ISSN 1028-7507 Акустичний вісник. 2003. Том 6, N 3. С. 34 – 43

УДК 621.391:534.78

МЕТОД СЛЕПОЙ ДЕКОНВОЛЮЦИИ РЕЧЕВЫХ СИГНАЛОВ, ОСНОВАННЫЙ НА АНАЛИЗЕ ЛИНЕЙНЫХ СПЕКТРАЛЬНЫХ ЧАСТОТ

А. Я. КАЛЮЖНЫЙ[*], **В. Ю. СЕМЕНОВ**[**]

[*] *Научно-производственное предприятие "Дельта", Киев*
[**] *Институт гидромеханики НАН Украины, Киев*

Получено 02.10.2003

Рассмотрена задача слепой деконволюции речевых сигналов при наличии фоновых шумов. Предложен эффективный метод детектирования и устранения влияния передаточной функции среды, основанный на анализе линейных спектральных частот искаженного сигнала. Его принципиальное преимущество перед существующими подходами состоит в адаптивном учете помехи в структуре алгоритма. Разработана эффективная процедура локализации посторонних резонансов, основанная на анализе разностей линейных спектральных частот. С целью устранения эффекта "усиления шума" использован алгоритм блочной калмановской фильтрации. Эффективность результирующего метода проверена на искусственных и реальных искажениях речевых сигналов. Предлагаемый подход характеризуется более низкими вычислительными затратами по сравнению с рядом современных методов слепой деконволюции речевых сигналов.

Розглянуто задачу сліпої деконволюції мовних сигналів в умовах присутності фонових шумів. Запропоновано ефективний метод детектування та компенсації впливу передаточної функції середовища, який базується на аналізі лінійних спектральних частот спотвореного сигналу. Його принципова перевага перед існуючими підходами полягає в адаптивному урахуванні шуму в структурі алгоритму. Розроблено ефективну процедуру локалізації сторонніх резонансів, засновану на аналізі різниць лінійних спектральних частот. З метою усунення ефекту "підсилення шуму" використано алгоритм блокової калманівської фільтрації. Ефективність результуючого методу перевірено на штучних і реальних спотвореннях мовних сигналів. Запропонований підхід характеризується більш низькими обчислювальними витратами, у порівнянні з низкою сучасних методів сліпої деконволюції мовних сигналів.

A problem of the blind deconvolution of speech signals at presence of noise background is considered. An effective method of detection and compensation of media transfer function is proposed. The method is based on the analysis of line spectral frequencies of a distorted signal. Its principal advantage over existing approaches is an adaptive noise compensation inside the structure of the algorithm. An effective procedure of localization of media resonances, based on the analysis of differences of line spectral frequencies, is developed. To diminish the "noise enhancement" effect a block Kalman filter is used. The effectiveness of the resulting method is verified on artificial and real distortions of speech signals. The proposed approach is characterized by lower computational expenses in comparison to the number of modern blind deconvolution of speech signals methods.

ВВЕДЕНИЕ

В настоящее время все более широкое развитие получают системы автоматического распознавания речи, а также системы верификации и идентификации дикторов [1, 2]. Их наиболее современные варианты допускают режим дистанционного применения (например, речевое управление по телефонным каналам системами жизнеобеспечения жилого дома, дистанционная верификация клиентов банка по речевому паролю и т. д.). Одной из наиболее сложных проблем при создании таких дистанционных систем является коррекция канала связи. Суть проблемы состоит в том, что в каждом конкретном случае речевые эталоны, используемые системой, получаются в совершенно определенных условиях с точки зрения как акустической обстановки, так и характеристик используемых линий связи. На практике же реальные условия при использовании системы могут заметно отличаться от эталонных, что должно приводить к ошибкам распознавания. Для преодоления этих трудностей в системы дистанционного распознавания в качестве начального устройства следует включать корректор (эквалайзер) канала [2], т. е. устройство, которое приводит речевой сигнал к условиям, максимально близким к эталонным.

В традиционной технике связи проблема коррекции обычно решается за счет использования на начальном этапе соединения специальных измерительных сигналов, по которым оценивается импульсная переходная характеристика (ИПХ) канала [3]. Однако такой подход не всегда возможно применить в системах дистанционного распознавания речи. Кроме того, он не обеспечивает учета акустических характеристик среды. Исходя из этого, единственной возможностью остается оценивание суммарной характеристики канала и акустической среды непосредственно по информационному акустическому сигналу. Поскольку этот сигнал заранее неизвестен, такая задача получила название слепого выравнивания (blind equalizati-

on) или слепой деконволюции. Формально постановку задачи слепой деконволюции можно представить соотношением

$$z(n) = s(n) \otimes h(n) + v(n), \qquad (1)$$

в котором регистрируемый сигнал $z(n)$ представляет собой сумму свертки речевого сигнала $s(n)$ с неизвестной ИПХ среды $h(n)$ и фонового шума $v(n)$.

Задачи слепой деконволюции известны и в архитектурной акустике (например, при устранении влияния реверберации помещений). В последнее время приобрела популярность идея восстановления сигнала с помощью кепстрального подхода [4, 5]. Заметим, однако, что используемые до сих пор в акустике методы базируются на многоканальном приеме сигнала – в нескольких (двух и более) точках акустической среды. В то же время, в силу специфики использования систем дистанционного распознавания, для обработки доступен, как правило, только один канал. Задача же одноканальной слепой деконволюции является чрезвычайно сложной и нетрадиционной для акустики.

Один из первых методов одноканальной слепой деконволюции, предложенный в работе [6], основан на идее относительной стационарности искажающего воздействия. Этот метод применялся, в частности, к восстановлению старинных звукозаписей. Однако в нем подразумевается наличие опорного сигнала, сходного по своим спектральным характеристикам с сигналом, подлежащим восстановлению. Несмотря на эффективность полученных частных результатов, очевидно, что данный подход имеет весьма ограниченную область применения.

Существенным продвижением в направлении решения задач одноканальной слепой деконволюции сигналов стал метод, предложенный в работах [7, 8]. Принципиальным его отличием было использование специальных математических моделей среды и полезного сигнала.

Полюсная модель передаточной функции среды

Дискретная передаточная функция линейного объекта в общем случае может быть представлена в следующем каноническом виде [9] (коэффициент усиления опущен, так как он может быть учтен в модели полезного сигнала):

$$H(z) = \frac{\sum\limits_{l=0}^{n} d_l z^{-l}}{\sum\limits_{k=0}^{m} c_k z^{-k}} = \frac{\prod\limits_{l=1}^{n} (1 - \delta_l z^{-1})}{\prod\limits_{k=1}^{m} (1 - \gamma_k z^{-1})}, \qquad (2)$$

т. е. она содержит как полюса γ_k, $k = 1, 2, \ldots, m$, так и нули δ_l, $l = 1, 2, \ldots, n$. Однако в очень широком классе задач слепой деконволюции речевых сигналов можно ограничиться рассмотрением дискретной передаточной функции, содержащей только полюса. Помимо того, что произвольный амплитудный спектр всегда может быть аппроксимирован спектром полюсной модели достаточно большого порядка [10], существуют и более веские “физические” причины, обосновывающие такое упрощение.

Так, в случае старинных звукозаписей искажения сводятся в основном к появлению в сигнале практически неизменных во времени посторонних резонансов, вносимых записывающим устройством. Как известно, значение резонансной частоты и ее интенсивность определяются парой комплексно сопряженных полюсов[1]. Поэтому передаточная функция искажающего воздействия может быть представлена полюсной моделью, порядок которой вдвое превосходит количество резонансных частот записывающего тракта.

Всеполюсная модель пригодна и для описания искажений, вносимых реверберацией помещений. Нули передаточной функции помещения характеризуют локальные взаимопогашения звуковых волн, распространяющихся внутри его в результате многократных отражений от стен [8, 11]. Поэтому они очень чувствительны к изменениям во взаимном расположении источника и приемника сигнала. В то же время, полюса передаточной функции помещения характеризуют резонансы данного замкнутого объема, которые практически не изменяются при изменении пространственной конфигурации системы “источник – приемник” [11]. Кроме того, порядок полюсной модели помещения может оказаться значительно меньшим по сравнению с порядком соответствующей всенулевой модели [8].

Дополнительным аргументом для использования всеполюсной модели передаточной функции среды служит то, что “провалы” в спектре результирующего сигнала, определяемые нулями

[1] Резонансная частота, соответствующая паре полюсов $\gamma_{1,2} = r \exp(\pm i\phi)$, определяется как $f_{\text{рез.}} = \phi f_s/(2\pi)$ (f_s – частота дискретизации), а соответствующая ширина полосы есть $B = -f_s \ln r/\pi$.

z-преобразования, значительно менее важны с точки зрения правильного восприятия звуков по сравнению с пиками АЧХ (они характеризуются полюсами) [13]. Другими словами, человеческий слух значительно менее восприимчив к спектральным нулям. Вследствие этого всеполюсная модель может быть принята для описания очень широкого класса искажающих воздействий сверточного типа[2].

Авторегрессионная модель речеобразования

Большинство современных методов обработки речи также основаны на использовании полюсной (авторегрессионной) модели речеобразования [12]. В ней речевой сигнал $s(n)$ представляется как результат прохождения управляющего (возбуждающего) процесса $w(n)$ через полюсной фильтр

$$H_s(z) = \frac{g}{1 + \sum\limits_{k=1}^{p} a_k z^{-k}}, \qquad (3)$$

где коэффициент усиления g характеризует уровень сигнала, а авторегрессионные (АР) коэффициенты a_k, $k=1,2,\ldots,p$ определяют форму голосового тракта в момент произнесения звука. Порядок АР модели p, как правило, выбирается в пределах от 8 до 20. Возбуждающий процесс моделирует поток воздуха на выходе голосовых связок человека. В задачах обработки речевых сигналов обычно считается, что параметры АР модели неизменны на временных интервалах длиной $10 \div 30$ мс (свойство квазистационарности).

Подход к идентификации ИПХ среды, представленный в работах [7,8], основан на том, что полюса z-преобразования искаженного сигнала, соответствующие передаточной характеристике среды, не меняют своего расположения внутри единичного круга с течением времени (или изменяются очень медленно). В результате накопления гистограмм полюсов искаженного сигнала отбирались те полюса, которые преобладали на общем фоне. Это давало возможность идентифицировать знаменатель дискретной передаточной функции (2).

Главным недостатком такого подхода является неучет влияния фонового шума. Общеизвестно, что оценки АР коэффициентов, полученные с помощью автокорреляционного и ковариационного методов линейного предсказания, становятся ненадежными даже при относительно небольшом уровне фоновых помех [16, 17]. Поэтому анализ полюсов искаженного сигнала, основанный на использовании традиционных методов, может изначально привести к принципиально неверному решению относительно присутствия в сигнале посторонних искажений или обеспечить в корне неправильную оценку передаточной характеристики искажающего воздействия.

Другой важный недостаток указанной методики состоит в том, что вычисление комплексных корней полиномов традиционно является нежелательным элементом для систем цифровой обработки сигналов (ЦОС), работающих в режиме реального времени, поскольку приводит к непредсказуемым временным задержкам и подвержено накоплению ошибок округления. Это становится особенно заметным при рассмотрении АР моделей относительно высоких порядков.

В работах [8, 18] представлены альтернативные подходы, основанные на оценивании АР параметров искаженного сигнала по методу максимума апостериорной вероятности, а также методе Монте-Карло и квантователе Гиббса [19]. Несмотря на серьезную теоретическую обоснованность, они имели очевидный недостаток, связанный с отсутствием сколько-нибудь определенной информации об априорных распределениях АР параметров речи. Отсутствие надежных методов глобальной максимизации функций многих переменных, с одной стороны, и недопустимо высокие (с точки зрения устройств реального времени) вычислительные затраты метода Монте-Карло, с другой, являются дополнительными объективными препятствиями для использования этих алгоритмов в системах ЦОС. Концептуально упомянутые методы основывались на построении итерационного процесса, на каждом шаге которого обновляется оценка АР параметров сигнала. Однако использование вектора АР коэффициентов в качестве переменной итерационного алгоритма крайне нежелательно, так как даже очень малые погрешности их вычисления могут привести к существенным изменениям в спектре восстановленного сигнала [20]. Кроме того, здесь по-прежнему не затрагивалась проблема учета фонового шума.

Отметим также, что ни в одном из перечисленных источников (за исключением частного случая, описанного в [6]) не учитывался эффект "noise enhancement" (усиления шума), заключающийся в том, что пропускание сигнала через фильтр, обратный идентифицированному, приводит к по-

[2]Использование полюсной модели искажающего воздействия избавляет нас от решения трудной задачи построения обратного фильтра [14, 15]. Действительно, таковым будет фильтр с конечной импульсной характеристикой, коэффициенты которого совпадают с оценками коэффициентов знаменателя дискретной передаточной функции.

явлению аддитивной помехи $v_1(n)$ со спектром $\widehat{V}_1(\omega)=V_1(\omega)/\widehat{H}(\omega)$. Даже если уровень исходного шума был приемлемым, возникающая аддитивная помеха способна настолько ухудшить качество сигнала, что улучшение, достигнутое благодаря устранению влияния передаточной функции среды, нивелируется.

В связи с указанными недостатками существующих подходов, предлагается новый эффективный метод детектирования и устранения влияния передаточной функции среды, основанный на анализе линейных спектральных частот (ЛСЧ) наблюдаемого сигнала. Его принципиальным преимуществом является адаптивный учет помехи в структуре алгоритма. Помимо повышения надежности получаемых результатов, это позволяет исключить возможность идентификации окрашенного шума с сильно выраженной резонансной структурой как "полюсного" искажения (это было свойственно, в частности, подходу, описанному в [7,8]). Кроме того, предложена эффективная процедура локализации резонансов, основанная на анализе разностей ЛСЧ. С целью устранения эффекта усиления шума введена фильтрационная процедура, использующая блочный фильтр Калмана, предложенный в работе [21]. Отметим, что обсуждаемый подход характеризуется принципиально более низкими вычислительными затратами по сравнению с методикой [7,8].

1. СВЯЗЬ ЛИНЕЙНЫХ СПЕКТРАЛЬНЫХ ЧАСТОТ С ФОРМАНТАМИ РЕЧЕВЫХ СИГНАЛОВ

В настоящее время наиболее популярным способом частотного представления АР параметров являются линейные спектральные частоты (см. обзоры [20, 22]). Формально ЛСЧ ω_k, $k=1,2,\ldots,p$ можно определить как аргументы корней полиномов $G_1(z)$ и $G_2(z)$, лежащие в диапазоне $(0,\pi)$. Упомянутые полиномы получаются из исходного отбеливающего полинома

$$A(z)=1+\sum_{k=1}^{p}a_k z^{-k}$$

следующим образом:

$$\begin{cases} G_1(z)=\dfrac{A(z)+z^{-p-1}A(z^{-1})}{1+z^{-1}}, \\ G_2(z)=\dfrac{A(z)-z^{-p-1}A(z^{-1})}{1-z^{-1}}. \end{cases} \qquad (4)$$

Заметим, что корни полиномов $G_1(z)$ и $G_2(z)$ лежат на единичной окружности и чередуются между собой [23]. Учитывая тесную связь величин ω_k с формантными частотами (см. ниже), будем подразумевать под ЛСЧ значения $f_k=\omega_k f_s/(2\pi)$, лежащие в диапазоне $(0, f_s/2)$, где f_s – частота дискретизации.

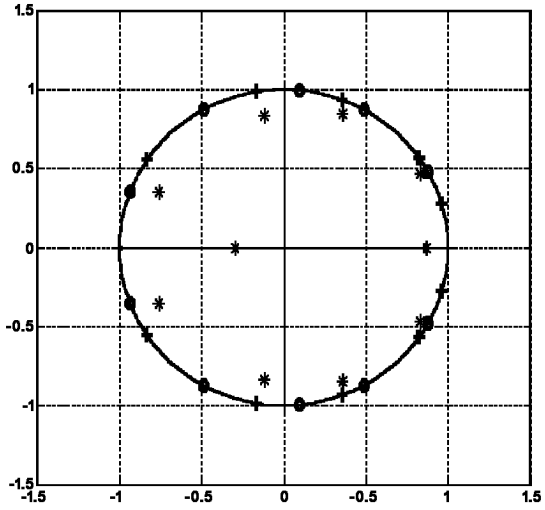

Рис. 1. Нули полиномов $A(z)$, $G_1(z)$ и $G_2(z)$

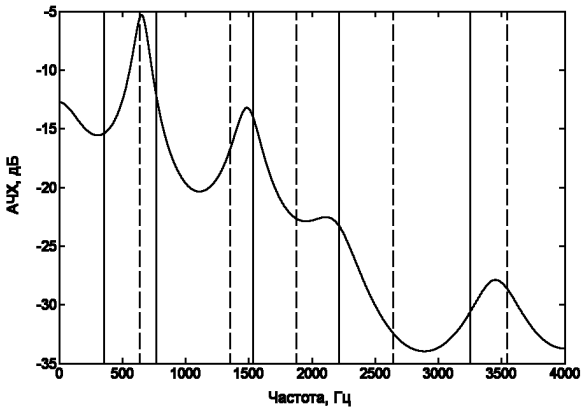

Рис. 2. Взаимное расположение формантных частот и соответствующих ЛСЧ

С точки зрения решения задачи компенсации передаточной функции среды, наиболее важным является свойство ЛСЧ, характеризующее их взаимосвязь с формантными частотами. Рассмотрим в качестве примера фрагмент речевого сигнала длиной 20 мс (160 дискретных отсчетов при $f_s=8000$ Гц), произносимого диктором-мужчиной, и вычислим соответствующие АР коэффициенты с помощью автокорреляционного метода [12]. На рис. 1 показано расположение ну-

лей исходного полинома $A(z)$ и нулей полиномов $G_1(z)$, $G_2(z)$, соответствующих ЛСЧ. Корни $A(z)$ обозначены маркерами "$*$", а нули $G_1(z)$ и $G_2(z)$ – "$+$" и "$\circ$" соответственно. Амплитудно-частотная характеристика АР фильтра $1/A(z)$ и расположение соответствующих ЛСЧ приведены на рис. 2 (ЛСЧ, соответствующие полиному $G_1(z)$ обозначены сплошной линией, а $G_2(z)$ – штриховой).

Из рис. 1 видно, что пары соседних ЛСЧ стремятся ограничить те комплексные полюса, которые находятся близко к единичной окружности, т. е. именно те, которые определяют резонансные частоты в спектре сигнала. Чем ближе к единичной окружности находится полюс (чем сильнее выражена в спектре сигнала соответствующая резонансная частота), тем точнее он аппроксимируется парой ЛСЧ. И наоборот, чем дальше от единичного круга находится полюс, тем дальше друг от друга отстоят ЛСЧ в соответствующей паре. Это подтверждает и рис. 2, показывающий, что каждая формантная частота окружена набором из двух или трех ЛСЧ, а ширина полосы соответствующего резонансного пика зависит от их близости. Указанная особенность ЛСЧ играет важную роль при разработке и обосновании предлагаемого метода детектирования и устранения искажений, вносимых средой передачи речевых сообщений.

Отметим, что подсчет ЛСЧ характеризуется принципиально более низкими вычислительными затратами по сравнению с поиском комплексных корней полиномов. Эффективный метод вычисления ЛСЧ, обладающий рядом существенных преимуществ перед аналогами, предложен в работе [20]. Создание метода компенсации влияния передаточной функции среды, основанного на анализе ЛСЧ, особенно важно, поскольку вычисление ЛСЧ является неотъемлемой частью подавляющего большинства современных систем сжатия речевых сигналов [22, 24], а также ряда алгоритмов распознавания речи и идентификации диктора [25, 26].

2. АЛГОРИТМ ПОМЕХОУСТОЙЧИВОЙ ДЕКОНВОЛЮЦИИ РЕЧЕВЫХ СИГНАЛОВ

Исходя из упомянутых свойств ЛСЧ, можно сделать вывод о характере их распределения в речевом сигнале, искаженном полюсным фильтром. Определенные пары ЛСЧ должны постоянно группироваться вокруг резонансов, характеризующих искажающий фильтр. Это обстоятельство позволяет эффективно локализовать частотные диапазоны, в которых присутствуют посторонние резонансы. Подход, основанный на построении общей гистограммы распределения ЛСЧ, не является универсальным решением, поскольку во многих случаях, не связанных с присутствием в сигнале полюсных искажений, некоторые ЛСЧ могут обладать очень малой дисперсией. Это приводит к преобладанию в общей гистограмме соответствующих средних значений. Более конструктивным является рассмотрение разностей, характеризующих близость смежных ЛСЧ:

$$d_k = f_k - f_{k-1}, \qquad k = 2, \ldots, p. \tag{5}$$

Перейдем к построению критерия, определяющего наличие или отсутствие на некотором временном фрейме резонансной структуры. Будем полагать, что наличие на фрейме пары ЛСЧ, удаленных друг от друга менее, чем на некоторую критическую величину $\Delta f_{\text{кр.}}$, свидетельствует о наличии резонанса, заключенного между этими частотами. Экспериментально установлено, что в качестве $\Delta f_{\text{кр.}}$ целесообразно принять 125 Гц. Относительно небольшие отклонения от этой пороговой величины не приводили к существенным отличиям в получаемых результатах. Отметим, что "заподозренный" таким образом резонанс может принадлежать как сигналу, так и передаточной функции среды. Однако при проведении накопления всех подозрительных полюсов те из них, которые соответствуют полезному сигналу, не должны быть заметны на общем фоне в силу нестационарности речи. Отметим, что в работе [27], в которой рассматривалось использование разностей ЛСЧ для компрессии речевых сигналов, отмечалось, что при отсутствии искажений величины d_k (5) характеризуются ограниченным диапазоном изменения и относительной инвариантностью для разных дикторов. Это облегчает обнаружение отклонений в распределении разностей ЛСЧ, вызванных наличием полюсных искажений.

Предлагаемый алгоритм обнаружения искажений, вносимых передаточной функцией среды, состоит из выполнения на каждом фрейме следующих действий.

1. Подсчитывается автокорреляционная функция (АКФ) зашумленного сигнала (1): $R_z(k)$, $k = 0, 1, 2, \ldots, p$.

2. Вычисляется оценка АКФ свертки $s \otimes h$:

$$\widehat{R}_{s\otimes h}(k) = R_z(k) - \widehat{R}_v(k), \qquad k = 0, 1, 2, \ldots, p, \tag{6}$$

где АКФ помехи $\widehat{R}_v(k)$ оценивается адаптивно по фреймам с наименьшей энергией. Помни-

мо этого, на каждом фрейме производится экспоненциальное усреднение полученной АКФ $\widehat{R}_{s\otimes h}$ с коэффициентом усреднения $\alpha=0.98$.

3. Путем применения процедуры Левинсона–Дарбина [12] к АКФ $\widehat{R}_{s\otimes h}$ определяются предварительные оценки $\widehat{\boldsymbol{b}}^{(0)}_{s\otimes h}$ АР коэффициентов сигнала $s\otimes h$, соответствующие данному фрейму.

4. Значения параметров $\widehat{\boldsymbol{b}}^{(0)}_{s\otimes h}$ уточняются с помощью одной итерации алгоритма [16]. В результате получаем оценки коэффициентов $\widehat{\boldsymbol{b}}^{(1)}_{s\otimes h}$.

5. Полученные коэффициенты $\widehat{\boldsymbol{b}}^{(1)}_{s\otimes h}$ преобразуются в набор ЛСЧ f_k, $k=1,\ldots,p$ в соответствии с алгоритмом, предложенным в работе [20].

6. В наборе ЛСЧ выделяются пары $\{f^{(1,1)},\ f^{(1,2)}\},\ldots,\{f^{(m',1)},\ f^{(m',2)}\}$, удаленные друг от друга менее чем на величину $\Delta f_{\text{кр.}}$.

7. Каждой из отобранных таким образом пар $\{f^{(k,1)},\ f^{(k,2)}\}$, $k=1,\ldots,m'$ ставится в соответствие пара резонансных полюсов $z^{(k)}_{1,2}=re^{\pm i\phi}$, где

$$\phi=\pi\,\frac{f^{(k,1)}+f^{(k,2)}}{f_s}\,,\qquad k=1,2,\ldots,m',\quad(7)$$

$$r=\sqrt{1+\cos\left(2\pi\frac{f^{(k,1)}}{f_s}\right)-\cos\left(2\pi\frac{f^{(k,2)}}{f_s}\right)}\,,\quad(8)$$

$$k=1,2,\ldots,m'.$$

Полученные комплексные полюса выводятся на общую гистограмму.

8. На завершающей стадии из общей гистограммы отбираются наиболее интенсивные полюса, преобразуемые затем по стандартным формулам в оценки коэффициентов знаменателя передаточной функции среды $\widehat{c}_k$, $k=0,2,\ldots,m$ [28, с. 38].

Из формулы (7) видно, что оценки резонансных частот искажающего воздействия вычисляются в виде полусуммы "окаймляющих" ЛСЧ. Что касается определения модулей резонансных полюсов, то соотношение (8), вообще говоря, в точности выполняется лишь в случае АР модели второго порядка. Однако, поскольку спектр сигнала в районе резонансного пика может быть описан моделью второго порядка с ЛСЧ, близкими к $(f^{(k,1)},\ f^{(k,2)})$, такая аппроксимация вполне оправдана.

Отметим, что предложенная методика детектирования полюсных искажений фактически включает в себя детектор пауз. Этот факт весьма важен, поскольку во многих случаях (например, в дереверберационных задачах) влияние передаточной функции среды не проявляется в паузах. Отсутствие же резонансной структуры приводит к относительно равномерному распределению ЛСЧ. Следовательно, предлагаемый метод просто "не принимает во внимание" паузы (они составляют, как правило, не менее $40\div50$ % от общей продолжительности речевых сигналов). В противоположность этому, метод построения полюсных гистограмм, предложенный в [7,8], осуществляет поиск нулей знаменателя передаточной функции на всех фреймах без исключения. Это, безусловно, затрудняет детектирование влияния передаточной функции среды. Помимо этого, согласно предлагаемому нами методу построение полюсных гистограмм осуществляется, минуя процесс вычисления комплексных корней уравнений, что соответствует принципиально более низким вычислительным затратам.

Следует заметить, что в рамках предложенного алгоритма происходит автоматическая идентификация порядка искажающего воздействия m. При этом делается допущение, что значение m является меньшим порядка АР модели p, с которым ведется анализ наблюдаемого сигнала. Это допущение справедливо во многих задачах обработки речевых сигналов, в которых типичные значения p от 8 до 30 превосходят удвоенное количество основных посторонних резонансов. В задачах акустической деревербераций речевых сигналов, где порядок искажающего воздействия может быть значительно выше, более конструктивным решением является использование отдельных АР моделей сравнительно небольших порядков в каждом из относительно узких частотных поддиапазонов [8,13].

После того, как идентифицирована передаточная характеристика среды, необходимо восстановить исходный речевой сигнал. Поскольку свертка $s\otimes h$ является АР процессом порядка $(p+m)$, то к оцениванию такого сигнала в присутствии шума $v(n)$ может быть применен блочный фильтр Калмана (БФК), разработанный в работе [21]. Одна из базовых идей данного алгоритма – использование квантов АР параметров $\{\boldsymbol{a}_r, g_r\}$ (r – номер кванта), вычисленных заранее по сформированным с участием различных дикторов тестовым речевым массивам. На каждом фрейме зашумленно-

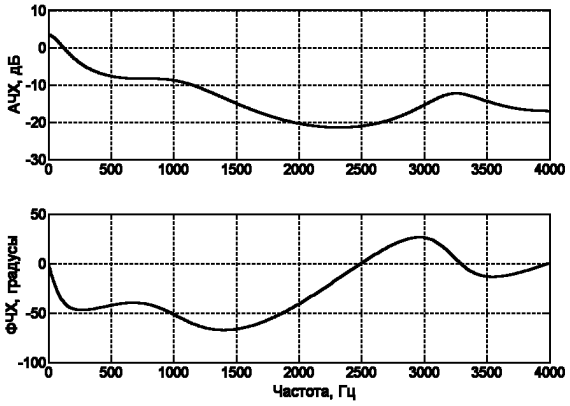

Рис. 3. Усредненный спектр помехи внутри салона автомобиля

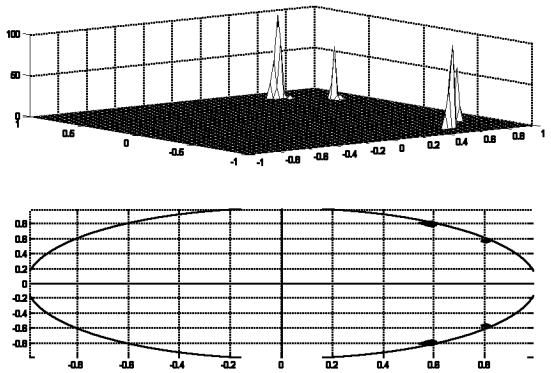

Рис. 4. Гистограмма распределения полюсов (предлагаемая методика)

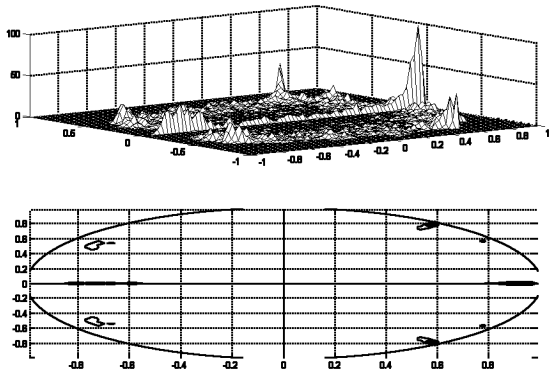

Рис. 5. Гистограмма распределения полюсов, построенная посредством нахождения всех нулей АР полиномов на каждом фрейме искаженного сигнала

го сигнала в качестве предварительной оценки АР параметров выбирался квант, максимизирующий записанный с учетом помехи блочный функционал правдоподобия (было показано, что для эффективного выполнения данной процедуры количество квантов, равное 16, является достаточным). Полученная таким образом оценка АР параметров уточнялась с помощью эффективной итерационной процедуры. Для восстановления сигнала на основе оцененных АР коэффициентов применялся алгоритм блочной калмановской фильтрации, обладающий преимуществом перед традиционными фильтрационными подходами как в контексте ошибки оценивания, так и по вычислительным затратам.

Покажем, как данная процедура блочной фильтрации может быть обобщена на случай искажений, характеризуемых формулой (1). Предварительно выполняется свертка квантов АР параметров $\{\boldsymbol{a}_r, g_r\}$ с коэффициентами знаменателя оцененной передаточной функции среды: $\boldsymbol{b}_r = \boldsymbol{a}_r \otimes \widehat{\boldsymbol{c}}$. Преобразованные таким образом кванты $\{\boldsymbol{b}_r, g_r\}$ будем называть модифицированными. Далее, на каждом фрейме выполняются следующие действия.

1. В качестве начального приближения $\{\boldsymbol{b}^{(0)}, g^{(0)}\}$ для АР коэффициентов свертки $s \otimes h$ выбирается модифицированный квант $\{\boldsymbol{b}_{r_0}, g_{r_0}\}$, максимизирующий блочный функционал правдоподобия.

2. Выполняется итерационное улучшение оценок $\{\boldsymbol{b}^{(0)}, g^{(0)}\}$. В результате получаем оценки АР коэффициентов $\{\boldsymbol{b}^{(1)}, g^{(1)}\}$.

3. Для восстановления свертки $s(n) \otimes h(n)$ к наблюдаемому сигналу $z(n)$ применяется БФК, основанный на значениях АР параметров $\{\boldsymbol{b}^{(1)}, g^{(1)}\}$.

4. Оценка искомого полезного сигнала $\widehat{s}(n)$ формируется путем пропускания полученной оценки свертки через фильтр с конечной импульсной характеристикой, имеющий коэффициенты $\widehat{c}_k$, $k = 0, 2, \ldots, m$.

Описанная процедура восстановления сигнала обеспечивает компенсацию передаточной функции среды и аддитивных фоновых помех, являясь эффективным средством решения проблемы усиления шума.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Предложенный алгоритм слепой деконволюции проверен экспериментально на моделях и на реальных сигналах.

В первой серии экспериментов в качестве исходного использовался речевой сигнал продолжительностью 6.75 с, произнесенный диктором-мужчиной. Этот сигнал подвергался искусственному искажающему воздействию АР фильтром 4-го порядка, определяемым полюсами $\gamma_{1,2}=0.99e^{\pm i\pi/5}$, $\gamma_{3,4}=0.995e^{\pm i3\pi/10}$. При $f_s=8000$ Гц данные полюса соответствуют внесению в сигнал посторонних резонансов с частотами 800 и 1200 Гц (подобные искажения свойственны некоторым аналоговым устройствам звукозаписи [8]). Модифицированный указанным образом сигнал был смешан с окрашенным шумом, аппроксимирующим помеху, записанную внутри салона движущегося автомобиля (АЧХ и ФЧХ ее спектральной огибающей приведены на рис. 3). При этом отношение сигнал/шум составляло 5 дБ.

Гистограмма распределения полюсов, построенная в соответствии с описанной методикой, представлена на рис. 4. Для порядка АР модели и длины фрейма были взяты значения $p=10$ и $L=256$ соответственно. На гистограмме полюсов четко проявлены пики, соответствующие посторонним резонансам, что обеспечивает их точную идентификацию. Для сравнения, на рис. 5 представлен результат применения к искаженному сигналу подхода, основанного на нахождении всех комплексных нулей АР полиномов на каждом фрейме искаженного сигнала [7, 8]. В гистограмме, полученной таким образом, велико влияние резонансов аддитивного шума. Это делает невозможным корректную идентификацию передаточной характеристики искажающего воздействия.

На рис. 6 приведено сопоставление АЧХ и ФЧХ исходного искажающего фильтра и результата его идентификации с помощью предложенной нами методики. Характеристики исходного фильтра представлены сплошными линиями, а идентифицированного – штриховыми. Как следует из рисунка, результат идентификации практически полностью совпадает с исходным фильтром.

Рассмотрим теперь результат применения разработанной методики к восстановлению старинной звукозаписи. Исходный аналоговый сигнал был предварительно оцифрован с частотой дискретизации $f_s=8000$ Гц. Анализ производился при порядке АР модели $p=10$ и длине фрейма $L=160$. Усредненный спектр помехи (фактическое

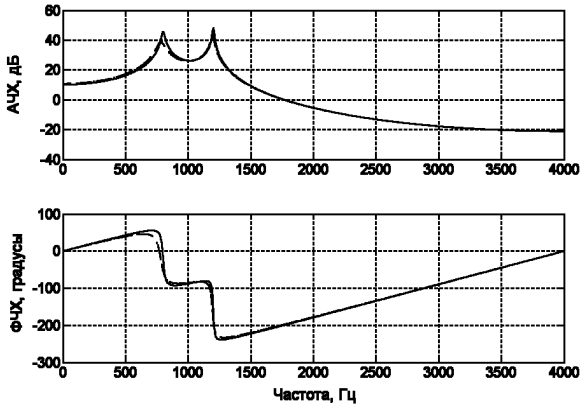

Рис. 6. Сопоставление АЧХ и ФЧХ исходного искажающего фильтра (сплошные) и результата его идентификации с помощью предлагаемой методики (штриховые)

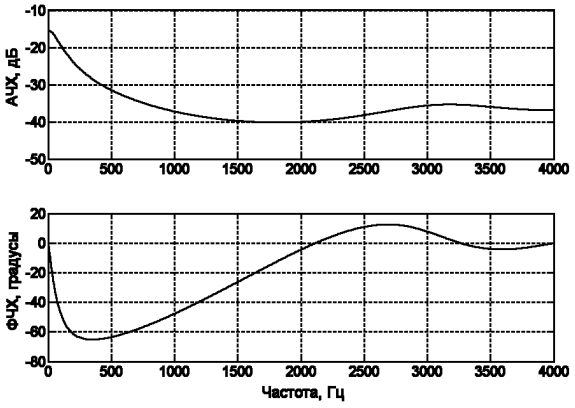

Рис. 7. Усредненный спектр помехи, характерной для старинной звукозаписи

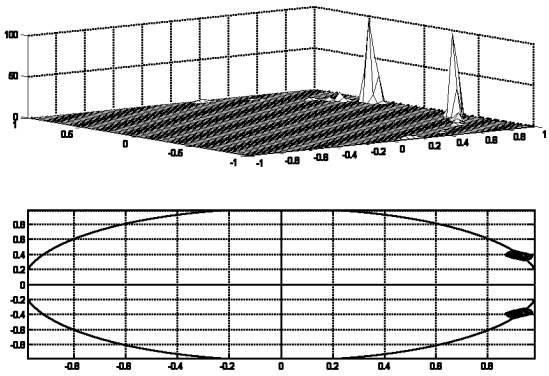

Рис. 8. Гистограмма распределения полюсов, соответствующая старинной звукозаписи

отношение сигнал/шум составляло 12.5 дБ) и гистограмма распределения полюсов представлены на рис. 7 и 8 соответственно. Из рис. 8 ясно видно, что в сигнале присутствуют стационарные резонансные полюса $z = 0.8 \pm 0.38i$ (т. е. на частоте $f_{\text{рез.}} = \text{arctg}\,(0.38/0.8)\, f_s/(2\pi) \approx 508$ Гц), определяющие передаточную функцию записывающего устройства.

Применение к рассмотренным сигналам предложенного обобщения алгоритма блочной калмановской фильтрации позволило устранить "металлическое" звучание, вызванное присутствием посторонних резонансов, и повысить субъективное качество сигнала вследствие существенного снижения уровня фонового шума.

ЗАКЛЮЧЕНИЕ

Рассмотрена задача одноканальной слепой деконволюции речевых сигналов в условиях присутствия фоновых шумов. Предложен эффективный метод детектирования и устранения влияния передаточной функции среды, основанный на анализе линейных спектральных частот искаженного сигнала.

Принципиальное преимущество предлагаемого метода состоит в адаптивном учете помехи в структуре алгоритма. Помимо повышения надежности получаемых результатов, это исключает возможность принятия окрашенного шума с сильно выраженной резонансной структурой за "полюсное" искажение, что было свойственно существующим методам одноканальной слепой деконволюции.

Предложена эффективная процедура локализации посторонних резонансов, основанная на анализе разностей линейных спектральных частот. Данная методика, в частности, отсеивает фреймы со слабо выраженной резонансной структурой и фактически включает в себя детектор речевой активности, что выгодно отличает ее от существующих методов слепого выравнивания. Создание метода компенсации передаточной функции среды, основанного на анализе линейных спектральных частот, особенно важно, поскольку их вычисление является неотъемлемой частью большинства современных систем цифровой обработки речевых сигналов.

С целью устранения эффекта "усиления шума" введена эффективная итерационная процедура, использующая блочный фильтр Калмана, которая обеспечивает компенсацию как передаточной функции среды, так и аддитивных фоновых помех. Эффективность результирующего метода слепой деконволюции проверена на искусственных и реальных искажениях речевых сигналов.

1. *Furui S.* Steps toward flexible speech recognition // Proc. 8-th Austral. Conf. SST-2000.– Canberra, 2000.– P. 19–29.
2. *Van Vuuren S.* Comparison of text-independent speaker recognition methods on telephone speech with acoustic mismatch // Proc. Int. Conf. ICSLP.– Philadelphia,1996.
3. *Forney G. D., Eyuboglu M. V.* Combined equalization and coding using precoding // IEEE Communic. Mag.– 1991.– **29**.– P. 25–34.
4. *Subramaniam S., Petropulu A. P., Wendt C.* Cepstrum-based deconvolution for speech dereverberation // IEEE Trans. Speech Audio Proces.– 1996.– **4**.– P. 392–396.
5. *Petropulu A. P., Nikias C. L.* Blind deconvolution using signal reconstruction from partial higher order cepstral information // IEEE Trans. Signal Proces.– 1993.– **41**.– P. 2088–2094.
6. *Стокхэм Т. Дж., Кэннон Т. М., Ингебретсен Р. Б.* Цифровое восстановление сигналов посредством неопределенной инверсной свертки // ТИИЭР.– 1975.– **4**.– С. 161–177.
7. *Hopgood J.* Blind deconvolution with application for reverberation cancellation in hearing aids (Final-year undergraduate project).– Cambridge: University of Cambridge, Dept Engng, 1997.– 50 p.
8. *Hopgood J.* Non-stationary signal processing with application to reverberation cancellation in acoustical environments (Ph. D. Thesis).– Cambridge: University of Cambridge, 2000.– 348 p.
9. *Astrom K. J., Hagander P., Sternby J.* Zeros of sampled systems // Automatica.– 1984.– **20**.– P. 31–38.
10. *Gray R. M., Buzo A., Gray A. H., Matsuyama Y.* Distortion measures for speech processing // IEEE Trans. Acoust. Speech Signal Proces.– 1980.– **28**.– P. 367–376.
11. *Haneda Y., Makino S., Kaneda Y.* Common acoustical pole and zero modeling of room transfer functions // IEEE Trans. Speech Audio Proces.– 1994.– **2**.– P. 320–328.
12. *Рабинер Л., Шафер Р.* Цифровая обработка речевых сигналов.– М.: Радио и связь, 1981.– 496 с.
13. *Маркел Дж., Грей А.* Линейное предсказание речи.– М.: Связь, 1977.– 308 с.
14. *Miyoshi M., Kaneda Y.* Inverse filtering of room acoustics // IEEE Trans. Acoust. Speech Signal Proces.– 1988.– **36**.– P. 145–152.
15. *Neely S. T., Allen J. B.* Invertibility of a room impulse response // J. Acoust. Soc. Amer.– 1979.– **65**.– P. 165–169.
16. *Lim J., Oppenheim A.* All-pole modeling of degraded speech // IEEE Trans. Acoust. Speech Signal Proces.– 1978.– **26**.– P. 197–210.
17. *Лим Дж. С., Оппенхайм А. В.* Коррекция и сжатие спектра зашумленных речевых сигналов // ТИИЭР.– 1979.– **12**.– С. 5-27.
18. *Hopgood J., Rayner P. J. W.* Bayesian single channel blind deconvolution using parametric signal and channel models // Proc. IEEE Workshop Appl. Signal Proces. Audio Acoust.– New York, 1999.– P. 151–154.

19. *Godsill S. J., Rayner P. J. W.* Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler // IEEE Trans. Speech Audio Proces.– 1998.– **6**.– P. 352–372.

20. *Семенов В. Ю.* Новый метод вычисления линейных спектральных частот речевых сигналов, основанный на универсальном алгоритме решения трансцендентных уравнений // Акуст. вісн.– 2002.– **5**, N 4.– С. 38–50.

21. *Калюжный А. Я., Семенов В. Ю.* Экономичный метод очистки речи от шума, основанный на блочном представлении сигнала в пространстве состояний и векторном квантовании // Акуст. вісн.– 2002.– **5**, N 3.– С. 28–34.

22. *Grassi S.* Optimized implementation of speech processing algorithms (Ph. D. Thesis).– Neuchatel: Universite de Neuchatel, 1998.– 211 p.

23. *Itakura F.* Line spectrum representation of linear predictive coefficients of speech signals // J. Acoust. Soc. Amer.– 1975.– **57**, N 1, Suppl. 1.– P. S35.

24. *Paliwal K. K., Atal B. S.* Efficient vector quantization of LPC parameters at 24 bits/frame // IEEE Trans. Speech Audio Proces.– 1993.– **1**.– P. 3–14.

25. *Paliwal K. K.* A study of line spectrum pair frequencies for speech recognition // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– New York, 1988.– P. 485–488.

26. *Liu C., Lin M.,Wang W., Wang H.* A study of line spectrum pair frequencies for speaker recognition // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– Alburquerque, 1990.– P. 277–280.

27. *Soong K. S., Juang B.-H.* Optimal quantization of LSP parameters // IEEE Trans. Speech Audio Proces.– 1993.– **1**.– P. 15–24.

28. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров.– М.: Наука, 1984.– 833 с.