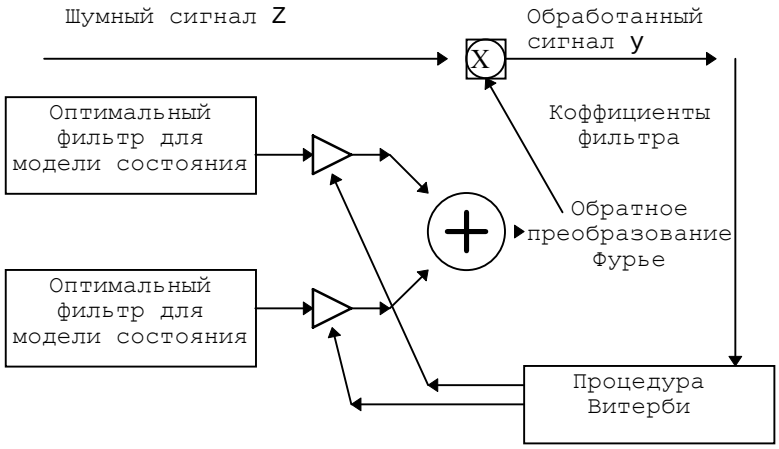


Рис 4.1. Алгоритм фильтрации речевого сигнала с использованием скрытой марковской модели

В исследованиях [Sheikhzadeh, 94] сначала, используя стандартную базу данных (марковские модели формировались на TIMIT) строились модели состояний для незашумленного речевого сигнала. Для каждого состояния модели β и каждой гауссовской составляющей (реально использовалась только одна составляющая на состояние) оценивался оптимальный фильтр Винера $H\beta(\theta)$. В процессе обработки сигнала на каждом шаге (кадре анализа)с помощью процедуры Витерби вычислялись правдоподобия состояний в соответствии с которыми выбирались весовые коэффициенты $W\beta$ для каждого фильтра $H\beta(\theta)$. Очистка сигнала затем производилась в частотной области в соответствии с:

$$y\ (iw,k) = \left[\sum_{\beta} W_{\beta}(k) H_{\beta}^{-1}(iw)\right]^{-1} z(iw,k)$$

где k - номер итерации (кадра анализа), Wβ - вес фильтра Hβ(θ), z(iw) - спектральная компонента зашумленного сигнала и y(iw) - спектральная компонента обработанного сигнала.

Для эффективной обработки нестационарных сегментов отдельно оценивалась марковская модель шума. В отличие от простых моделей состояний полезного сигнала шум моделировался набором состояний, каждое из которых содержало несколько гауссовских компонент.

Во время обработки зашумленного сигнала при определении отсутствия полезного сигнала выполнялось декодирование сегмента паузы процедурой Витерби для выбора оптимальной модели шума. Модель шума, обеспечивающая максимальное правдоподобие наблюдаемой последовательности использовалась далее для обработки сигнала.

Для сохранения ”преемственности” между итерациями применялись инерционная схема фильтра Винера.

В реальных экспериментах по фильтрации речевого сигнала описанными методами число состояний марковской модели речевого сигнала выбиралось равным 5, то есть фактически речевой сигнал моделировался как последовательность фонов, соответствующих широким фонетическим категориям (звонкий-шумный-глухой и т.п.).

Сравнительные исследования выполнялись на различных типах шумов (белый, симуляция шума вертолете, одновременный разговор нескольких мешающих дикторов) Обьективные измерения (изменение отношения сигнал/шум на входе и выходе системы) показали очевидное превосходство описанной методики (в среднем улучшение +2.5 децибела, в диапазоне от 0 до +20 дб, причем при отношении сигнал/шум > 10 дб. превосходство составило в среднем +7 дб) над стандартной системой фильтрации, построенной по методу вычитания амплитудных спектров. Субьективные тесты (оценка качества звучания обработанного сигнала на слух по пятибальной шкале) также показали превосходство марковских моделей над стандартными методиками. По мнению авторов, разборчивость речи в результате обработки также повысилась, вероятно вследствие того, что низкоэнергетические звуки в данном случае обрабатываются существенно аккуратнее.

Последовательное развитие этого подхода привело к использованию алгоритма минимизации среднеквадратической ошибки[Ephraim, 1990] вместо процедуры Витерби [Ephraim, 1992; Sheikhzadeh, 1995] с улучшеннвми возможностями.

Следует отметить, что очевидным недостатком подхода является необходимость иметь априорную информацию о возможных типах шумов (в виде предварительно обученных марковских моделей состояний). Типов возможных шумов для различных практически важных условий много и требование наличия заранее вычисленных моделей представляется маловыполнимым.

Кроме того, известно, что качество обработки сигнала ухудшается в тех случаях, когда помеха имеет существенно нестационарный характер.

В связи с этим дальнейший прогресс в этом направлении может быть достигнут за счет использования более гибких способов моделирования помех. В работе [McKinley, 1996] предлагается эффективный алгоритм для переоценки параметров вероятностной модели шума, основанный на адаптивной подстройке элементов кодовой книги (которая состоит из матрицы корреляционных коэффициентов для авторегрессионной скрытой марковской модели) по методу скользящего среднего. Показано, что подобная методика приводит к дополнительному выигрышу примерно 2,3 дб (по сравнению с исходным алгоритмом фильтрации, основанным на марковской модели[Ephraim, 1992]) на нестационарных шумах и не ухудшает качества обработки для стационарных помех.

Использование марковских моделей речевого сигнала оказывается выигрышным при наличии априорной информации синтаксического характера о зашумленном речевом сигнале. Довольно часто (например, при анализе черных ящиков с борта самолета или реставрации аудиозаписей) время обработки сигнала не играет определяющей роли.

В тех случаях, когда аудитор может приблизительно указать (дать гипотезы), что именно было произнесено, или на что похоже зашумленное высказывание, можно улучшить качество сигнала с помощью использования аппарата марковских моделей, построенных для предстазанных звуков.

Анализ метода фильтрации речевых сигналов с помощью марковских моделей при наличии гипотез о произносимом тексте выполнен в работе [Hansen, 1997]. Было продемонстрировало улучшение, в результате обработки, качества звучания сигнала в условиях различных типов помех и в широком диапазоне отношений сигнал/шум.

1.5. Методы, основанные на использовании, отдельных характерных свойств речевого сигнала.

К методам этого типа относятся прежде всего класс методов обработки зашумленных речевых сигналов, которые используют квазипериодичность речевого сигнала.

Первая группа методов использует периодичность речевых сигналов для построения адаптивного компенсатора помех, с помощью которого обрабатывается зашумленный речевой сигнал. Предполагается, что исходный речевой сигнал s(n) в (3.5) строго периодичен с периодом T, кратным частоте дискретизации, а случайный аддитивный шум v(n) некоррелирован с s(n). В качестве опорного сигнала для адаптивной компенсации помехи используется

$r(n) = z(n) - z(n+T), n = \ldots, -1, 0, +1, \ldots$

Результаты применения описанного метода приводятся в работе Сэмбера [Sambur, 1978]. Отношение сигнал/шум может быть увеличено на 7 - 10 дб., однако разборчивость отфильтрованной речи при этом несколько понижается.

Вторая группа методов, использующих периодичность звонких звуков основанан на представлении сигнала в кепстральной области. В этом случае периодический характер речевого сигнала используется для синтеза адаптивной гребенки фильтров [Lim, 1978, Lyon, 1982].

Как известно, периодичность звонких звуков выражается в частотной области в том, что их спектр имеет линейчатый характер, причем соседние пики (спектральные максимумы) отчтоят друг от друга на интервал (в частотной области) равный частоте основного тона.. Поэтому, если гребенка фильтров такова, что гармоники основного тона (спектральные пики) попадают в полосы пропускания, то можно расссчитывать на повышение качества речевого сигнала. Во временной области фильтрацию речевого сигнала гребенкой фильтров с равноразнесенными по частоте каналами можно представить соотношением [Lyon, 1982]:

$$y(n) = \sum k=-L+La(k)s(n-kT),\ n=\ldots-1,0,1,..;\ -L<= k<=+L \quad (5.1)$$

Экспериментальное исследование адаптивной гребенки фильтров показало, что достигаемое улучшение качества речевого сигнала невелико, а в тех случаях, когда помеха носит структурированный характер, фильтация такого рода гребенкой фильтров вообще неэффективна. Усовершенствованная модификация рассмотренного выше подхода предложена Кохом и Малахом и заключается в том, что коэффициенты a(k) в соотношении (5.1) зависят не только от k, но и от n(mod T).

Исследования проведенные на синтетических гласных звуках показали, что при надлежащем выборе взвешивающих коэффициентов a(k,n) можно добиться значительного эффекта для улучшения восприятия речи в тех случаях, когда помеха или шум являются структурированными [Malah, 1982].

Обобщение метода (5.1) является также известный адаптивный фильтр Фрезиера [Frazier, 1976]. В этом случае учитывается изменение чаастоты основного тона на интервале времени, равном длине импульсной характреистики гребенки фильтров. Исследование характеристик такого фильтра показали, что он может дать выигрыш в отношении сигнал/шум до 10 дб, однако при этом несколько снижается разборчивость речи.

Только что описанные в этом разделе методы имеют существенные недостатки. Помимо того, что все методики, кроме фильтра Фрезиера, не учитывают изменений частоты основного тона, эти методы непригодны для фильтрации глухих звуков.Например, существенные для разборчивости речевого сигнала звуки “п”, “т”, “к” не могут быть успешно обработаны такими методами. Наконец, качество обработки сигнала зависит от точности оценки частоты основного тона в зашумленном речевом сигнале, что само по себе не всегда возможно.

То обстоятельство, что на разборчивость речи существенно влияет правильное восприятие согласных звуков, в частности, взрывных “п”, “т”, “к” и шумного “с”, было использовано в системе [Drucker, 1968]. Фильтрация речевого сигнала заключяалась в том, что перед взрывными звуками вставлялась короткая пауза, - смычка, а согласный “с” фильтровался специально подобранным фильтром.

Испытания показали, что такая обработка повышает разборчивость речевых сигналов. Существенным недостатком предложенной методики является необходимость в априорной информации о местонахождении взрывных звуков и звука “с”.

Работа [Drucker, 1968] оказалась важной потому, что это была первая работа по фильтрации речевого сигнала, в которой было предложено обрабатывать различные типы звуков (было выделено пять категорий звуков - фрикативные, взрывные, назальные, гласные и глайды) разными процедурами оценивания параметров и фильтрации.

1.6. Методы, основанные на оценке спектральных характеристик шума.

Наиболее часто используемыми методами, основанными на использовании спектральных характеристик шума, являются методы, реализующие различные модификации алгоритма вычитания амплитудных спектров[Boll, 1979, Sondhi, 1981].

В качестве обоснования этих методов приводятся следующие соображения. Если стационарный сигнал s(t), t=…-1,0,1,.. со спектральной плотностью мощности $P_{ss}(iw)$ искажен аддитивным стационарным шумом n(t) со спектральной плотностью мощности $P_{nn}(iw)$, который предполагается некоррелированным с x(t), то спектральная плотность мощности зашумленного сигнала x(t) - $P_{xx}(iw)$ равна:

$P_{xx}(iw) = P_{ss}(iw) + P_{nn}(iw)$

следовательно спектральная плотность мощности полезного сигнала s(n) может быть оценена как:

$P_{ss}(iw) = P_{xx}(iw) - P_{nn}(iw)$ (6.1)

В силу нестационарности речевых сигналов использовать соотношение (6.1) непосредственно нельзя. На практике, при обработке речи на достаточно коротких участках, напрмер, квазистационарных участках гласных звуков, величины $P_{xx}(iw)$, $P_{nn}(iw)$ аппроксимируют с помощью усредненных квадратов кратковременных амплитудных спектров наблюдаемого сигнала и шума. Спектр шума при этом должен оцениваться в моменты пауз. Полученная таким образом оценка соответствует квадрату амплитудного спектра сигнала. Восстановление речевого сигнала во временной области осуществляется с помощью обратного преобразования Фурье, причем фазовый спектр для восстановленного сигнала берется таким же, как и у наблюдаемого сигнала.

В наиболее общем виде операция спектрального вычитания может быть выражена соотношением:

$$|S(t,iw)|^2 = \begin{cases} |X_i(t,iw)|^2 - A(t)|N(t,iw)|^2, & \text{если } |X_i(t,iw)|^2 \geq (A(t)+B)|N(t,iw)|^2 \\ B|N(t,iw)|^2, & \text{в противном} \end{cases}$$

Здесь коэффициент A(t)(фактор переоценивания), вообще говоря, зависит от

соотношения сигнал/шум на сегменте анализа, И имеет типичные значения близкие к 0.7 - 0.95, а коэффициент В (спектральный порог)- выбирается в диапазоне 0.01 - 0.1.

Блок-схема алгоритма вычитания амплитудных спектров привенена на следующем рисунке.

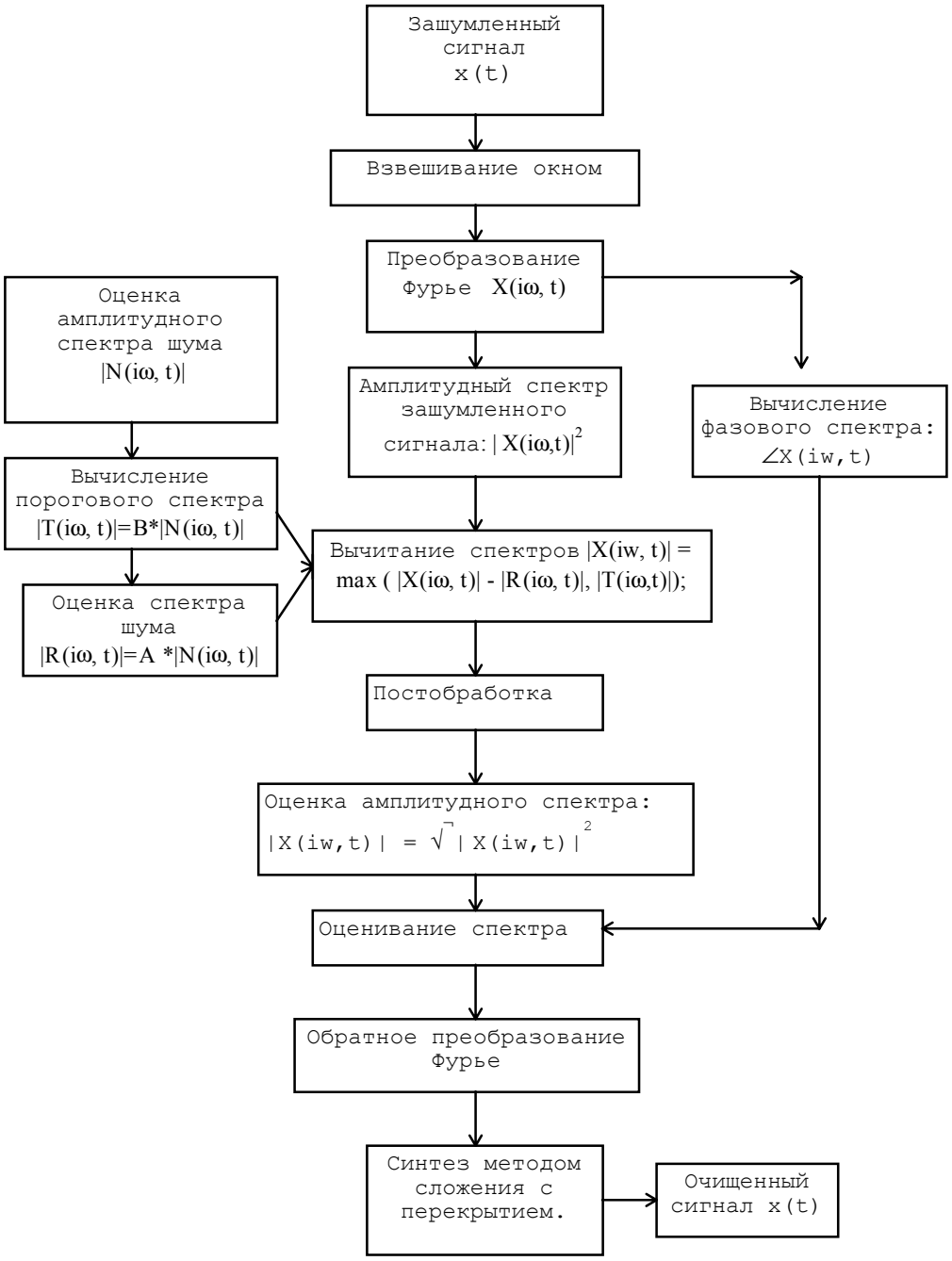


Рис 6.1. Алгоритм вычитания амплитудных спектров.

Исследования качества и разборчивости речи, получаемой в результате применения описанной методики, показали {Sondhi, 1982], что в тех случаях, когда шум или помеха имеют стационарный (или квазистационарный) характер и их

спектр имеет гармоническую структуру, достигается значительное на слух повышение как качества так и разборчивости речи. Однако, в случае шумов с быстроизменяющимися спектральными характеристиками такая обработка малоэффективна.

По мнению аудиторов, такая речь звучит чище и приятнее (несмотря на наличие характерных эффектов обработки - так называемых “музыкальных тонов”, заключающихся в случайных кратковременных выбросах в спектре обработанного сигнала) чем до обработки, однако заметного повышения разборчивости в случае аддитивных широкополосных шумов не происходит, хотя отношение сигнал/шум повышается на 3 - 6 дб. [Ноу, 1983].

В целом, методы, основанные на вычитании спектров считаются одними из лучших - они приводят к удовлетворительным результатам обработки и не требуют больших вычислительных ресурсов. Что же касается музыкальных тонов, которые существенно ухудшают восприятие обработанного сигнала, то для их подавления разработаны различные алгоритмы, основанные на эмпирических и эвристических соображениях [Boll 1979, Sondhi, 1982].

К классу методов, основанных на оценке спектральных характеристик шума, относятся также методы коррекции спектра речевого сигнала, основанные на винеровской фильтрации [Овчинникова, 1977].

В этих методах зашумленный речевой сигнал z(n) фильтруется фильтром с частотной характеристикой, рассчитанной из условия минимизации среднеквадратической ошибки фильтрации, то есть, если H(iw) - частотная характеристика такого фильтра, то

$$H(iw) = P_{xx}(iw) / (P_{xx}(iw) + P_{vv}(iw)), \qquad (6.2)$$

где $P_{xx}(iw)$ - спектральная плотность мощности сигнала
$P_{vv}(iw)$ - спектральная плотность мощности шума

В реальных условиях применения частотную характеристику (6.2) аппроксимируют как:

$$H(iw) = |X(iw)2| / (|X(iw)2| + |V(iw)2|),$$ где

|X(iw)2| и |V(iw)2| - усредненные квадраты амплитудных спектров сигнала x(n) и шума v(n), соответственно, причем оценка величин |X(iw)2| и |V(iw)2|осуществляется так же как и в методе вычитания амплитудных спектров.

Как было упомянуто, одним из основных недостатков спектрального вычитания является наличие артифактов в обработанном сигнале. Музыкальные тона существенно ухудшают качество сигнала, поэтому неудивительно, что одним из приоритетных направлений в развитии этого подхода стало создание пост-процессоров, снижающих эффект музыкальных тонов без дальнейших искажений в сигнале.

В большинстве случаев речь идет о пост-обработке сигнала в спектрально-временной области. Идея состоит в том, что музыкальные тона в спектрально-временном представлении представляют собой локальные (во времени и по частоте) спектральные максимумы, которые, как правило, можно найти из полу-эвристических соображений. В работе [Whipple, 1994] поиск спектральных

максимумов, соответствующих музыкальным тонам осуществляется методами обработки изображений.

Анализируя проводимые авторами упомянутых работ результаты, характеризующие эффективность очистки сигнала от шума, следует отметить, что практически во всех случаях при использовании методик типа вычитания спектров в качестве характристики работоспособности алгоритма используются меры качества звучания сигнала или обьективные меры типа усредненных евклидовых расстояний между отфильтрованным сигналом и незашумленным сигналом (предполагается, что он доступен).

Разборчивость в результате применения описанных методов к сигналу, содержащему случайный аддитивный шум (типа белого или розового, или нестационарные помехи с широкополосным спектром) по видимому, не изменяется.

Исключения к только что сформулированному тезису составили измерения, проведенные на методе вычитания кепстров (то есть вместо спектра сигнала использовался кепстр, а вместо логарифма при оценке спектральных амплитуд использовали кубический корень), когда для речевого сигнала с аддитивным шумом (двигатели самолета F-16) наблюдалась не только улучшаение качества, но и разборчивости речи.

Кроме того, при сравнительных испытаниях методов фильтрации речевых сигналов [Gualtiero, 1984] реализованных как препроцессоры на входе ЛПК-вокодера, в результате артикуляционных испытаний для метода вычитания амплитудных спектров было зарегистрировано повышение разборчивости передаваемой речи на 22.8% при исходном отношении сигнал/шум 2 дб.

Интересная модификация вычитания спектров - вычитание сигналов во временной области предложена в работе [Arlsan, 1997]. Там же продемонстрировано, что этот метод, в отличие от вычитания амплитудных спектров и винеровской фильтрации дает хорошие результаты (в том числе, повышает как качество так и разборчивость сигнала при исходных соотношениях сигнал/шум -10дб и -30 дб) на нестационарных и коррелированных с речевым сигналов шумах типа мешающего диктора, гармонических помехах, одновременном фоновом разговоре (бормотании) нескольких дикторов.

Метод предполагает доступ к передаваемому незашумленному сигналу на передающей стороне и основан на технике добавления нулевых отсчетов в передаваемый речевой сигнал (то есть сигнал квантуется с удвоенной частотой, причем каждый второй отсчет - нулевой). На принимающей стороне характеристики шума оцениваются исходя из величин дополнительных (нулевых в начале передачи) отсчетов. Поскольку большинство практически встречающихся шумов (например все речеподобные сигналы) коррелированы на интервалах между соседними отсчетами, оценка шума выполненная для дополнительных отсчетов вполне пригодна для фильтрации сигнала.

Очевидными недостатками предлагаемой методики является условие доступа к данным на передающей стороне и удвоение скорости передачи данных, поэтому область практического применения алгоритма существенно сужена.

1.7. Метод оценивания минимальной средне квадратической ошибки.

Описываемый алгоритм (оригинальное название Minimum Mean-Square Error estimation) впервые был предложен в работе [Ephraim, 1984]. Как и вычитание спектров алгоритм основан на оценке амплитудного спектра сигнала и общая блок-схема алгоритма в целом соответствует рисунку (6.1). Среди других методов фильтрации, предполагающих наличие только одного микрофона, алгоритмы, основанные на минимуме среднеквадратической ошибки являются одними из наиболее полезных. Их использование приводит к значительному сокращению уровня шума в сигнале без внесения остаточных искажений типа музыкальных тонов[Cappe, 1994].

Допустим, что s(t) и b(t) обозначают, соответственно, речевой сигнал и аддитивный шум, а y(t)- наблюдаемый сигнал, то есть y(t) = s(t)+b(t). Пусть также S(iw), B(iw) и Y(iw) обозначают соответственно спектральные компоненты речевого сигнала, шума и зашумленного сигнала, оцененные на интервале анализа, на котором предполагается квазистационарность речевого сигнала.

Оцениватель амплитудного спектра сигнала по минимуму среднеквадратичной ошибки (MMSE) определяется из двух следующих (апостерирного и априорного) локальных отношений сигнал/шум:

$$SNR_{post}(f) = \frac{|Y(iw)|^2}{E\{|B(iw)|^2\}}$$

и

$$SNR_{prio}(f) = \frac{E\{|S(iw)|^2\}}{E\{|B(iw)|^2\}}$$

Передаточная функция шумоподавителя определяется формулой:

$$N(iw) = \frac{\Lambda(iw)}{1+\Lambda(iw)} N_0(iw)$$

где Λ(iw) - это обобщенное отношение правдоподобия, которое принимает во внимание величину неопределенности присутствия полезного сигнала (речи) в зашумленном сигнале. Отношение правдоподобия оценивается как

$$\Lambda(f) = \frac{1-q(iw)}{q(iw)} \frac{\exp\left\{\frac{SNR_{post}(iw)SNR_{pri}(iw)}{1+SNR_{prio}(iw)}\right\}}{1+SNR_{prio}(iw)},$$

а “обычный” коэффициент подавления шума N0(iw) на частоте iw равен

$$N_0(iw) = \frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{SNR_{post}(iw)}\left(\frac{SNR_{prio}(iw)}{1+SNR_{prio}(iw)}\right)} \times F\left[SNR_{post}(iw)\left(\frac{SNR_{prio}(iw)}{1+SNR_{prio}(iw)}\right)\right]$$

где F[x]=exp(-x/2)[(1+x)I_0(x/2) + xI_1(x/2)], I_0(.), I_1(.) - обозначают модифицированные функции Бесселя нулевого и первого порядка, а q(iw) (0<= q(iw) <= 1)- вероятность отсутствия полезного сигнала в соответствующей спектральной компоненте.

Приведенные формулы выведены при неявном предположении, что априорное отношение сигнал/шум известно. В реальных условиях, однако, этот параметр априори неизвестен, при этом предлагается его оценивать соотношением:

$$SNR_{prio}(t,iw) = (1-\beta)P\left[SNR_{post}(t,iw)-1\right] + \beta\frac{\left|S(t-1,iw)\right|^2}{P_B(iw)} \qquad (1)$$

где t - индекс времени и P[] - обозначает операцию клиппирования полуволны. Параметр β выбирается из эмпирических соображений и обычно β = 0.98.

В недавно проведенных исследованиях[Scalart, 1996] утверждается, что в значительной мере превосходство метода оценивания минимальной среднеквадратической ошибки над методиками типа Винеровской фильтрации или вычитания амплитудных спектров связано именно с введением априорной оценки сигнал/шум в каждой спектральной полосе. В связи с этим, были предложены модификации стандартных подходов (винеровской фильтрации, вычитания амплитудных спектров и оценок максимального правдоподобия) использующие априорные отношения сигнал/шум типа (7.1), что привело к существенному улучшению результатов фильтрации.

1.8. Методы, основанные на искусственных нейронных сетях.

Разработка аппарата искусственных нейронных сетей привела к появлению нового типа алгоритмов для для обработки зашумленных речевых сигналов, основанных на использовании моделей нейронных сетей. Подобных работ пока еще немного и получаемые в этом направлении результаты пока хуже, чем достигаемые более традиционными методами, однако, поскольку нейронные сети обладают потенциально огромными возможностями по непараметрическому моделированию различных типов плотностей, можно ожидать в этом направлении появления мощных алгоритмов фильтрации.

Исследование свойств многослойного персептрона как нелинейного фильтра во временной области выполнено в [Le, 1996]. В качестве помехи рассматривались аддитивный гауссовский шум и нелиенйный шум, моделирующий артифакты низкоскоростного CELP кодера. Персептрон обучался на зашумленном сигнала, роль сигнала-учителя выполнял чистый сигнал (доступный на этапе

обучения). В качестве речевого материала использовались записи гласного “e” (120 записей от 40 дикторов).

Наилучшие результаты были продемонстрированы на трехслойном персептроне (на каждом слое по 60 нейронов): улучшение отншения сигнал/шум составило 3 дб для шума кодека и 6 дб для белого шуса при начальном уровне сигнал/шум = 7.4 дб. Эти результаты, показывают, что пока выигрыша от использования многослойного персептрона по сравнению с многими стандартными методиками нет (отметим. что в более ранней работе [Fechner, 1993] утверждалось, что многослойный персептрон с успехом может быть использован для фильтрации гауссовского шума.

1.9. Методы, основанные на использовании закономерностей восприятия речи человеком.

Некоторые алгоритмы анализа речевых сигналов основаны на использовании свойств слухового анализатора человека. В основе развития этого класса методов лежит утверждение, что анализ речи, основанный на модели слуха человека, будет возможно более успешным, чем анализ, основанный на довольно абстрактных моделях речеобразования или статистических марковских моделях. В частности, утверждается, что системы цифровой обработки речевых сигналов, построенные на таких принципах, будут устойчивы по отношению к мешающему шуму и дикторам.

В пионерской работе [Laughans, 1978], которая во многом обусловила появление такого алгоритма предобработки речевого сигнала как RASTA, описано экспериментальное исследование новой методики обработки зашумленных речевых сигналов, которая основана на гипотезе о том, то слух человека наиболее чувствителен к модуляциям в спектральной огибающей сигнала с частотой 2-3 герца.

Если рассматривать амплитуду кратковременного преобразования Фурье речевого сигнала |X(n, iw)| как функцию времени n при фиксированной частоте w, то наиболее важными для восприятия модуляциями |X(n, iw)| оказываются те, которые имеют частоту 2 - 3 герца. Поскольку аддитивный шум уменьшает модуляционную глубину |X(n,iw)|, то один из способов повышения разборчивости речи состоит в том, чтобы искусственно увеличить модуляционную глубину сигнала в определенном диапазоне частот. Экспериментальная проверка этой методики показала, что существеного увеличения разборчивости речевого сигнала можно добиться путем увеличения модуляционной глубины речевого сигнала до зашумления. Применение же метода к зашумленному сигналу разборчивости речи не повысило.

Одним из наиболее специфических механизмов слуха является эффект маскировки в слуховой системе, который во многом определяет свойства помехоустойчивости, присущие слуху.

Поскольку традиционные методы фильтрации типа вычитания спектров или оптимального среднеквадратического оценивания сопровождаются наличием пост-процессорных искажений сигнала, основным направлением использования моделей эффекта маскировки является их использование в качестве дополнительных блоков, имитирующих маскировку слабых сигналов более сильными в критических полосах слуха.

Результаты применения модели слуха для усовершенствования алгоритма вычитания спектров рассмотрены в работе [Virag, 1996], а соответствующие усовершенствования для модели оптимального среднеквадратического оценивания предложены в работе [Azirani, 1996].

В обоих случаях использована модель обработки сигнала в слуховой системе, предложенная в [Johnston, 1988]. В экспериментах отмечалось некоторое улучшение качества обработанного сигнала, особенно при низких (-7 дб), для[Virag, 1996], и высоких (10-20 дб) - для метода [Azirani, 1996] отношениях сигнал/шум.

В работе [Curtis 1978] описано исследование метода повышения разборчивости зашумленной речи, котрый основан на том что речевой сигнал сначала подвергался высоночастотной фильтрации таким образом, чтобы ослабить первую форманту, тем самым повышая удельный вес высших формант в спектре речевого сигнала. Далее отфильтрованный речевой сигнал подвергается клиппированию. Авторы утверждают, что операция клиппирования увеличивает амплитуду речевой влоны на участках, которые соответствуют важным для восприятия согласным по отношению к амплитуде гласных звуков.

Одним из часто применяемых на практике методов обработки зашумленных речевых сигналов является фильтрация этих сигналов полосовыми или режекторными фильтрами. Если спектральные характеристики шума известны заранее, то этот метод скорее можно отнести к классу методов, использующих априорные сведения о спектральных характеристиках шума.

В работе [Hanson, 1983] описано экспериментальное исследование методики улучшения разборчивости речи в условиях мешающего диктора. Восстановление речи осуществлялось фильтром по спектральной огибающей зашумленного речевого сигнала и траектории частоты основного тона голоса выделяемого диктора. Этот метод привел к заметному улучшению качества речевого сигнала, а то мнению аудиторов, знакомых с содержанием речевого сообщения, также улучшил разборчивсть речи. Однако более измерения разборчивости показали, что реально разборчивость речи в результате применения этой методики понизилась.

Потенциально весьма многообещающие результаты получены при испытаниях систем анализа и обработки зашумленных речевых сигналов, построенных на представлении речевых сигналов с помощью так называемых волновых функциях-вейвлетах[Teolis, 1994, Pinter, 1996].

Волновой (вейвлетный, wavelet) анализ речи широко применяется при анализе и обработке речевых сигналов последние 10-15 лет.

Семейство волновых функций описыватся соотношением:

$$w_{a,b}(t) = \frac{1}{\sqrt{a}} w(\frac{t-b}{a}); a > 0, b \in R$$

где t обозначает время, параметры b и a регулируют перенос (трансляцию) по времени и сжатие/растяжение.

Так же как и в других базисах, эти ортогональные функции используются для декомпозиции речевого сигнала в сумму элементарных сигналов.

Для перцептивных волновых функций[Pinter, 1996] порождающая волновая базисная функция w(t) обладает преобразованием Фурье вида:

$$W_0^c(b) = \begin{cases} \cos^2(\frac{\pi}{2} b), -1 \le b \le 1 \\ \hline 0, else \end{cases}$$

Семейство волновых функций $W_k(b)$: $W_k(b) = W_0(b-b_1-k\Delta b)$, k=0,1,…,K-1 выбирается так, что $c\sum_{k=0}^{K-1} W_k(b) = 1$, где $0<b_1<=b<=b_2$ и $[b_1,b_2]$ - анализируемый интервал в частотной области, который выбирается на основе частот критических полосок слухового анализатора.

Основная идея предложенного метода фильтрации навеяна свойствами робастности по отношению к помехам, которыми обладает человеческая аудиторная система и заключется в моделировании уже упомянутого эффекта маскировки, когда слуховая система суммирует сигналы в критических полосках и, при одновременном присутствии двух сигналов разного уровня, сигнал с более высоким уровнем подавляет сигнал меньшего уровня.

В данном случае, если m-ый сегмент (речевой сигнал обрабатывается посегментно, с перекрытием 1:1) зашумленного сигнала $z^m(t)$:

$$z^m(t) = x^m(t) + v^m(t),$$

где $x^m(t)$ - незашумленный сигнал, $v^m(t)$- аддитивный шум, то оценка обработанного сигнала $x\hat{}^m(t)$ на m-ом сегменте анализа вычисляется как:

$x\hat{}^m(t) = \sum_{k=0}^{K-1} x^m(t) T_k(e_k^m, e_k, \sigma_k)$, где коэффициент

$$T_k(e_k^m, e_k, \sigma_k) = \frac{1}{1 + \exp\left\{-1/2[e_k^m - (e_k + \frac{\sigma_k}{2})]\right\}},$$

а e_m^k - энергия m-го сегмента в k-ой полосе частот, e_m^k - усредненная энергия и σ_k - дисперсия усредненной энергии.

Приведенная формула дает основание рассматривать изложенный метод как модификацию алгоритма спектрального вычитания, выполняемую в спектральных полосах, определяемых волновыми функцияими, в данном случае - аппроксимацией критических полос слухового анализатора.

Характеристики метода оценивались субьективными методами, посредством прослушивания записей исходного и обработанного сигнала. При этом аудитор постепенно добавлял к обработанному речевому сигналу шум до тех пор, пока перцептивно получаемый сигнал не становился таким же шумным как и исходный зашумленный сигнал. Уровень шума, который необходимо было добавить к обработанному сигналу (чтобы получить такой же по качеству сигнал, как и до обработки), являлся мерой качества обработки.

Тестирование выполнялось на шести типичных для систем речевой технологии шумах: гауссовском шуме, транспортном шуме (запись с уличного перекрестка), кратковременных выключениях описанных шумов (40 прерываний в секунду), шума неречевого происхождения, записанного со старого фонографа и речеподобного шума, сгенерированного как сумма пяти высказывания, взятых из речевой базы данных. Для каждого типа шума делалось восемь тестовых записей с различным уровнем шума.

Применение изложенной методики обеспечило выигрыш около 26 децибел (уровень субьективно добавленного шума) в случае гауссовского шума, 18 децибел в случае речеподобной помехи и 20-22 дб. в остальных случаях. В результате испытаний на собранной базе данных автор утверждает, что выигрыш при использовании метода составляет не менее 18 дб. Эти результаты представляются достаточно высокими и заслуживающими внимания, хотя не совсем понятно как можно пересчитать эти цифры в более привычную методику оценки выигрыша в отношении сигнал/шум.

Как было отмечено, повышение качества и комфортности звучания очищщегнного сигнала вовсе не означает улучшения его разборчивости. Одной из причиной такого положения является то, что большинство из предложенных методов ориентированы на подавление шума, а не на выделение полезного речевого сигнала, что по мнению [Hansen, 1996] не одно и то же. Как результат, подавление шума в переходных сегментах речи и низкочастотных полосах, незначительно и не сопровождается улучшением разборчивости.

Сравнительно недавно [Hansen, 1996] был разработан класс итеративных алгоритмов помехоподавления, ориентированных на лучшее использование характерных свойств речевых сигналов. Результатом явилось существенное улучшение (по результатам субьективных тестов типа попарного сравнения) качества обработки сигнала в условиях помех типа белого и розового шумов.

Исходные алгоритмы были получены на базе моделирования речевого сигнала s(n) моделью авторегрессии с локально постоянными параметрами a(t):

$$s(n) = a^T s_{n-1}^{n-p} + gw(n),$$

где s(n), w(n) - отсчеты наблюдаемого речевого сигнала и возбуждения, соответственно, а a - вектор коэффициентов модели.

Для оценки параметров модели сигнала была предложена итерационная процедура оценивания по методу максимума апостериорной вероятности, которая максимизировала $\max p(a, g, s_{n-1}^{n-p} \mid y_{n-1}^{n-p})$, где y - означает вектор отсчетов

наблюдаемого сигнала. По вычисленным параметрам а оценивался спектр речевого сигнала:

$$P_s(w) = \frac{g^2}{|1 - \sum_{k=1}^{p} a_k e^{-jkw}|^2}$$

Спектр шума $P_D(w)$ оценивался по паузам, а фильрация выполнялась фильтром Винера с характеристикой

$$H(w) = \frac{P_s(w)}{P_d(w) + P_s(w)}$$

Существенным обстоятельством алгоритма является то, что новый подход существенно использовал допустимые ограничения на последовательность кратковременных амплитудных спектров сигнала с тем, чтобы получить более правдопободные траектории формант и сократить дребезг полюсов модели при переходе от одного кадра анализа к следующему. Внутри кадра анализа ограничения накладывались на величины автокорреляционных коэффициентов модели, а между кадрами ограничения накладывалить на параметры линейных спектральных пар.

При исходном отношении сигнал/шум = -5 дб результаты попарных сравнений с алгоритмом вычитания спектров (в оригинальной форме [Boll, 1979], показали существенное преимущество представленного подхода, при незначительном увеличении вычислительных затрат.

Дополнительные ограничения, вытекающие из свойств речевого сигнала использованы также и в методе [Hansen, 1994]. Здесь дополнительно использована процедура, которая обеспечивает адаптивное выделение границ сигнала и затем использует т.н. морфологические спектральные ограничения. Этот метод работает в частотной области, используя, аналогично вычитанию спектров, переменный коэффициент подавления в каждой спектральной полосе. Соотношение для оценки отфильтрованного сигнала имет вид:

$$\overline{S}(jw) = \left\{ \frac{1}{M_1 + M_2 + 1} \sum_{i=-M_1}^{M_2} |S(jw)^{\beta} - \alpha(i) E\left[|D(jw)|^{\beta} \right] \right\}^{\frac{1}{\beta}} e^{j\theta(jw)}$$

где M1, M2 - обозначают число сегментов (кадров) анализа из прошлого и будущего, соответственно, которые используются для усреднения амплитудного спектра, D(jw) - оценка спектра шума, S(jw) - спектр зашумленного сигнала, α(i) - модифицируемый коэффициент, регулирующий уровень вычитания сигнала (этот коэффициент устанавливается на основании значений границ, выделяемых адаптивным алгоритмом поиска границ сегментов), β - степенной коэффициент, также регулирующий степень подавления. Величины обоих коэффициентов α(i) и β определяются совместно для всех комбинций типов речевых сегментов (звонкий-глухой-переходный).

Для минимизации остаточных эффектов (типа музыкальных тонов) использовано множество морфологических операторов, выполняемых над сигналом в спектрально-временной системе координат (то есть над спектрограммой). Морфологические операторы применяются с учетом заранее предопределенных структурных элементов. Базовыми морфологическими операциями являются операции типа распространения, эрозии, открытиия и закрытия.

Утверждается, что использование морфологических фильтров (полученных как комбинации морфологических операций) в спектрально-временной области приводит к существенно лучшим результатам, чем амплитудное усреднение сигнала в вычитании спектров, особенно при обработке переходных сетментов сигнала.

Результаты испытаний (SNR в диапазоне от -5 до +10 db) подтверждают превосходство метода над стандартным вычитанием спектров и Винеровской фильтрацией.

Выводы

Обзор методов повышения качества и разборчивости зашумленных речевых сигналов показывает, что существует много различных подходов к обработке зашумленной речи. Такое разнообразие методов обусловлено как важностью проблемы так и отсутствием достаточно надежных методов ее решения.

Объективное сравнение этих методов и выбор наиболее приемлемых сделать весьма затруднительно, так как перед системами коррекции речевых сигналов ставятся различные задачи. Например, можно в качестве главного критерия использовать повышение разборчивости речи, допуская при этом возможность искажений в тембре голоса или появление артифактов в виде структурированного шума. Можно поставить целью понижение утомляемости аудитора или сохранение натуральности голоса диктора, что достигается в основном за счет повышения качества речевого сигнала. Наконец, могут быть известны заранее важные априорные сведения, например тип или параметры шума, характеристики голоса диктора, наконец, гипотезы о произносимом тексте, что также может определяющим образом повлиять на выбор метода фильтрации.

Важно отметить, что универсальных методов обработки, которые одинаково хорошо боролись бы с существенно нестационарными и стационарными, аддитивными и мультипликативными шумами, существенно повышали бы качество и одновременно разборчивость речи, сейчас нет, и возможно не будет. Как типичная (за редкими, указаными в обзоре исключениями, наблюдается обратная тенденция: если сравнивать системы обработки зашумленной речи по двум показателям - повышению качества звучания речевых сигналов и повышению разборчивости, то системы, повышающие качество и натуральность звучания, скорее всего снижают разборчивость и наоборот, повышение разборчивости приводит к понижению качества и натуральности звучания. Поэтому, многие из названных методов фильтрации нужно рассматривать как взаимнодополняющие, и в идеальном случае нужно иметь библиотеку из нескольких методов фильтрации.

Рассматривая последние тенденции в области обработки зашумленных сигналов, следует особенно выделить высокие результаты, полученные за счет использования математических моделей речевых сигналов (авторегрессионные скрытые марковские модели), а также использование нейроподобных структур (многослойный персептрон) для фильтрации аддитивных стационарных шумов, хотя первые результаты в этом направлении проигрывают более традиционным методам типа минимальной среднеквадратической оценки.

2.0 Программное обеспечение для обработки зашумленных речевых сигналов

В разделе рассмотрены алгоритмы, команды и режимы программного обеспечения для цифровой обработки зашумленных речевых сигналов. На основе классификации основных типов помех были выбран набор алгоритмов которые обеспечивают обработку наиболее распространенных типов помех. Разработка системы преследовало двоякую цель. С одной стороны предполагалось создать среду для разработки, тестирования и сравнения различных подходов к подавлению помех и предобработке речевого сигнала. С другой стороны, в результатв работы был создан работоспособный инструментарий для очистки зашумленных речевых сигналов.

Программы фильтрации реализуют следующие типы алгоритмов:

- подавления аддитивного квазистационарного шума методов вычитания амплитудных спектров
- подавления широкополосного аддитивного шума методом оценки минимальной среднеквадратической ошибки
- подавение периодической помехи
- подавление аддитивного фонового шума
- подавление импульсных помех
- подавление эхо-сигнала и реверберации
- синтез цифровых фильтров и цифровую фильтрацию

Программы обработки функционируют в составе и под управлением компьютеризованной речевой лаборатории (Computerized Speech Laboratory-CSL4300B) производства фирмы KAY Elemetrics Corp. Выбор среды обусловлен тем, что в CSL реализованы все необходимые базисные функции по вводу-выводу и записи речевого сигнала, программное обеспечение открыто для наращивания его новыми модулями, командный язык системы достаточно удобен, расширяем и позволяет удобно организовать пакетную обработку. Наконец, наличие процессоров цифровой обработки сигналов и средств их программирования позволяет перенести основные вычислительные операции в аппаратуру CSL, вести обработку в реальном масштабе времени и освободить центральный процессор компьютера. Ниже рассмотрены более подробно основные программы входящие в состав системы.

2.1. Подавление аддитивного квазистационарного шума методом вычитания амплитудных спектров

Программа спектрального вычитания предназначена для удаления нежелательного аддитивного квазистацинарного шума из аудио записей.

Эта программа наиболее приспособлена для обработки многих типов шумов, спектральные характеристики которых изменяются достаточно медленно в течение интервала наблюдения. В качестве типичного примера можно привести шумы кондиционеров, видеокамеры, автотрансформаторов и усилителей. Поведение алгоритма контролируется набором параметров, включая предварительно измеренные характеристики шума. Для того, чтобы пользлватель смог полностью использовать все возможности, заложенные в алгоритме, предусмотрен диалоговый режим работы, когда выбор и изменение параметров и режимов обработки выполняется в ходе обработки, причем оператор контролирует качество работы метода прослушиванием обработанного сигнала.

Алгоритм вычитания амплитудных спектров

Программа реализует алгоритм фильтрации, в соответствии с соотношением:

$$|S(t,iw)|^2 = \begin{cases} |X_i(t,iw)|^2 - A(t)|N(t,iw)|^2, & \text{если } |X_i(t,iw)|^2 \geq (A(t)+B)|N(t,iw)|^2 \\ B|N(t,iw)|^2, & \text{в противном} \end{cases}$$

Блок-схема алгоритма вычитания амплитудных спектров привенена на следующем рисунке.

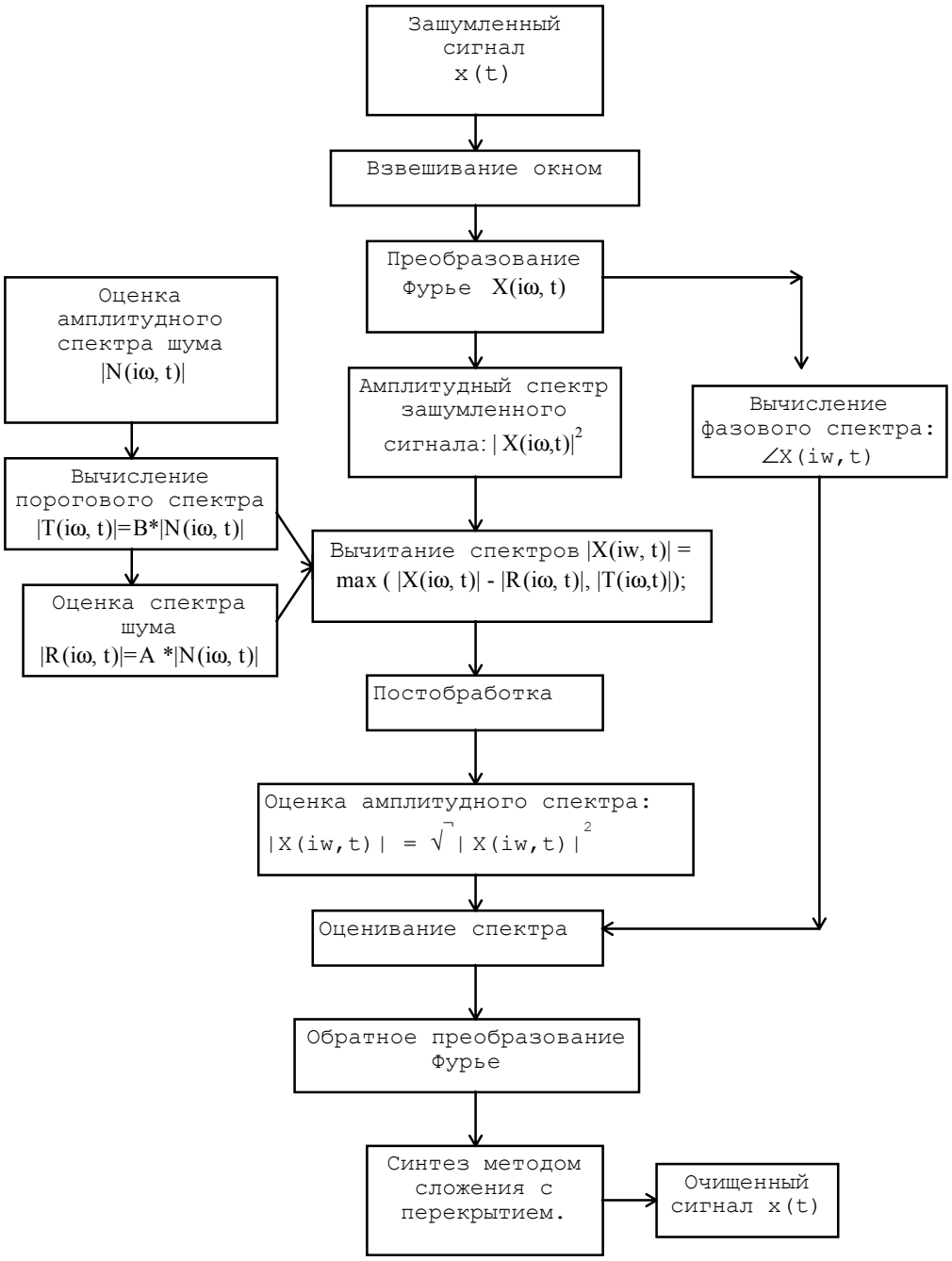


Рис 2.1. Алгоритм вычитания амплитудных спектров

Взаимодействие пользователя с программой осуществляется в соответствии со следующей диаграммой

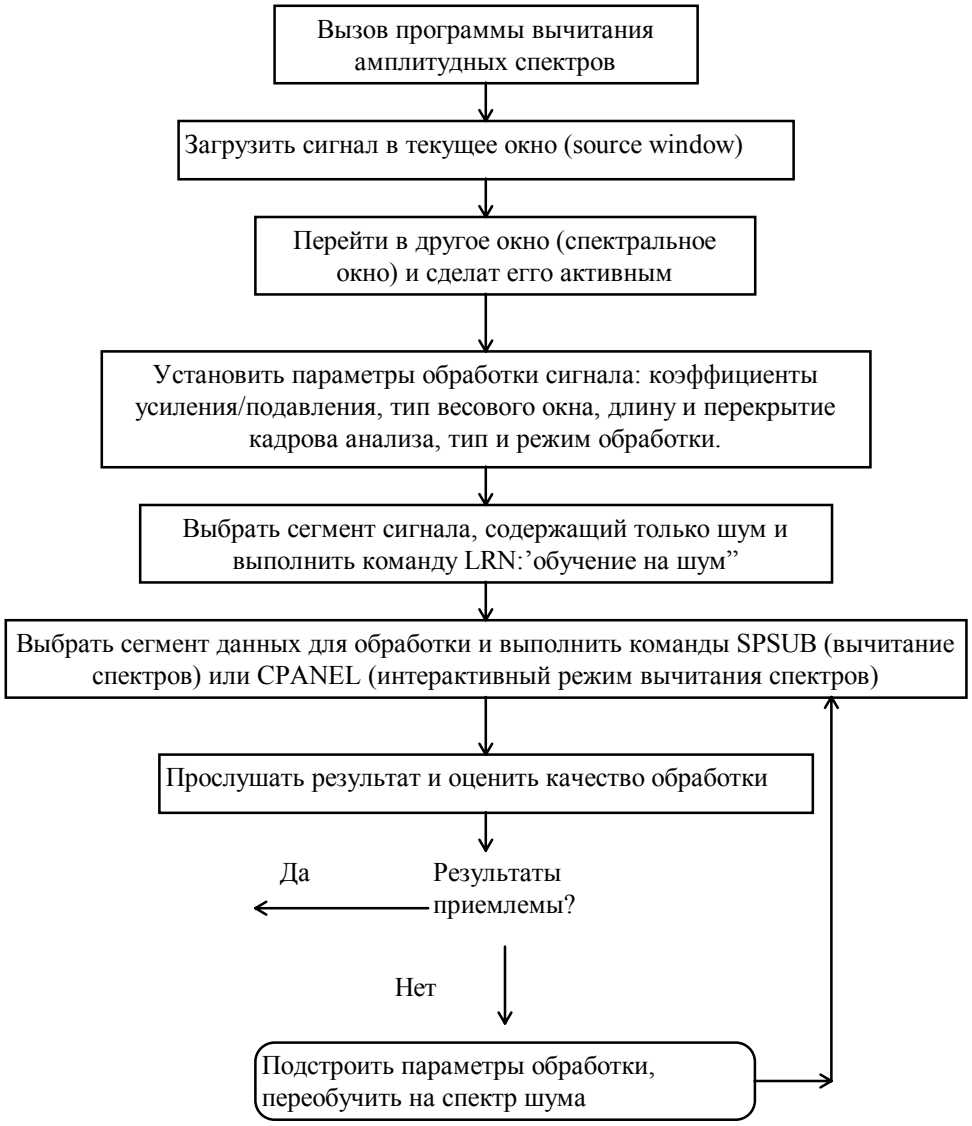


Рис 2.2. Сценарий использования вычитания амплитудных спектров

Как вытекает из рисунка, сценарий использования вычитания спектров состоит из трех главных частей:

1. Выбора параметров обработки (если требуется),
2. Выполнения обучения на спектр шума
3. Выполнения процедуры вычитания амплитудных спектров.

При этом оператор может регулировать процесс обработки данных путем перманентной подстройки параметров во время обработки. Это осуществляется с помощью режима Контрольной панели

Установка параметров осуществляется через командуы "SET parameter" ядра CSL.

Команды установки параметров вычитания спектров

Для того, чтобы использовать команду SET установки параметров вычитания амплитудных спектров, необходимо выполнить следующие операции:

Мышь. Выберите на член **SPSUP** главного меню программы. Затем выберите член **OPTION** в раскрывающемся подменю и наконец, выберите член, указывающий на параметр обработки, значение которого предполагается изменить. Подменю SPSUB показаны ниже на рисунке 2.3.

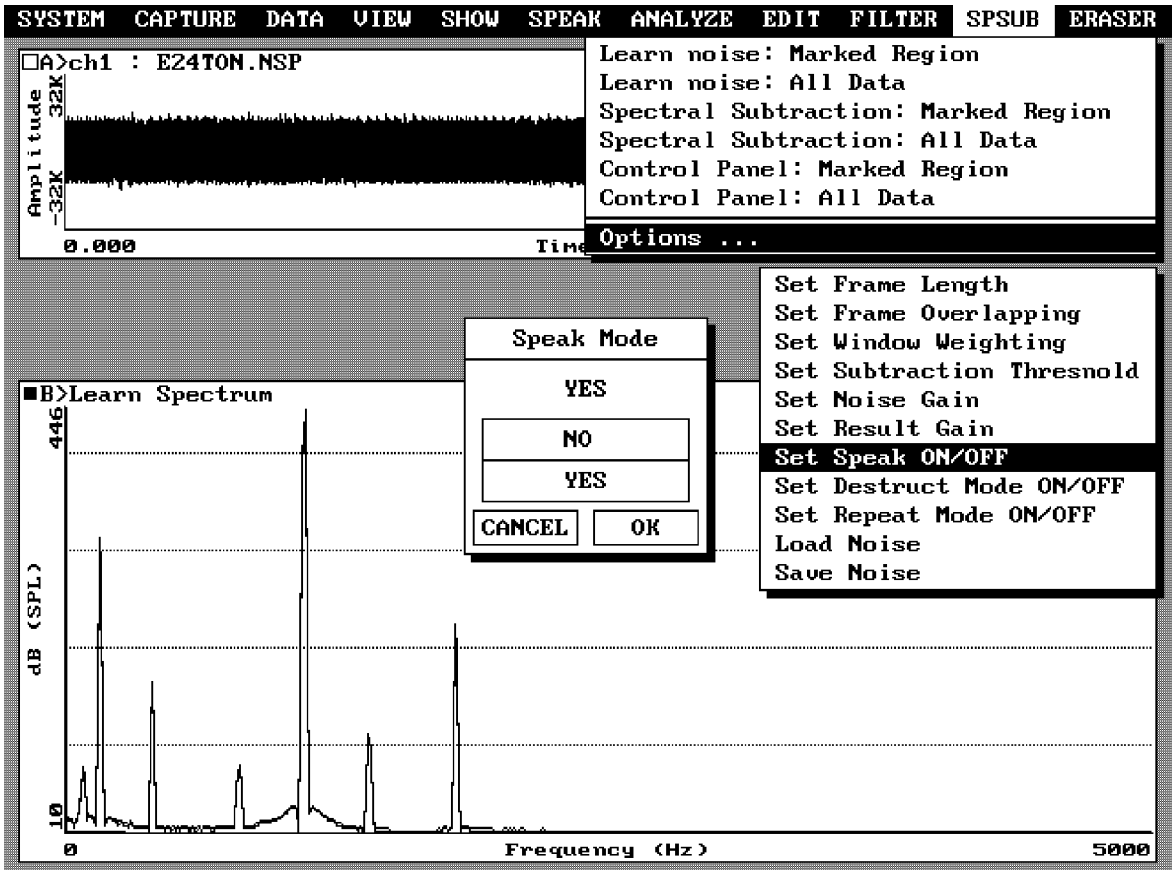


Рис 2.3. Подменю программы вычитания амплитудных спектров

Клавиатура. Введите команду SET для выбранного параметра и его новое значение в соответствии со следующей таблицей

Член меню	Реализуемая команда	Обьяснение
Frame Length	SET Frame Length [points]	Установка длины кадра анализа. Это длина блока данных для процедуры быстрого преобразования Фурье. По умолчанию

		значение этого параметра выбирается равным 1024.
Frame Overlapping	SET Frame Overlapping [2|4]	Устанавливает перекрытие соседних кадров анализа. Допустимые значения это1, 2, 4 и 8. Значение 2 означает, что соседние кадры анализа перекрываются наполовину. Вообще говоря, чем больше перекрытие, тем лучшие результаты можно ожидать, однако, большее перекрытие кадров означает увеличение вычислительной загрузки процессора, так, что режим реального времени может не быть осуществимым. По умолчанию перекрытие равно 2.
Window	SET window [window_type]	Выбор типа весового окна. По умолчагию используется окно Хемминга
Speak [On/Off]	SET Speak [On/Off]	Устанавливает или сбрасывает режим синхронного прослушивания обрабатываемого сигнала во время обработки. Таким образом можно контролировать процесс фильтрации.
Repeat [On/Off]	SET Repeat [On/Off]	Зацикливание процесса обработки. Оператор может установить разрушающий (**Destruct Off**) или неразрушающий (**Destruct**

		On) режим работы. Означает, что данные обрабаьываются виртуально - сигнал в буфере обработки реально не изменяется. Цель такого режима - тонкая настройка параметров перед действительной обработкой данных. Если выбран разрушающий режим работы (**Destruct** установлен в **On**) данные в буфере реально изменяются на каждом цикле обработки. Подобный режим хорошо использоват для выполнения нескольких итераций алгоритма на одном и том же материале. Чтобы остановить обработку при режиме зацикливания (**REPEAT=On)** достаточно нажать любую клавишу клавиатуры
	SET Destruct [On/Off]	Режим **On** означает, что обработка разрушает данные в буфере, результат обработки заменяет исходные данные. **Off** означает неразрушающую обработку, когда реально данные остаются неизменными и результат обработки только выводится на прослушивание..
Set Threshold	SET	Устанавливает

	Threshold[value]	значение коэффициента порога спектра B. По умолчанию B = 0.05
Set Estimate	SET Estimate[value]	Устанавливает значение коэффициента усиления оценки спектра A По умолчанию значение A = 0.95
Load	SET Load [file]	Загружает предварительно обученный спектр шума из файла
Save	SET Save [file]	Сохраняет обученный спектр шума в файле

Таблица 2.4. Команды установки параметров вычитания спектров

Обучение параметрам шума

Следующий шаг заключается в обучении алгоритма: измерении спектра спектра мощности шума, котрый будет затем использован для построения фильтра. Обучение выполняется с помощью команды LRN - Обучение спектра шума. Для этого необходимо идентифицировать участок (не менее 250-300 миллисекунд) аудиоматериала, где нет полезных данных, а присутствует (или доминирует) шум. Установить метки в начале и конце выбранного сегмента данных и выполнить команду LRN.

По этой команде программа рассчитает усредненный спектр мощности шума и отобразит его в активном окне. Непосредственно операции реализуются следующим образом.

Mouse. Выберите член **SPSUP** в главном меню программы, затем выберите член **LEARN MARKED SEGMENT** если сегмент содержащий шум отмечен маркерами или **LEARN ALL DATA** если все данные в текущем окне могут быть использованы для обучения параметрам шума.

Keyboard. Введите команду **LRN ? M1 M2** для выполнения процедуры обучения параметрам шума для данных в активном окне, находящихся между метками M1 и M2. Любые типы аргументов для определения границ сегмента, поддерживаемые CSL могуи быть использованы в данном случае.

Обученный спектр мощности будет вычислен и отображен в активном окне в виде двух контуров голубого и красного цвета. Глубой контур соответствует измеренному спектру мощности, а красный - оценке спектра мощности шума, который будет использован при фильтрации сигнала. Если параметр A = 1.0 то оба контура совпадают.

Процесс фильтрации инициируется следующим образом.

Выделяется порция данных в окне исходных данных, далее активизируется спектральное окно, в котором находится оценка спектра шума и запускается процедура вычитания спектров.

Mouse. Выбрать член **SPSUP** в главном меню. Затем выбрать член **PROCESS ALL DATA**, если нужно обработать все данные в окне исхоных данных или **PROCESS MARKED SECTION**, чтобы отфильтровать только выбраную порцию сигнала.

Keyboard. Введите команду **SPSUB ? M1 M2** чтобы обработать все данные, находящиеся между метками или **SPSUB ? 0 *** чтобыto обработать все данных в окне с исходным сигналом. Как и в случае обучения спектру шума, для определения данных, подлежащий фильтрации, можно использовать полный синтаксис командной строки **CSL**.

После ввода команды начинается процесс фильтрации. Если параметр **SPEAK** установлен в **ON**, то поток обрабатываемых данных после фильтрации немедленно посылается на динамики для оперативного контроля. Курсор в активном окне движется, показывая текущий обрабатываемый блок данных. После завершения обработки данные сигнал в окне исходных данных перерисовывается в соответствии результатами фильтрации. При этом следует помнить, что реальные данные изменяются в памяти только если установлен разрушающий режим фильтрации (парамеетр **DESTRUCT** установлен в **YES**)

Приборная панель вычитания спектров

Программа вычитания спектров включает специальный режим приборной панели для того, чтобы позволить оператору выполнять настройку параметров спектрального вычитания в интерактивном режиме, одновременно с обработкой данных.

Такая возможность важна, если у оператора есть достаточно времени, чтобы аккуратно подбирать и настраивать параметры, опираясь при этом на качество сигнала, оцениваемое на слух.

Вызов режима приборной панели осущетвляется следующим образом.

Допустим что окно исходных данных содержит сигнал для обработки и в нем уже выделен сегмент сигнала, предназначенный для фильтрации. Активное окно должно содержать обученный спектр шума..

Mouse. Выберите на главном меню член **SPOUSE**. Затем выберите член **CONTROL PANEL ALL DATA** для обработки всех данных, изображенных в окне или **CONTROL PANEL MARKED SECTION** для обработки выделенного фрагмента.

Keyboard. Введите команду **SSPANEL ? M1 M2 LAN** для обработки указанного фрагмента или коменду **SSPANEL ? 0 *** Б чтобы обработать все данные в окне.

После вызова команды вместо меню программы CSL будет изображено меню приборной доски. Это меню показано на следующем рисунке 2.5.

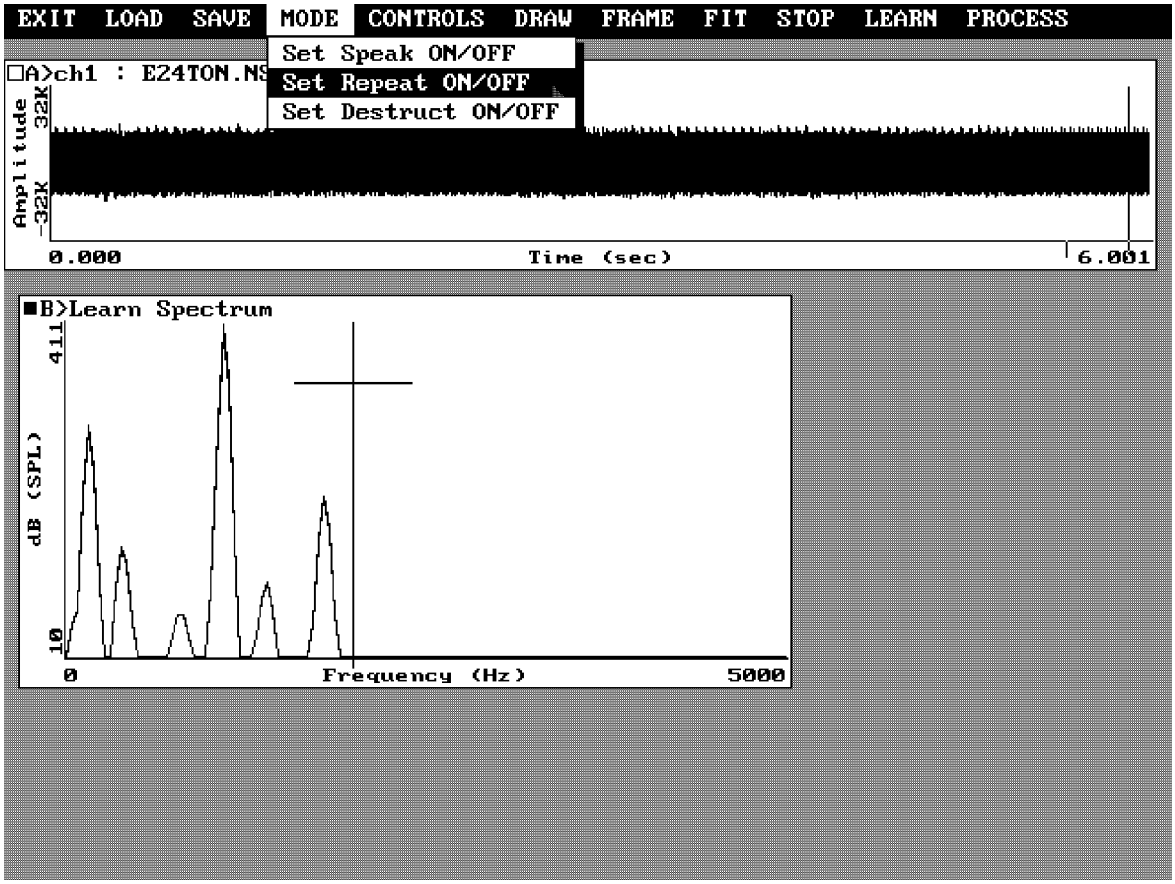


Рис 2.5. Меню режима приборной доски

Сценарий работы пользователя для режима приборной доски соответствует рассмотренному ранее (см. рис 2.2.). Единственное отличие, что в режиме приборной доски опреатор может изменять спектральные характеристики шума и параметрвы обработки непосредственно при обработке данных. Чтобы обеспечить интерактивный режим работы главное меню приборной панели содержит все команды SET установки параметров, показанные ниже на рисунке 2.6. Оператору доступны также пять дополнительных команд, описание которые дано ниже.

Имя члена меню	Обьяснение
Draw	Устанавливает режим рисования, который позврляет рисовать контур спектра шума в активном окне
Fit	Заменяет текущий (то есть используемый в настоящее время) спектр шума на контур, отображенный в активном окне. Например эта команда применяется после того как спектр шума был вручную поправлен или нарисован.
Exit	Выход из режима приборной панели и возврат в главное меню CSL
Process	Запуск обработки данных
Stop	Остановка обработки данных. Эта команда имеет смысл, когда используется несколько итераций обрабо (обработка зациклена установкой режима REPEAT в YES)

Рис 2.6. Дополнительные команды для режима приборной доски

Также как и в обычном режиме, можно установить параметры размера и типа окна, перекрытия блоков и т.п. Чтобы начать обработку данных, выберите в меню **PROCESS**. При этом если параметр **REPEAT** установлен в Y**ES,** то процесс обработки буден зациклен до тех пор, пока не будет введена команда **STOP**.

Замечание. Не все члены меню приборной панели разрешены во время обработки данных (так как одновременная обработка данных, перерисовка окна и выбор членов меню вообще говоря в режиме реального времени выполнимы на достаточно мощных машинах). Неактивные членыменю в период, когда их выполнение не разрешено, изображены серым цветом, в отличен от активных членов меню. Неактивные члены меню не могут быть выбраны в процессе обработки, они не подсвечиваются при выполнении.

Чтобы нарисовать контур спектра выберите член меню **DRAW**, переместите мышь в окно и нажмие левую кнопку. Вмесо курсора мыши появится перекрестье. Установите его в точку, откуда неужно начать коррекцию спектра или его рисование. Нажмите левую кнопку и, держа ее в нажатом состоянии, рисуйте

контур спектра шума как вы его себе придставляете. Перемещением мыши вы можете рисовать контур пока кнопка находится в нажатом состоянии. Результат рисования изображен точечной линией серого цвета. После того как левая кнопка отжата, программа выходит из режима рисования и теперь любое перемещение мыши просто приводит к перемещению курсора. Если нарисованный или скорректированный спектр шума устраивает оператора, он может быть подставлен в программу фильтрации вместо текущего спектра шума. Для этого выбирается член меню **FIT**. После выбора **FIT** спектр шума немедленно перерисовывается и посылается в программу фильтрации в сигнальном процессоре. После окончания фильтрации данные в окне сигнала также будут перерисованы.

2. Программа подавления фонового шума

Программа подавления фонового шума предназначена для удаления из аудиоматериала аддитивной периодической помехи.

Типичные ситуации для применения программы:
- удаление сетевой помехи частотой 50-60 кГц;
- удаление периодических помех с прямоугольными или треугольными импульсами;

Программа подавления фонового шума использует следующий алгоритм очистки сигнала:
Выходной сигнал y(t) связан с входным сигналом x(t) следующим соотношением:

$$y(t) = 0.5 * (x(t)-x(t-T)),$$

где T - текущий или усредненный период сигнала помехи, оцененный с помощью CSL

Программа фильтрации фонового шума приводит к достаточно хорошим результатам при удалении сложных многотональных помех в диапазоне частот 50-500 Гц.

Интрефейс пользователя с программой подавления фонового шума осуществляется в соответствии со следующим сценарием

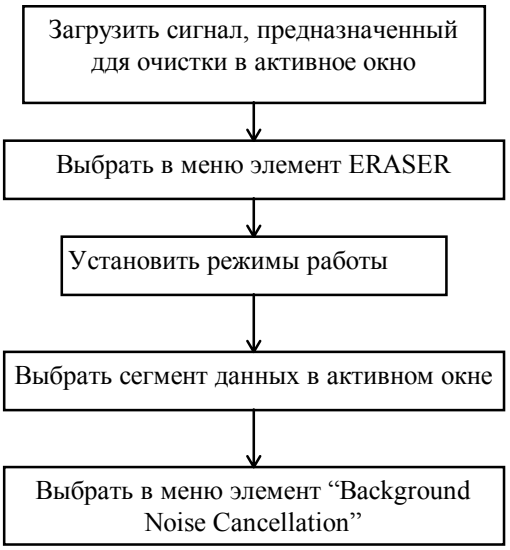


Рис 2.7. Сценарий начала работы с программой подавления фонового шума

Как следует из сцеария, процедура работы с программой состоит из двух отдельных шагов:
1. Установка режимов работы программы,
2. Вызов программы подавления фонового шума.

Задание режимов работы программы

Задание режимов работы программы выполняется через выполнение команды SET установки параметров, котоая может выть выполнена через командную строку или выбором соответствующего элемента меню.

Мышь. Выберите член ERASER в главном меню. На экране появится раскрывающееся подменю. Выберите соответствующий элемент (например **Set Speak ON/OFF** или **Set Destruct Mode ON/OFF**) нажатием кнопки мыш.

Клавиатура. Введите команду SET для выбранного параметра в соответствии со следующей таблицей

Элемент меню	Соответствующая команда	Обьяснение
Speak [On/Off]	SET FonSpeak [On/Off]	В процессе обработки данных выходной сигнал посылается на наушники. Оператор может слышать уже обработанй сигнал “как он есть” в реальном времени.
Destruct [On/Off]	SET SaveResult[On/Off]	“On” означает, что обработка разрушает исходные данные. Если

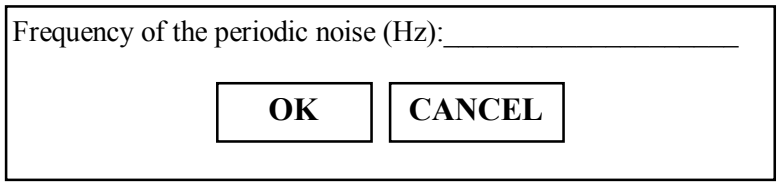
		выбран “Off” , данные в памяти остаются неизменными. В этом случае оператор может много раз проверяить как выбор тех или иных значений параметров обработки отражается на качестве получаемого сигнала. Выбрав лучший вариант можно выполнить окончательную, разрушающую обработку данны.

Fig 2.7. Background Noise Cancellation SET parameter commands

Запуск программы подавления фонового шума

Загрузите аудиоматериал в активное окно. Выберите сегмент данных для обработки.

Выберите элемент **ERASER** в главном меню CSL. Появится подменю, в котором нужно указать, в зависимости от того, обработываются ли все данные в окне или только выбранный сегмент, элемент
REMOVE BACKGROUND : ALL DATA или элемент
REMOVE BACKGROUND : MARKED SECTION. Появится следующий диалог:



Введите значние оцененной частоты помехи в нерцах и нажмите кнопку OK. Программа стартует.

Введите командную строку вида:

GROUND = 0 * или аналогичную, для задания другого типа границ сегмента (используйте обычный синтаксис командных строк CSL). В результате на экране появится приведенный выше диалог.

2.3. Программа подавления импульсных помех

Программа подавления импульсных помех предназначения для подавления резких кратковременных амплитудных искажений в аудиоматериале, вызванных, например царапинами на магнитном носителе или внешней помехой.

Подавление импульсных помех реализуется каскаднам фильтром. Общая идея алгоритма состоит в оценке выходного сигнала y(t) по входному сигналу в соответствии:

$$y(t) = \begin{cases} x(t), \text{ if } |x(t)| < P(t) \\ p(t)*sign[x(t)] \text{ if } |x(t)| > P(t), \end{cases}$$

$$\text{где } P(t) = A * LPF(|x(t)|);$$

здесь аббревиатура LPF означает операцию низкополосной фильтрации с подстраиваемой частотой фильтра, а коэффициент “A” регулирует уровень выходного сигнала.

Интерфейс пользователя с программой осуществляется в соответствии с рассмотренной выше на рис. 2.7 блок-схемой.

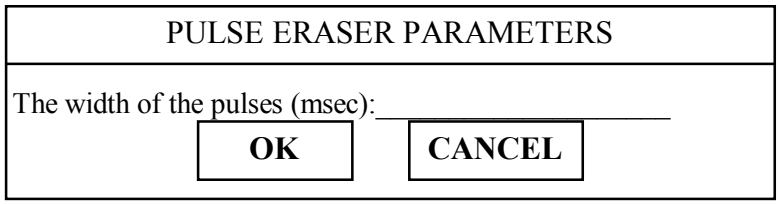
Процедура запуска программы состоих из двух частей:
1. Оценки средней ширины (скважности) импульсов,
2. Запуска программы фильрации

Эта программа не имеет каких либо параметров, устанавливаемых с помощью команды SET.

Единственым настраиваемым параметром программы является оценка ширины импульсов, которая выполяется экспертным методом с помощью предварительного просмотра сигнала на CSL. Максимально возможная щирина импульсов не должна превышать 3.0 миллисекунд.

После того, как ширина импульсов оценена, аудиоматериал, подлежащий обработке, загружается в активное окно, и, если необходимо, выбирается фрагмент, на котором будет выполняться фильтрация.

Мышь Выберите член ERASER в главном меню. На экране появится раскрывающееся подменю. Выберите соответствующий элемент
DELETE IMPULSES : ALL DATA или элементо
DELETE IMPULSES : MARKED SECTION (в зависимости от того все данные подлежат обработке или только выбранный фрагмент). На экране появится следующий диалог:



Введите оцененную среднюю ширину импульсов в миллисекундах (это значение должно быть в диапазоне [0 - 3.0 ms] и нажмите кнопку OK. Программа стартует.

Эта программа требует существенно больше вычислительных ресурсов и поэтому в реальном времени выполняется только на достаточно мощных машинах.

Ввведите команду NOPULSE = 0 * или аналогичную (используя обычный синтаксис командной строки CSL) и вы увидете на экране диалог, приведенный выше.

2.4. Програма подавления эхо и реверберации

Эхо помехи и реверберация - одни из наиболее часто всречающихся типов помех. Эхо сигнал присутствует в аудиоматериале каждый раз, когда запись делается в обычном помещении. Во время телефонных разговоров на большом расстоянии также присутствует это сигнал. Другим распространенным случаем является так называемый копир-эффект. Когда при длительном хранении магнитных материалов сигналс одного участвам магнитной записи копируется (при соприкосновении) на смежные. В общем случае эхо сигнал довольно сложно подавить полностью. Однако эффект его присутствия может быть существенно уменьшен. Наиболее удачный, с точки зрения качества фильтрации, вариант подавления эхо сигнала можно выполнить если время задержки примерно постоянно и может быть оценено по аудиоматериалу.

Программа пожавления эхо-сигнала имеет два режима работа. Программа Remove_Echo удалает реверберацию из аудиосигнала. Программа AddEcho наоборот, добавляет эхо к аудиоматериалу, выполняя обратную функцию по отношению к первой программе.

Идея алгоритма фильтрации эхо сигнала основана на том, что результарующий сигнал y(t) оценивается из исходного, содержащего эхо, сигнала x(t) как:
`y(t) = x(t) + A * x(t-del), где` “del” это времся задержки, а “A” это коэффициент усиления.

Программа добавления это-сигнала выполняет обратную функцию: **`y(t) = x(t) - a * x(t-del);`**

Интерфейс пользователя с программой поавления эха соответствует рассмотренному ранее на рис 2.7. Как и в предыдущих случаях процесс взаимодействия с программой состоит из двух шагов:
1. Установки необходимых режимов работы,
2. Вызова программы на выполнение.

Установка параметров работы программы эхо-подавления

Для установки параметров эхо-подавления можно использовать режим командной строки или мышь:

Мышь. Выберите элемент главного меню **ERASER**. Появится подменю программы фильтрации. Выберите необходимый элемент установки (а именно **Set Speak ON/OFF** либо **Set Destruct Mode ON/OFF**)

Клавиатура. Введите команду SET для выбоанного параметра обработки в соответствии с рассмотренной выше таблицей 2.7.

Запуск программы подавления эха

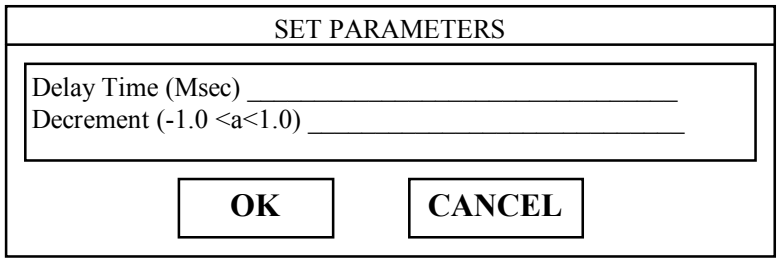
Загрузите аудиометриал в активное окно. Выберите сегмент данных для фильтрации.

Мышь. Выберите в главном меню CSL элемент ERASER. Появится подменю программы фильтрации.

Если треуется удалить эхо (программа RemoveEcho) выберите элементы подменю **REMOVE ECHO : ALL DATA** или
REMOVE ECHO : MARKED SECTION, в зависимости о того, требуется удалсить все данные или только выделенный сегмент.

Если требется программа добавления эхо, выберите элемент подменю **ADD ECHO : ALL DATA** или **ADD ECHO : MARKED SECTION**.

На экране появится следующий диалог:



введите величину оцененной задержки (в миллисекундах) and коэффициента затухания (в диапазоне от-1.0 to +1.0) и выберите кнопку **OK**. Процесс фильтрации начался.

Клавиатура. Для того, чтобы вызвать эхо-подавитель, введите командную строку типа: **NOECHO = 0 *** или аналогичную, использую стандартный синтаксис CSL. Для добавления эхо-сигнала введите команду **ECHOP = 0 ***.

5. The Butterworth Digital Filters

Introduction

The Butterworth digital filter program provides the user with possibilities for synthesis and processing the data by the digital Butterworth filters. Namely the four subroutines are included:
- low pass filter;
- band pass filter;
- band stop filter;
- high pass filter

Butterworth Digital Filter Program Flowchart.

The interface of the user with the Butterworth digital filter procedure is performed with accordance to the following Flowchart

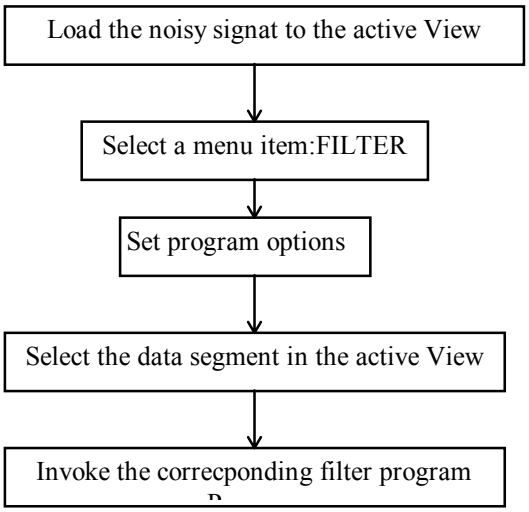


Fig 4.7. The Butterworth digital filter Flowchart

As it follows from the Flowchart the procedure consists of two separate steps:
1. Set the necessary program options ,
2. Invoke the Butterworth digital filter procedure.

The parameter setting is accomplished via the CSL SET parameter command option.

Butterworth digital filter SET parameter commands

In order to use SET parameter option with the Butterworth digital filter procedure you should make the following:

Mouse. Click the item **FILTER** of the Main Menu. The **FILTER** pull-down submenu will appear on the screen. Click the desired item (namely **Set Speak ON/OFF** or **Set Destruct Mode ON/OFF**) in the pull-down submenu.

Keyboard. Enter the command **SET** for the properly selected parameter with accordance to the following Figure

Menu Selection	Implemented Command	Explanation
Speak [On/Off]	SET FonSpeak [On/Off]	While data processing the output signal will be sent to the speakers. User can listen the enhanced signal “as it is”in the real time.
Destruct [On/Off]	SET SaveResult[On/Off]	On indicate that processing is destructive. If Off is chosen, the data in the memory will not changed. The user can estimate how algorithm is work but signal handled virtually: no real data changes in the memory are made.

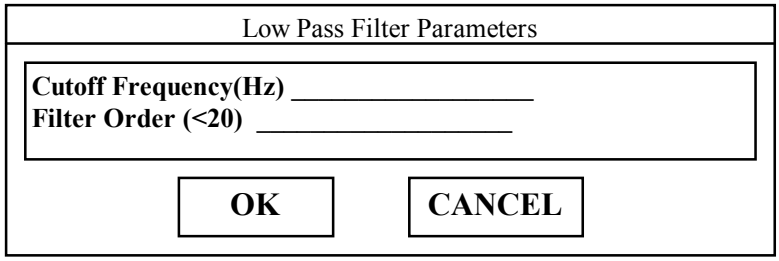
Fig 4.7. Butterworth digital filter SET parameter commands

How to launch the Butterworth digital filter program

Load the noisy audio material into the active window. Select the segment of data for processing.

Click the item FILTER in the CSL Main Menu. The pull-down submenu will appear.

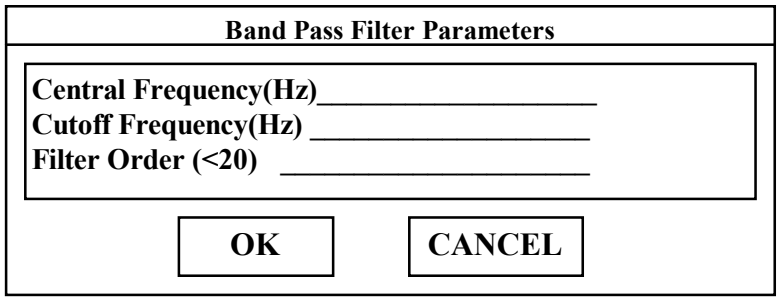
If the low pass filter is required click the submenu item **Low Pass: All Data** or **Low Pass: Marked Section** depending on all data or the only marked section of data should be processed. You should see the following dialog box:



Enter the values of the filter cutoff frequency (in hertz) and the filter order (ranges from 0 to 20) and click OK button. Process will start.

If the high pass filter is required click the submenu item **High Pass: All Data** or **High Pass: Marked Section** depending of the all data or the only marked section of data should be processed. You should see the dialog box for the high pass filter that looks like the previous one. Enter the required values of the cutoff frequency and the filter order and click OK button.

If the band pass filter is required click the submenu item **Band Pass: All Data** or **Band Pass: Marked Section** depending on all data or the only marked section of data should be processed. You should see the following dialog box:



Enter the values of the filter central frequency, cutoff frequency (here the term cutoff frequency stands for the half of the width of the band pass filter) and the filter order (ranges from 0 to 20) and click OK button. Process will start.

If the band stop filter is required click the submenu item **Band Stop Pass: All Data** or **Band Pass: Marked Section** depending on all data or the only marked section of data should be processed. You should see the dialog box like in the previous case but for the Band Stop filter.

Enter the values of the filter central frequency, cutoff frequency (here the term cutoff frequency stands for the half of the width of the band reject filter) and the filter order (ranges from 0 to 20) and click OK button. Process will start.

To invoke the Low Pass filter enter the command like: **LowP = 0 *** or similar (use the ordinary CSL command line syntax to define the data segment). Analogously to invoke the High Pass filter via keyboard it is necessary to use the command **HighP = 0 *** (for all data in the active view screen) or **HighP = M1 M2** for marked section of the data in the active view screen.

To invoke the Band Pass filter use command **BandP** and for Band Stop filter use command BandS.

Examples:

BandP = 0 * for processing of all data in the active view screen

BandS = M1 M2 for processing the marked section of data in the active view screen

BandP ? 0 = for processing the data from beginning to cursor in the source view screen

In all cases before processing will start you should see the corresponding dialog box and enter the selections of parameters.

Литература

1. Гурьев Ю.Ю. Прохоров Ю.Н. Алгоритм рекуррентной фильтрации речевых сигналов. Материалы Всесоюзного семинара АРСО-12. Киев, 1982, с.39-42.
2. Лим Дж, Опперхайм А.В. Коррекция и сжатие спектра зашумленных речевых сигналов, ТИИЭР, т.67, 12, 1979
3. Назаров М.В., Ковязин В.И. Марковская модель речевого сигнала. Материалы Всесоюзного семинара АРСО-12. Киев, 1982, с.44-49.
4. Назаров М.В., Ковязин В.И. Марковская модель речевого сигнала. Материалы всесоюзного симинара АРСО-13. Новосибирск, 1984, с.153.
5. Пономарев Е.П., Прохоров Ю.Н. Адаптивная линейная фильтрация при первичной обработке речевых сигналов. Материалы Всесоюзного семинара АРСО-10, Тбилиси, Мецниереба, 1978.
6. Прохоров Ю.Н. Рекуррентное оценавание параметров. - В кн.: Проблемы построения систем понимания речи. М., Наука, 1980, с. 97-109.
7. Прохоров Ю.Н. Статистические модели и рекуррентное предстазание речевых сигналов. М., Радио и Связь, 1984
8. Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М. Методы оптимизации. М.. Наука, 1978.
9. Овчинникова О.П. Повышение разборчивости речи путем цифровой фильтрации. 9 Всесоюзная акустическая конференция, М., 1977, Выпуск Ф. с.33-36
10. Санников В.Г., Журавский Ю.И. Прохоров Ю.Н. Формирование банка априорных данных о речи диктора. Материалы всесоюзного семинара АРСО-12, Киев, 1982, с.49-52.
11. Сейдж Э., Мелс Дж. Теория оценивания и ее применение в теории связи и управления. М., Связь, 1976.
12. Akbari Azirani A., R.J. Bouquin Optimizing Speech Enhancement by Exploiting Masking Properties of the Human Ear. Proc Int.Conf on Acoustics, Speech and Signal Proc. ICASSP-96, pp.800-803, 1996
13. Arslan M. Levent, Hansen John H.L. Speech Enhancement for Crosstalk Interference. IEEE Signal Processing Letters, Vol. 4, No.4, April 1997.
14. Boll S.F. Suppression of Acoustic Noise in Seech Using Spectral Subtraction. IEEE Trans. ASSP, Vol.27, No.2, pp 113-120, 1979.
15. Cappe O. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. IEEE Trans Speech Audio Process., Vol 2., No.2, pp.345-349.
16. Curtis R.A. Niederjohn R.E. Several Frequency Domain Processing Methods for Enhansing the Intelligibility of Speech in Wideband Random Noise, Proc.1978 IEEE Int Conf on ASsP, pp602-605.
17. Drucker H. Speech Processing in a High Ambient Noise Environment. IEEE Trans. On ASSP, ICASSP-76, pp.251-253.

18. Ephraim ,Y. , Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans Acoust, Speech and Signal Process, Vol ASSP-32, pp.1109-1121.
19. Ephraim Y. A Minimum Mean Square Error Approach for Speech Enhancement" Proc. ICASSP-90, pp.829-832.
20. Ephraim Y. Statistical model-based speech enhancement systems" Proc. IEEE, Vol 8, No 10, pp. 1526-1555, 1992
21. Fechner T. Nonolinear noise filtering with neural networks: comparison with Wiener optimal filtering" Proceedings of IEE conf. On Artificial Neural Networks, 1993, p 143-147.
22. Frasier R.E. , etc. Enhancement of Speech by Adaptive Filtering. Proc.1976 IEEE Int. Conf. ASSP, ICASSP-76, pp.251-253.
23. Gualtiero A., Giancarlo P. Noisy Speech Enhancement: a comparative analysis of three different techniques, Alta Frequenza, 1984, 535, No 3, pp.190-200.
24. Johnston J.D. Transform Coding of Audio Signals Using Perceptusl Noise Criteria. IEEE Journal on Selected Areas in Communications, Vol.6, No.2, pp 314-323, 1988
25. Junqua J.C. Haton J.P. Robustness in automatic speech recognition, Kluwer Academic Publishersm 1996
26. Hanson B.A., Wong D.Y., Yuang B.N. Speech Enhancement With Harmonic Synthesis. Proc. 1983 IEEE Int . Conf. ASSP, pp 1122-1125.
27. Hansen John H.L. Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. Speech Communication, Vol 20, 1996, 151-173.
28. Hansen J.H.L. Morphological constrained enhancement with adaptive cpestral compensation (MCE-ACC) for speech recognition in noise and Lombard effect. IEEE Trans.on Speech and Audio Processing. Special Issue: Robust Speech Recognition. Vol.2, No.4, pp.598-614.
29. Hansen G.H.L, Pellom B.L. Text-directed speech enhancement employing phone class parsing and feature map constrained vector quantizaion. Speech Communication, Vol 21, 169-189, 1997.
30. Hermansky H., Morgan N. RASTA Processing of Speech. IEEE Trans. on ASSP, Vol.2 Oct.1994, pp.578-589.
31. Hoy L.D., etc. Noise Suppression Methods for Speech Applications. Proc. 1983 IEEE Int .Conf.ASSP, ICASSP-83, pp.1133-1136.
32. Malah D., Cox R.V. A Generalized Comb Filtering Technique for Speech enhancement. Proc 1982 IEEE Int. Conf ASSP, pp. 160-163.
33. Petersen T.L. , Boll S.F., etc. Acoustic Noise Suppression in the Context of a Perceptual Model. Proc 1981 IEEE Int. conf. ASSP, ICASSP-81, pp.1086-1088.
34. Pinter I. Perceptual wavelet-representation of speech signals and its application to speech enhancement. Computer Speech and Language, 1996, 10, 1-22.
35. Laughans T., Strube H.W. Speech Enhancement by Nonlinear Multiband Envelop Filtering. Proc, 1978 IEEE Trans, ASSP, pp 156-159.

36. Le T.T. Mason J.S. Artificial neural networks for nonlinear time-domain filtering of speech. IEE Proc on Vis Image Signal Processing, vol. 143, No 3, pp 149-154, 1996

37. Lee K.Y., Lee B.-G., Song I. etc. Recursive Speech Enhancement Using the EM Algorithm with Initial Conditions Trained by HHM's. Int Conf on Acoustics, Speech and Signal Proc, ICASSP-96, 1996, pp.621-624.

38. Lim J.S., at al. Speech Enhancement. IEEE Trans. ASSP, Vol ASSP -26, No 9, 1979, pp.357-358

39. Lyon R.F. A computational Model of Filtering, Detection and Compression in the Cocklea. Proc, Int.Conf. on Acoust.Speech and Signal Proc. ICASSP-1982, pp.1282-1283.

40. Mammone R.J.Robust Speaker Recognition. IEEE Signal Processing Magazine, Sept.1996, pp.68-71.

41. McKinley B.L., Whipple G.H. Noise Model Adaptation in Model Based Speech Enhancement. Int.Conf on Acoustics, Speech and Signal Processing, ICASSP-96, 1996, pp.633-636.

42. Rahim M.G., Juang B.-H. etc. Signal Conditioning Techniques for Robust Speech Recognition. Signal Processing Letters, Vol.3, No. 4, April 1996, pp. 107-109.

43. Sambur M.R. Adaptive Noise Cancelling for Speech Signals. IEEE Trans. ASSP, vol.ASSP-26, 1978, pp.419-423.

44. Scalart P. Speech Enhancement Based on a Priori Signal to Noise Estimation. Proc. Int. Conf on Acoustics, Speech and Signal Proc. ICASSP-96, 1996, pp.629-632.

45. Scalart P., Benamar A. A system for speech enhancement in the context of hands-free radiotelephony with combined noise reduction and acoustic echo cancellation. Speech Communication, Vol.20, pp.203-214, 1996.

46. Sheikhzadeh H., Sameti H., Deng L. Comparative Performance of Spectral Subrtraction and HMM Based Speech Enhancement Strategies with Application to Hearing Aid Design. Proc. ICASSP-94, p.13-17.

47. Sheikhzadeh H., Sameti H, Brennan R.L. Real-Time Implementation of HMM-Based MSE Algorithm For Speech Enhancement in Hearing Aid Applications. Proc of Int.Conf on ASSP, ICASSP-95, 1995, p.808-811.

48. Sondhi M.M., Schmidt C.E., Rabiner L.R. Improving the Quality of Noisy Speech Signal, Bell Syst. Tech Journ, Vol.60, No.8. 1981, pp.1847-1858.

49. Teolis A., Benedetto J.J. Noise Suppression Using a Wavelet Model. Int Conf on ASSP, ICASSP-94, 1994, pp.17-20.

50. Virag N. Speech Enhancement Based on Masking Properties of the Auditory System. Proc of Int.Conf of Acoustics, Speech and Signal Proc., ICASSP-96, 1996, pp.796-799.

51. Widrow B., at al. Adaptive Noise Candelling: Principles andApplications. Proc IEEE, Vol63, No. 12, 1975, pp.1672-1716.

52. Whipple G. Low Residual Noise Speech Enhancement Utilizing Time-Frequency Filtering, Proc of ICASSP'94, p5-9.

53. McWhirer J.S., Palmer K.J., Robers J.B. A Digital Adaprive Noise-Canceller Based on a Stabilizer Version of the Widrow L.M.S. Algorithms, Proc. 1982, IEEE Int. Conf. ASSP, pp.1384-1387.
54. Yoo C. Selective All Pole Modeling of Degraded Speech Using M-Band Decomposition. Int.Conf on Acoustics, Speech and Signal Proc. ICASSP-96, 1996, pp.641-644.