

УДК 621.31:534.4

ИССЛЕДОВАНИЕ СВОЙСТВ ИСКАЖЕННЫХ РЕЧЕВЫХ СИГНАЛОВ

А.В.Кельманов, А.Г.Хайретдинова

В в е д е н и е

В большинстве существующих систем распознавания речи задача выделения признаков речевого сигнала решается путем аппроксимации сигнала моделью (например, при помощи параметров авторегрессии, преобразования Фурье и т.п.) с использованием критериев, которые в конечном итоге сводятся к минимуму среднеквадратической ошибки. К проверке этого же критерия фактически сводится сравнение контрольной реализации с эталонами. Как показывают результаты последних лет, подобный подход к выделению признаков (спектральных или временных) позволяет распознавать речь (изолированные и слитно произнесенные команды при настройке на диктора) с надежностью, близкой к 100%. Однако этот подход обладает весьма серьезным недостатком, состоящим в том, что после обучения системы распознавания нелинейные искажения в тракте связи становятся недопустимыми. Если же эти искажения имеют место, то, в отличие от верного восприятия речи человеком, надежность системы распознавания падает до такого уровня, что эта система становится практически непригодной [1]. Описанная ситуация часто встречается при работе устройств распознавания на борту различных технических систем.

Главная причина резкого снижения надежности распознавания после искажений сигнала состоит в том, что признаки, которые

используются в системах распознавания, не обладают свойством инвариантности к искажениям. Инвариантные же признаки могут быть найдены лишь после исследования свойств искаженных сигналов. Поэтому, с одной стороны, целью настоящей работы является качественная оценка разборчивости искаженной речи. С другой стороны, поскольку для систем распознавания речи одним из ключевых моментов является оценивание признаков, представляет интерес качественное и количественное исследование свойств этих оценок после искажений. В данной работе предметом исследования является поведение оценок спектральной плотности и параметров авторегрессии.

§1. Два класса нелинейных искажений

К наиболее распространенным искажениям относятся нелинейные неинерционные преобразования амплитуды сигнала в интервале его знакопостоянства в том числе такие, которые речевой сигнал, как непрерывную функцию времени (или ее отсчеты в дискретном случае), переводят в кусочно-непрерывную функцию [2-7]. Зачастую эти искажения неоднозначны, заранее неизвестны, носят случайный характер и, таким образом, необратимы. В этих условиях задача поиска признаков, инвариантных к нелинейным искажениям в тракте связи, приобретает особую актуальность. Следует отметить, что случай, когда вид искажений заранее известен и они обратимы, тривиален и не представляет особого интереса, поскольку в этом случае помеха может быть легко компенсирована путем обратных преобразований.

Очевидно, что инвариантные признаки могут быть найдены конструктивным способом лишь для определенного класса нелинейных искажений. Представляется естественным на первом этапе ограничить этот класс только такими искажениями, которые сохраняют разборчивость речи. Таким образом, возникает задача поиска

и формального описания класса допустимых искажений сигнала, сохраняющих разборчивость.

Пусть $x(t)$, $-\infty < t < \infty$, - непрерывный речевой сигнал (случайный процесс с нулевым матожиданием и ограниченной дисперсией), а $x_n = x(nT)$, $n = 0, \pm 1, \pm 2, \dots$, - отсчеты этого сигнала через равные промежутки времени T . Будем считать, что на коротких участках сигнала длительностью T_a , содержащих N отсчетов, $x_n \in P_x$, $n = 1, 2, \dots, N$, т.е. что каждый отсчет сигнала является выборочным значением из одномерного распределения P_x . Под искажением будем понимать нелинейное преобразование $y(t) = f[x(t)]$ или $y_n = f(x_n)$. Класс детерминированных искажений \mathcal{F}_0 , охватывающий широкий спектр нелинейных помех, встречающихся на практике, может быть описан в виде:

$$\mathcal{F}_0 = \{f: f(x) \leq \text{const}, x=0 \Rightarrow [f(x)=0] \forall [f'(x)=c\delta(x)]\},$$

$$f(x) > 0 \Leftrightarrow x > 0, f(x) < 0 \Leftrightarrow x < 0, |c| = \text{const}\}. \quad (1)$$

Из определения (1) видно, что в класс \mathcal{F}_0 включаются всевозможные искажения (преобразования) сигнала, которые имеют в качестве неподвижных либо множество точек $\{x(t) = 0\}$ таких, что на каждом полуоткрытом интервале знакопостоянства имеется только одна неподвижная точка $x(t) = 0$, либо скачки в момент перехода сигнала через нулевой уровень. В этот класс искажений входят полиномиальные искажения вида:

$$f(x) = \sum_{i=1}^k a_i x^{2i-1}, \quad (2)$$

которыми можно аппроксимировать многие из реальных искажений.

В [3] показано, что для речевых трактов связи характерными искажениями являются клиппирование (см. §2) и преобразование $y = ax^3$. В работах [3-7] проведены детальные исследования свойств искажений вида (2). При этом показано [3,4],

что дискретный спектр искаженной речи изменяется "несущественно" (сохраняется "форма" спектра) и разборчивость речи не ухудшается, а если и ухудшается, то весьма незначительно. Возникают вопросы: 1) останется ли разборчивым сигнал, если он претерпевает произвольные (случайные) искажения амплитуды в интервалах знакопостоянства? 2) почему для искаженных сигналов, сохраняющих разборчивость и имеющих спектр, несущественно отличный от исходного, надежность распознавания значительно ниже, чем для неискаженного? 3) какова количественная оценка несущественных изменений спектральной картины в терминах, используемых в системах распознавания?

Для ответа на эти вопросы необходимо рассмотреть более широкий класс стохастических искажений, который можно представить в виде:

$$\mathcal{F}_1 = \{f: x \geq 0 \Leftrightarrow f(x) \in P^+, x < 0 \Leftrightarrow f(x) \in P^-\}. \quad (3)$$

Здесь P^+ и P^- есть некоторые искажающие одномерные распределения, отличные от P_x . Очевидно, что $\mathcal{F}_0 \subset \mathcal{F}_1$.

Класс \mathcal{F}_1 содержит такие искажения речевого сигнала, которые получаются путем подмены истинного одномерного распределения. В определении (3) разбиение искажений по участкам знакопостоянства на первый взгляд выглядит слишком условным. Однако это сделано специально, поскольку, с одной стороны, уже давно известно [2,3], что в речевом сигнале большая часть информации содержится в моментах его перехода через нулевой уровень, а с другой - потому, что подобное разбиение удобно при моделировании искажений.

При тестировании искаженных сигналов на разборчивость в дополнение к известным [2-9] детерминированным способам "восстановления" речи во временной области был опробован еще ряд способов, которые в соответствии с введенными определениями (1) и (3) можно условно разделить на два класса: детерминированные и стохастические. Под "восстановлением" здесь понимает-

ся доопределение (по выбранному закону) значений сигнала в интервале между его нулевыми пересечениями. По существу, "восстановленный" таким образом сигнал является искаженным речевым сигналом. Ниже приведены модели детерминированных и стохастических искажений.

§2. Примеры искажений

2.1. Детерминированные искажения. Под клиппированием обычно понимается следующее нелинейное преобразование:

$$y(t) = \begin{cases} a_1, & \text{если } x(t) \geq 0, \\ -a_2, & \text{если } x(t) < 0. \end{cases} \quad (4)$$

Здесь $a_1 > 0$ и $a_2 > 0$ - постоянные (часто $a_1 = a_2 = a$).

Известно, что клиппирование речевого сигнала практически не снижает разборчивость, если только моменты пересечения нуля сигналом измеряются с точностью не менее 0.1 мс [2]. В то же время клиппирование, как кусочно-постоянное необратимое искажение, является серьезной (встречающейся в трактах связи довольно часто) помехой для систем распознавания, в которых в качестве признаков не используются интервалы между переходами сигнала через нулевой уровень.

Пусть $\left[-\frac{T}{2}, \frac{T}{2}\right]$ - какой-либо интервал знакопостоянства речевого сигнала. Для определенности пусть $x(t) \geq 0$. На этом интервале рассмотрим следующие искажения (подмена истинных значений):

$$y(t) = a(\tau) \cos \frac{\pi t}{\tau}, \quad -\frac{T}{2} \leq t < \frac{T}{2}, \quad (5)$$

$$y(t) = a(\tau) \left[1 - \left(\frac{2t}{\tau} \right)^2 \right], \quad -\frac{T}{2} \leq t < \frac{T}{2}, \quad (6)$$

$$y(t) = a(\tau) e^{-\alpha \tau^2} \cos \frac{\pi t}{\tau}, \quad -\frac{\tau}{2} \leq t < \frac{\tau}{2}. \quad (7)$$

Формула (5) описывает косинусоидальные искажения, (6) - параболические, (7) - экспоненциально-косинусоидальные. В зависимости от длины интервала знакопостоянства τ амплитуда $a(\tau)$ может изменяться, например, по линейному, квадратичному, показательному и другим законам, а также быть постоянной. Кроме того, возможна "стыковка" соседних интервалов знакопостоянства из условия гладкости искаженного сигнала.

В некоторых трактах связи может возникать мультипликативная помеха. Например,

$$y(t) = x(t) \frac{b \cos \frac{\pi t}{\tau_n} + c}{b+c}, \quad \begin{matrix} 0 < b < c, \\ -\infty < t < \infty. \end{matrix} \quad (8)$$

Интересным с практической точки зрения является преобразование

$$y(t) = \begin{cases} 2^{\lfloor \log_2 x(t) \rfloor}, & \text{если } x(t) \geq 1, \\ -2^{\lfloor \log_2 -x(t) \rfloor}, & \text{если } x(t) \leq -1, \\ 0, & \text{если } |x(t)| < 1. \end{cases} \quad (9)$$

Это преобразование принадлежит классу \mathcal{F}_1 и не принадлежит классу \mathcal{F}_0 , так как в точки $y(t)=0$ отображается интервал значений исходного сигнала. Но при цифровой обработке сигналов, когда $x(t)$ принимает лишь целочисленные значения, преобразование (9) включается в класс искажений \mathcal{F}_0 .

2.2. Стохастические искажения. В большинстве случаев при цифровой обработке речи для адекватного представления искаженного речевого сигнала как случайного процесса оказывается

достаточным использовать лишь одномерное распределение вероятностей и ковариационную функцию (или спектральную плотность). Известно, что выборочные оценки одномерного распределения вероятностей речевого сигнала хорошо аппроксимируются гамма-распределением и несколько хуже распределением Лапласа [10]. Тем не менее во всех без исключения известных методах выделения признаков используется лишь гауссовская аппроксимация сигнала, а чаще всего вид распределения вообще не учитывается. Правомерность подобного подхода к обработке неискаженной речи практически не обсуждается. В связи с этим в работе исследуются свойства речевого сигнала, искаженного путем замены истинного распределения вероятностей на ряд других распределений.

В качестве первого примера случайных искажений рассмотрим следующее преобразование:

$$y_n = \begin{cases} y_n^+ \in U(0, a), & \text{если } x_n \geq 0, \\ y_n^- \in U(-a, 0), & \text{если } x_n < 0. \end{cases} \quad (10)$$

Здесь $U(0, a)$ обозначает равномерное распределение на интервале $[0, a)$.

Другим примером стохастических искажений является преобразование

$$y_n = \begin{cases} y_n^+ \in \Phi(a, \sigma^2), & \text{если } x_n \geq 0, \\ y_n^- \in \Phi(-a, \sigma^2), & \text{если } x_n < 0. \end{cases} \quad (11)$$

Здесь $\Phi(a, \sigma^2)$ обозначает нормальное распределение с параметрами a и σ^2 .

Следующий пример искажений:

$$y_n = \begin{cases} y_n^+ \in \Phi_{\text{mod}}(0, \sigma^2), & \text{если } x_n \geq 0, \\ y_n^- \in -\Phi_{\text{mod}}(0, \sigma^2), & \text{если } x_n < 0. \end{cases} \quad (12)$$

Здесь символом $\Phi_{\text{mod}}(0, \sigma^2)$ обозначено распределение модуля гауссовской случайной величины, для которой параметрами распределения являются 0 и σ^2 . Символ $-\Phi_{\text{mod}}(0, \sigma^2)$ соответственно обозначает распределение отрицательных значений модуля гауссовской случайной величины.

Наконец, последний пример:

$$y_n = \begin{cases} y_n^+ \in \Gamma_{a,1}, & \text{если } x_n \geq 0, \\ y_n^- \in \Gamma_{-a,1}, & \text{если } x_n < 0. \end{cases} \quad (13)$$

В этой формуле через $\Gamma_{a,1}$ обозначено экспоненциальное распределение с параметром a .

§3. Экспериментальные результаты

Для проведения экспериментов речевой сигнал $x(t)$ вводился в ЭВМ через 8-разрядный АЦП при частоте квантования 10 кГц ($T = 0.1$ мс). В экспериментах участвовало 2 диктора - мужчина и женщина.

3.1. Субъективные оценки разборчивости. В качестве речевого материала использовался фразарь, состоящий из 39 тестовых фраз. Начитанный и оцифрованный неискаженный речевой сигнал вводился в ЭВМ. Для получения искаженного сигнала исходный сигнал разбивался на примыкающие по времени интервалы (сегменты или интервалы анализа) фиксированной длительности $T_a = 25.6$ мс, содержащие по $N = 256$ отсчетов. На каждом таком сегменте отыскивались интервалы знакопостоянства, сигнал искажался в ЭВМ в соответствии с формулами (4)-(13) и выводился через ЦАП на прослушивание четырех аудиторов.

Для детерминированных искажений (5)-(7) амплитуда сигнала в интервалах знакопостоянства была постоянной, а также изменялась по линейному, квадратичному, показательному законам в зависимости от длительности этих интервалов. Осуществлялась "стыковка" соседних участков знакопостоянства из условия гладкости искаженного сигнала. При моделировании мультипликативных искажений (8) варьировались постоянные b, c и τ_n . При клиппировании по формуле (4) $a_1 = a_2 = a$. Для моделирования стохастических искажений (10)-(13) применялись датчики случайных чисел. Значения параметров распределений приведены в следующем разделе.

Комбинируя параметры искажающих функций, можно получить достаточно много разнообразных искажений. В общей сложности на предмет разборчивости было исследовано около 50 таких искажений, в том числе с применением медианной фильтрации при окнах из трех, пяти и семи отсчетов. Во всех случаях получены разборчивые сигналы, более глухие для детерминированных искажений вида (5)-(7) и более звонкие, напоминающие речь в шумах, при стохастических искажениях (10)-(13) и клиппировании (4). Всеми аудиторами отмечено, что замена истинного одномерного распределения вероятностей на другие распределения приводит к искаженным сигналам, которые между собой имеют едва различимые на слух оттенки, т.е. окраска сигнала практически не зависит от вида одномерного распределения генерировавшегося искаженного случайного процесса.

При моделировании мультипликативной помехи (8) путем изменения параметров b, c и τ_n были получены как сигналы, которые на слух не отличались от исходного, так и сигналы с ощутимыми искажениями. Любопытно то, что женский голос удалось трансформировать в голос, похожий на мужской с характерной охриплостью. Все остальные искажения (детерминированные и стохастические) практически не изменяют особенностей речи диктора.

Наконец, искажения (9) не дали заметных отличий от исходного сигнала.

3.2. Оценки искажений признаков. В этом разделе дана как качественная и количественная оценка изменений признаков речевого сигнала после нелинейных искажений сигнала. Степень искажения признаков оценивается во временной и спектральной областях. При этом во временной области рассматривается поведение оценок параметров процесса авторегрессии (на примере частных корреляций), аппроксимирующего искаженные и неискаженные сигналы, а в спектральной - поведение оценок спектральной плотности этого процесса.

На рис.1-2 в логарифмическом масштабе приведены графики оценок авторегрессионной спектральной плотности для фонемы /а/ (мужской и женский голоса) при длительности интервала анализа $T_a = 25.6$ мс (длина выборки $N = 256$). Порядок аппроксимирующей модели равен 12. Аналогичные данные для фонемы /ш/ приведены на рис.3-4. На этих же рисунках рядом с каждым графиком приведены по 2 значения среднеквадратической невязки. Верхнее значение есть невязка между частными корреляциями, а нижнее - между нормированными авторегрессионными спектральными плотностями. Для исходного неискаженного сигнала (первый график на каждом рисунке) эти невязки подсчитывались между двумя реализациями одного и того же случайного процесса, описывающего данную фонему. На остальных графиках вместо абсолютных значений невязки приведены коэффициенты, показывающие, во сколько раз эта невязка увеличивается для данного искажения.

Данные на рис.1-4 получены при следующих значениях параметров искажающих функций: для клипированного сигнала $a_1 = a_2 = 64$; для равномерного датчика случайных чисел (формула (10)) $a = 64$. При искажении гауссовским датчиком (формула (11)) использовались следующие параметры: $a = 64$, $3\sigma = 127$; для модуль-гауссовского распределения (формула (12)) $3\sigma =$

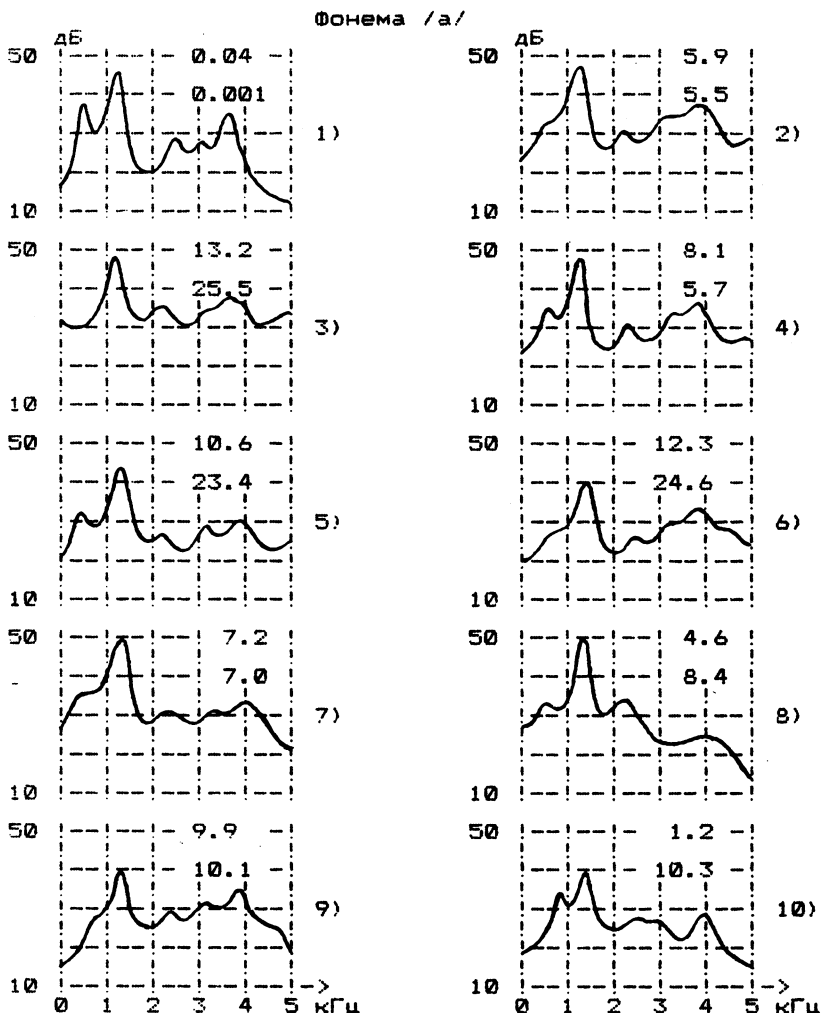


Рис.1. Оценки авторегрессионной спектральной плотности исходного и искаженных речевых сигналов. Диктор мужчина; 1) исходный сигнал, 2) - клиппированный, 3) искаженный равномерным датчиком случайных чисел, 4) гауссовским, 5) модуль-гауссовским, 6) экспоненциальным; 7) косинусоидальное искажение, 8) параболическое, 9) экспоненциально-косинусоидальное, 10) мультипликативное.

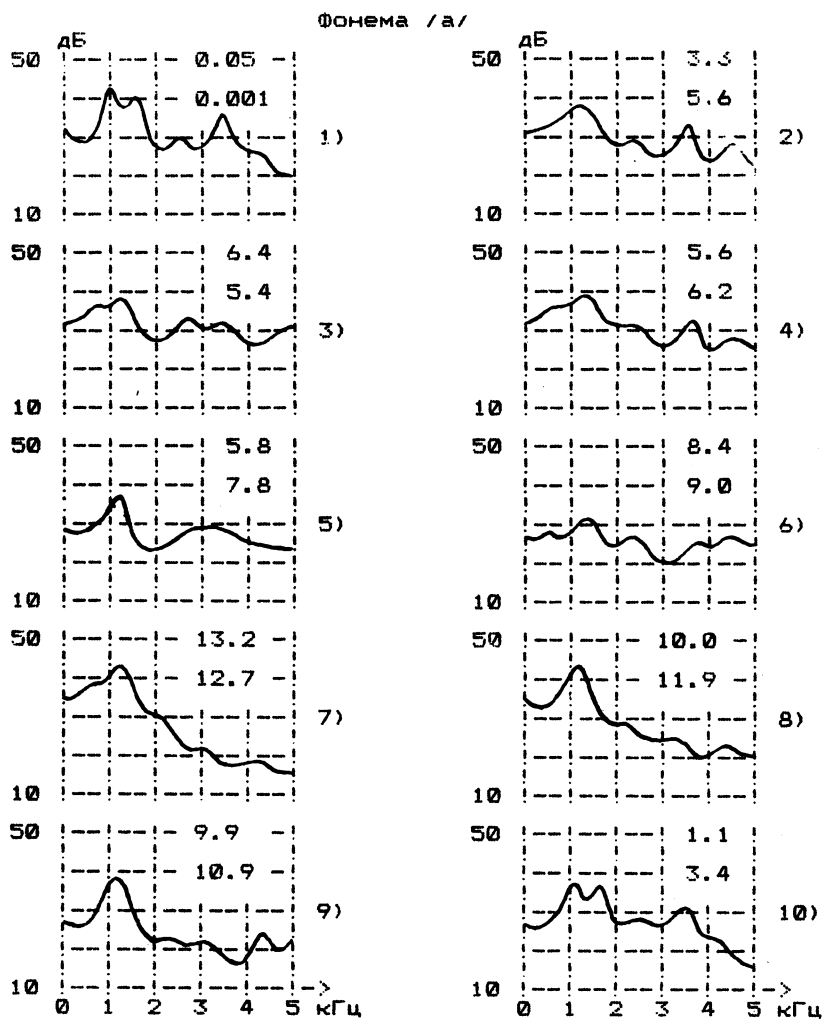


Рис. 2. Оценки авторегрессионной спектральной плотности исходного и искаженных речевых сигналов. Диктор женщина; 1) исходный сигнал, 2) - клиппированный, 3) искаженный равномерным датчиком случайных чисел, 4) гауссовским, 5) модуль-гауссовским, 6) экспоненциальным; 7) косинусоидальное искажение, 8) параболическое, 9) экспоненциально-косинусоидальное, 10) мультипликативное.

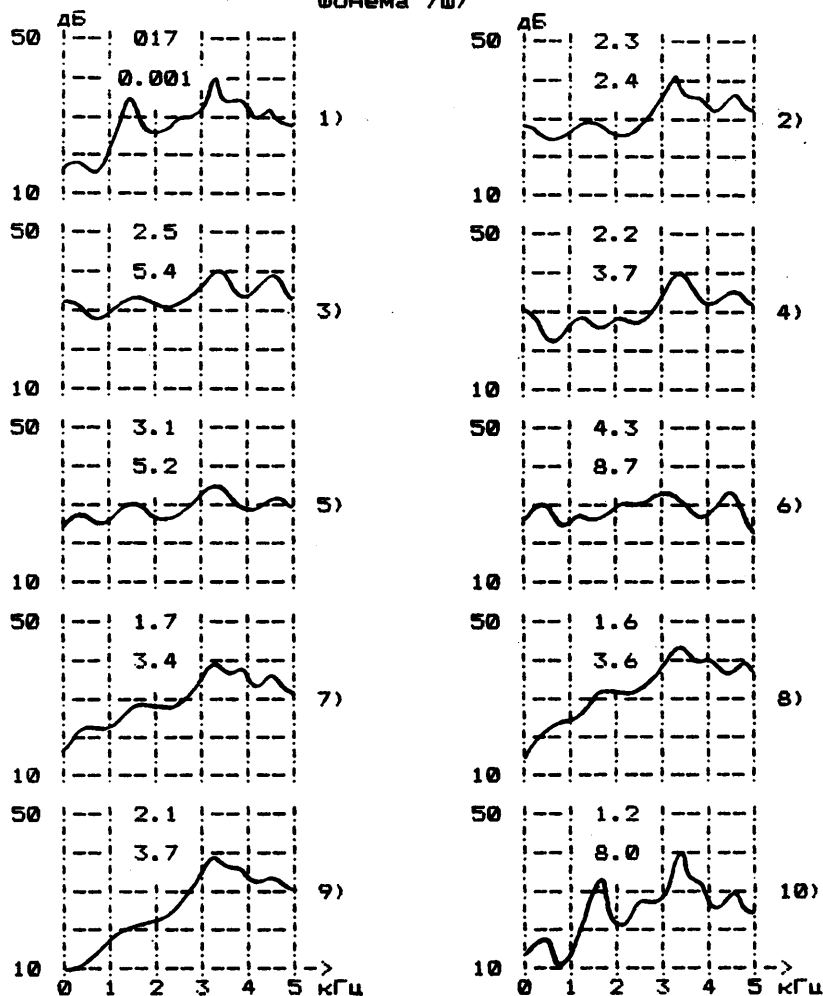


Рис. 3. Оценки авторегрессионной спектральной плотности исходного и искаженных речевых сигналов. Диктор мужчина; 1) исходный сигнал, 2) - клиппированный, 3) искаженный равномерным датчиком случайных чисел, 4) гауссовским, 5) модуль-гауссовским, 6) экспоненциальным; 7) косинусоидальное искажение, 8) параболическое, 9) экспоненциально-косинусоидальное, 10) мультипликативное.

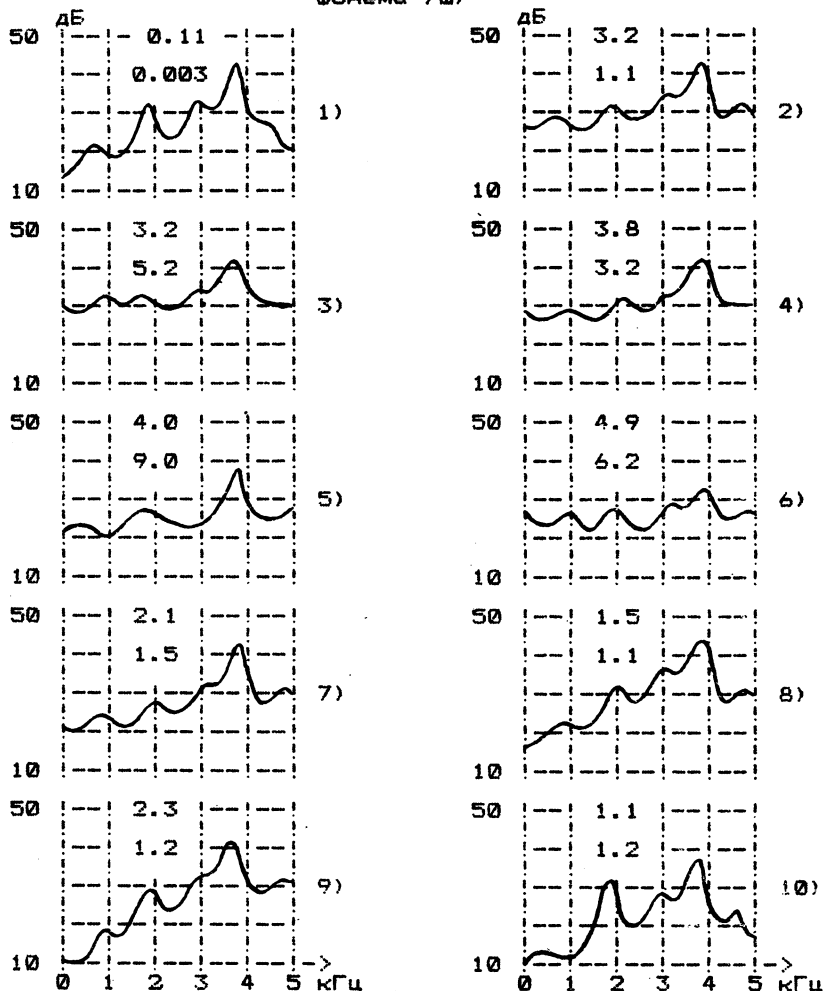


Рис. 4. Оценки авторегрессионной спектральной плотности исходного и искаженных речевых сигналов. Диктор женщины; 1) исходный сигнал, 2) - клипированный, 3) искаженный равномерным датчиком случайных чисел, 4) гауссовским, 5) модуль-гауссовским, 6) экспоненциальным; 7) косинусоидальное искажение, 8) параболическое, 9) экспоненциально-косинусоидальное, 10) мультипликативное.

$= 127$; а для экспоненциального (формула (13)) $3/a = 127$. При моделировании косинусоидальных и параболических искажений (формулы (5) и (6)) амплитуда сигнала на последующем интервале знакопостоянства вычислялась, исходя из условия "гладкости". Для экспоненциально-косинусоидальной помехи (7) $\alpha = 1$, а для мультипликативной $b = 10000$, $c = 9999$, $\tau_n = T_a/2$.

§4. Обсуждение результатов. Выводы

Прежде чем приступить к анализу результатов экспериментов, напомним, что главная цель данных исследований состоит в том, чтобы продвинуться по пути создания признаков, инвариантных к нелинейным искажениям сигнала. Причем к таким искажениям, вид которых заранее неизвестен, но известно, что они принадлежат заданному классу.

Класс нелинейных искажений \mathcal{F}_1 , рассматриваемый в этой работе, шире класса \mathcal{F}_0 , исследованию которого посвящено достаточно много работ. В этих работах ставилась цель получения количественных оценок разборчивости и качественных оценок изменения спектральных свойств сигнала. В данной работе цель другая. Поэтому здесь вместо количественных оценок разборчивости искаженной речи приводятся количественные оценки изменений спектральных и временных признаков и качественные оценки разборчивости.

Как уже упоминалось выше, класс \mathcal{F}_1 включает всевозможные случайные искажения сигнала, оставляющие неподвижными точки перехода неискаженного сигнала через нулевой уровень. Ограничение на неподвижность этих точек сделано не случайно, а в силу того, что, во-первых, во многих исследовательских работах указывается их исключительная информативность: незначительные сдвиги этих точек заметно снижают разборчивость речи. Во-вторых, потому, что в большинстве существующих трактов связи (микрофон - устройство распознавания) искажения таковы, что речь по-

сле этих искажений остается разборчивой (иногда эти искажения на слух незаметны), а процент ошибок распознавания резко возрастает и, в-третьих, потому, что данное несущественное ограничение позволяет естественным образом пополнить класс \mathcal{F}_0 до \mathcal{F}_1 .

Стохастические искажения, введенные и исследованные в этой работе, являются весьма сильными искажениями, так как дискретный речевой сигнал на каждом интервале знакопостоянства заменяется последовательностью независимых одинаково распределенных случайных величин. Хотя эти искажения и вносят дополнительные точки перехода сигнала через нулевой уровень, речь все же остается разборчивой. Поскольку вид одномерного распределения (т.е. характер изменения сигнала внутри интервала знакопостоянства) не влияет на уровень разборчивости, а незначительные смещения моментов перехода снижают этот уровень, из вышеизложенного и результатов исследований следует

УТВЕРЖДЕНИЕ 1. *Последовательность моментов перехода речевого сигнала через нулевой уровень является минимальным набором данных для восстановления разборчивой речи. Дальнейшее сокращение избыточности на сигнальном уровне невозможно.*

Ранее было известно более слабое утверждение, указывающее лишь на важность этих точек в информационном смысле.

Из результатов моделирования по формулам (10)-(13) вытекает зависимость от времени (на интервалах анализа, превышающих один интервал знакопостоянства) одномерной и двумерной функции распределения усгенерированных разборчивых сигналов. Так как традиционные интервалы анализа (на которых выделяются признаки) содержат множество интервалов знакопостоянства, справедливо

УТВЕРЖДЕНИЕ 2. *На традиционных интервалах анализа речи, содержащих более одного интервала знакопостоянства,*

ства, представление речевого сигнала как нестационарного случайного процесса с изменяющимся во времени одномерным равномерным, гауссовским, модуль-гауссовским и экспоненциальным распределением адекватно.

В этом утверждении представляется самым главным, что оно фактически обосновывает возможность гауссовской аппроксимации речевого сигнала.

УТВЕРЖДЕНИЕ 3. *Текстовая информация в речи полностью определяется моментами перехода речевого сигнала через нулевой уровень.*

Чтобы еще раз подтвердить это утверждение, был проведен следующий эксперимент. На прослушивание выводился сигнал $\tilde{y}_n = y_n - y_{n-1}$, $n = 1, 2, \dots$; где y_n - клиппированный сигнал. Результатом подобного преобразования является последовательность одиночных импульсов (или δ -импульсов), сменяющих знак в моменты перехода речевого сигнала через нулевой уровень. В интервалах между импульсами преобразованный сигнал \tilde{y}_n равен нулю. Полученный таким образом сигнал на слух практически ничем не отличался от клиппированного или от сигналов, полученных при помощи датчиков случайных чисел.

Наконец, из приведенных выше утверждений вытекает

УТВЕРЖДЕНИЕ 4. *Для построения признаков, инвариантных к нелинейным искажениям из классов \mathcal{F}_0 , \mathcal{F}_1 и к другим искажениям, сохраняющим неподвижными точки перехода неискаженного сигнала через нулевой уровень, необходимо и достаточно использовать последовательность этих точек.*

Перейдем теперь к выявлению причин снижения надежности распознавания искаженной, но разборчивой речи. Для этого обратимся к данным, приведенным на рис. 1-4. Качественный анализ графиков оценок спектральной плотности неискаженного и искаженных сигналов показывает заметные изменения в поведении оценок,

хотя речь при этом остается разборчивой. Численные данные (связки) показывают, что в зависимости от искажающей функции среднеквадратическое расстояние между исходным и искаженным сигналами в пространстве признаков может практически не измениться, но может и увеличиться почти в 26 раз. Эти же данные свидетельствуют о том, что степень увеличения расстояния различна для отдельных фонем и зависит от вида искажений. В результате для искаженных сигналов некоторые контрольные реализации оказываются ближе к чужим (для каждого искажения разным) эталонам, чем к своим. В этом состоит главная причина резкого увеличения ошибок распознавания (40% для словаря из 200 слов [1]).

Если вспомнить, что среднеквадратический критерий лежит в основе правил принятия решения практически всех систем распознавания речи, становится очевидным

УТВЕРЖДЕНИЕ 5. Принятие решения по минимуму среднеквадратической ошибки (или по однозначным связанным с ней критериям) в пространстве классических корреляционно-спектральных признаков неадекватно отражает процесс восприятия речи.

Ценой за среднеквадратическую близость эталонных и контрольных реализаций в пространстве классических признаков является неинвариантность систем распознавания к нелинейным искажениям. Выход из данной ситуации лежит либо в замене критерия близости, либо в переходе к другим математическим моделям сигнала.

Таким образом, для построения систем распознавания, инвариантных к нелинейным искажениям, необходим принципиально отличный от существующего подход. По результатам исследований можно сделать вывод, что этот подход должен основываться на распознавании речи по последовательности интервалов знакопо-

стоянства сигнала. Хотя ранее предпринимались попытки задействовать интервалы знакопостоянства в лабораторных системах распознавания речи [11,12], в настоящее время ни одна из действующих промышленных систем распознавания не использует эти интервалы в качестве признаков. Невысокие показатели надежности у подобных лабораторных систем распознавания ограничили их практическую пригодность и привели к преждевременному свертыванию исследований в данном направлении. Представляется, однако, что это направление далеко не исчерпало себя. Поэтому для решения проблемы инвариантности признаков к нелинейным искажениям необходимы дальнейшие исследования. В качестве первоочередного направления исследований следует указать на установление функциональной связи между моментами перехода речевого сигнала через нулевой уровень и классическими спектрально-временными признаками.

Наконец, заметим, что при построении систем распознавания речи в шумах задача фильтрации помех может быть сведена к поиску моментов изменения знака сигнала. И, вероятно, алгоритмы оценивания этих моментов будут существенно проще традиционных алгоритмов оптимальной фильтрации и оценивания. При этом, однако, придется исследовать вопросы устойчивости оценок моментов перехода, а также более детально оценить влияние на разборчивость и в особенности на надежность распознавания речи таких возможных эффектов алгоритмов обработки, как появление лишних переходов, пропадание моментов перехода и смещение их оценок от истинных значений.

Л и т е р а т у р а

1. КЕЛЬМАНОВ А.В. Первичная обработка сигналов в системах распознавания речи. //Настоящий сборник. -С. 96-130.
2. ФЛАНАГАН Дж. Анализ, синтез и восприятие речи: Пер. с англ. /Под ред. А.А.Пирогова. -М.: Связь, 1986. - 396 с.
3. САПОЖКОВ М.А. Речевой сигнал в кибернетике и связи. -М.: Связьиздат, 1963. - 450 с.

4. РЕПИНА О.И. Искажения в телефонном тракте. - М.: Связь, 1978. - 174 с.
5. ЛЕВИН Б.Р. Теоретические основы статистической радиотехники. Т.1. - М.: Сов.радио, 1974. - 552 с.
6. ГОРЯИНОВ В.Т., ЖУРАВЛЕВ А.Г., ТИХОНОВ В.И. Статистическая радиотехника. -М. Сов.радио, 1980. - 544 с.
7. ТИХОНОВ В.И. Нелинейные преобразования случайных процессов. -М.: Радио и связь, 1986. - 296 с.
8. ВАЙНШТОК А.П. Информационные компоненты речевого сигнала //Тез.докл. и сообщ. 12-го Всесоюз. семинара по автоматическому распознаванию слуховых образов (АРСО-12), Киев-Одесса, сентябрь 1982. - Киев, 1982. -С. 72-94.
9. ВИНЦЮК Т.К., ДЫС Л.И., КУЛЯС А.И. Экономное представление сигнала в памяти ЭВМ //Там же, с. 74-75.
10. РАБИНЕР Л., ШАФЕР Р. Цифровая обработка речевых сигналов. - М.: Радио и связь, 1981. - 495 с.
11. ЦЕМЕЛЬ Г.И. Опознавание речевых сигналов. - М.: Наука, 1971. - 148 с.
12. БУСЛЕНКО Н.П., ДЕЕВ В.В., ФРОЛОВ Г.Д. Система автоматического речевого общения с ЭВМ //Программирование, - 1977. -№3. - С. 79-83.

Поступила в ред.-изд.отд.

2 августа 1990 года