АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
(Вычислительные системы)
1991 год Выпуск 141

УДК 621.31:534.4

# ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ, ИСКАЖЕННЫХ ВИБРОПОМЕХАМИ

А.В.Кельманов, А.Г.Хайретдинова, С.А.Хамидуллин

## В в е д е н и е

Известны ситуации, когда голосовой ввод информации в систему распознавания сопровождается помехой, состоящей в воздействии вибраций на оператора и его речеобразующий тракт. Подобные помехи могут возникнуть при работе устройств распознавания на борту различных технических систем, например, в кабине автомобиля. Представляется естественным, что для обработки речи в условиях вибропомех потребуются специальные методы и алгоритмы. Однако к настоящему времени о воздействии вибропомех на диктора-оператора практически ничего не известно. Не описаны даже на качественном уровне изменения характеристик речи, возникающие из-за виброколебаний. Отсутствуют сведения о стабильности известных методов выделения признаков по отношению к данному классу помех. Открытым остается вопрос о степени влияния этих помех на надежность распознавания отдельных фонем, изолированных слов и слитной речи. И наконец,в литературе никак не отражены средства компенсации помех типа вибраций.

Цель данной работы заключается в выявлении характерных особенностей речи в условиях вибраций, построении модели вибропомех,оценивании влияния виброколебаний на устойчивость оценок

спектральной плотности речевого сигнала и на надежность распознавания речи.

## 1. Модель вибропомех

Как уже упоминалось во введении, о воздействии вибропомех на речеобразующий тракт практически ничего не известно. Для получения хорошей модели, вероятно, потребуются обширные физиологические исследования. Однако на первом этапе можно ограничиться построением искажающей функции, которая трансформирует обычную речь в такую, которая воспринимается на слух как речь в условиях виброколебаний.

1.1. Качественный анализ реальной виброречи. Чтобы подобрать подходящую искажающую функцию, необходимо найти характерные особенности виброречи. Поэтому для составления качественной картины речи при наличии вибраций проводились следующие исследования. Производилась запись фонограмм при воздействии на дикторов реальных вибропомех. Реальными считаются помехи, полученные на вибростенде либо в ходе эксплуатации соответствующего технического устройства. В результате по трем специально составленным словарям ((объемом 16,10 и 8 слов) семью дикторами-мужчинами были сформированы фонограммы при частотах виброколебаний 4,5,6,7,8,10,15 и 25 Гц. Через микрофоны двух типов (обычный конденсаторный и микрофон типа ДЭМШ, обычно располагаемый близко к губам) на магнитофон записывались по три реализации каждого слова.

Анализ фонограмм показал, что в условиях вибраций, особенно в тех случаях, когда их частота совпадает с частотой колебаний органов речеобразующего тракта (7-8 Гц), речь приобретает эмоциональную окраску, иногда становится похожей на речь плачущего человека, а в некоторых случаях у речи наблюдается оттенок охриплости. Совпадение частоты виброколебаний с частотой артикуляции приводит к резонансным явлениям и в конечном

итоге к нежелательным искажениям речи. При этом заметны колебания в громкости сигнала как на протяжении слова, так и на протяжении отдельных фонем. Уровень громкости колеблется в достаточно широких пределах для всех звуков речи. Эти эффекты объясняются тем, что вместе с виброколебаниями происходит усиленное или ослабленное проталкивание воздуха из легких через речеобразующий тракт. На звонких звуках громкость увеличивается (или уменьшается), если начало (или окончание) колебаний голосовых связок фазируется с виброколебанием. Несмотря на виброколебания, сохраняется опознаваемость диктора по голосу. Кроме того, было обнаружено, что при вдохе или выдохе через микрофон, в особенности тот, что расположен близко к губам (ДЭМШ), на систему распознавания может поступать речеподобный сигнал. Этот сигнал образуется путем модуляции шумов дыхания виброколебаниями.

1.2. Искажающая функция. Слуховой анализ фонограмм речи в условиях вибраций и визуальная обработка осциллограмм показали, что одной из главных характеристик виброречи является изменение уровня громкости или амплитуды сигнала с частотой виброколебания. Изменение амплитуды сигнала пропорционально силе (или амплитуде) вибраций. Следовательно, искажения можно моделировать путем модуляции речевого сигнала низкочастотным колебанием. Если обозначить через $x(t)$ непрерывный речевой сигнал, а через $y(t)$ - сигнал, искаженный вибропомехой, то процесс искажения можно описать в виде воздействия низкочастотной мультипликативной помехи $v(t)$ на сигнал $x(t)$:

$$y(t) = x(t)v(t),$$

$$v(t) = \frac{1}{b+c}\left[b\cos\left(\frac{2\pi t}{\tau} + \varphi\right) + c\right], \quad 0 < b < c, \quad -\infty < t < \infty. \quad (1)$$

Здесь $\tau$ и $\varphi$ - постоянные, определяющие период и фазу вибро-

колебания, а $b$ и $c$ - параметры, позволяющие регулировать глубину модуляции или силу вибрации.

Для моделирования вибропомех оцифрованный речевой сигнал $x_n = x(nT)$ $(n = 0, \pm 1, \pm 2, \ldots, T$- интервал дискретизации$)$ вводился в ЭВМ с восьмиразрядного аналого-цифрового преобразователя при частоте дискретизации 10 кГц. Виброискажения осуществлялись программным путем в соответствии с формулой (1) при $b = 9999$, $c = 10000$. Частоты виброколебаний были те же, что и у реальных сигналов, а фаза задавалась датчиком случайных чисел. В ЭВМ вводились фразы, слова, отдельные фонемы и шумы дыхания. После искажений сигнал выводился на прослушивание через восьмиразрядный цифро-аналоговый преобразователь.

При проведении слухового анализа четырьмя аудиторами и обработке осциллограмм модельного сигнала были отмечены все эффекты, свойственные реальному сигналу. На этом основании можно считать, что формула (1) достаточно хорошо моделирует процесс виброискажений.

## 2. Влияние вибропомех на оценки признаков

В системах цифровой обработки речи сигнал часто аппроксимируют моделью авторегрессии:

$$x_n = \sum_{i=1}^{p} a_i x_{n-i} + \varepsilon_n, \qquad (2)$$

где $\varepsilon_n$ - последовательность независимых, одинаково распределенных случайных величин с нулевым матожиданием и дисперсией $\sigma_\varepsilon(0)$. При этом спектральная плотность сигнала аппроксимируется спектральной плотностью модели авторегрессии:

$$E(\omega) = \frac{\sigma_\varepsilon(0)}{2\pi\left|1 - \sum_{i=1}^{p} a_i e^{-j\omega i}\right|^2}, \quad -\pi \le \omega \le \pi. \qquad (3)$$

Оценки параметров авторегрессии $\{a_i\}$ и однозначно связанной с ними спектральной плотности $E(\omega)$ лежат в основе большинства известных систем первичной обработки речевых сигналов. Поэтому устойчивость признаков к виброколебаниям можно рассмотреть на примере этих оценок. В данной работе это сделано на примере оценок спектральной плотности модели авторегрессии.

При проведении экспериментов указанные оценки вычислялись для фонем |а|, |ш| и шумов дыхания (вдох и выдох) при участии двух дикторов - мужчины и женщины. На рис. 1-4 для диктора-мужчины приведены графики функции

$$\hat{E}_1(\omega) = -20\log\left|1 - \sum_{i=1}^{p} \hat{a}_i e^{-j\omega i}\right|, \quad 0 \le \omega \le \pi,$$

т.е. графики оценок нормированной спектральной плотности (3) (при $\sigma_\varepsilon(0) = 2\pi$) в логарифмическом масштабе (в дБ). Вычисления производились по сигналам, дискретизированным с частотой 10 кГц, при ширине окна анализа 25,6 мс (длина выборки 256 отсчетов); порядок модели авторегрессии $p = 12$.

На каждом из рисунков первый график - это оценка нормированной авторегрессионной спектральной плотности фонемы или звука, вычисленная для одного окна анализа неискаженного сигнала. Эта оценка в дальнейшем рассматривается как эталонная.

На втором графике совмещены 64 оценки для сигнала из этого же окна анализа, искаженного восемью модельными вибропомехами при частотах $1/\tau$, равных 4,5,6,7,8,10,15 и 25 Гц. Фаза $\varphi$ виброколебаний принимала 8 значений в диапазоне от 0 до $7\pi/4$ с шагом $\pi/4$. Поскольку при уменьшении интенсивности вибропомех разброс оценок также уменьшается, величины параметров $b$ и $c$ у искажающей функции (см. п.1.2) были выбраны так, чтобы обеспечить максимальную интенсивность модельных вибропомех. Таким образом, вторые графики на каждом из рисун -

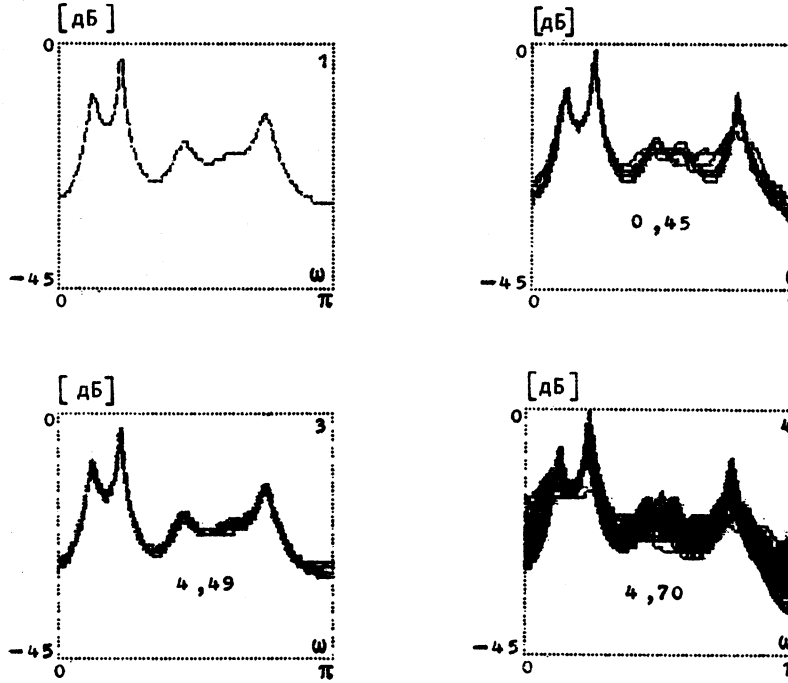


Рис.1. Оценки авторегрессионной спектральной плотности для фонемы |а|

ков иллюстрируют возможный разброс оценок из-за виброколебаний для одного отдельно взятого сегмента сигнала.

Третьи графики на каждом из рисунков показывают величину разброса оценок на восьми соседних участках сигнала (начиная с участка, оценка для которого приведена на первом графике) в условиях, когда искажения отсутствуют. Иными словами,эти графики иллюстрируют возможный разброс оценок для различных реализаций одной и той же фонемы или звука в отсутствие вибраций.Второй и третий графики позволяют сравнить разброс оценок из-за виброколебаний с разбросом оценок от реализации к реализации, т.е. с разбросом, который обычно имеет место при выделении приз-

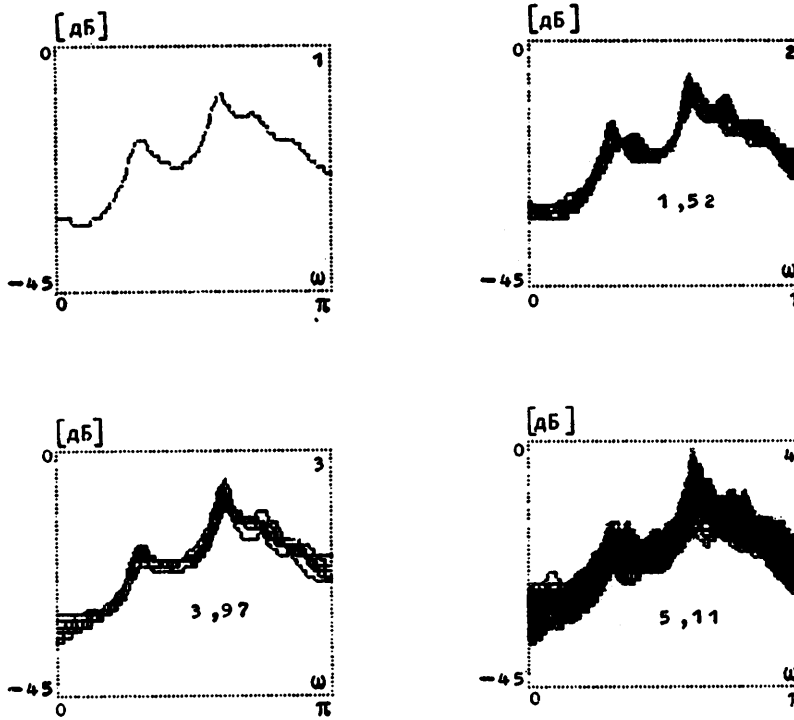


Рис. 2. Оценки авторегрессионной спектральной плотности для фонемы |ш|

наков в системах распознавания речи в отсутствие виброискажений.

На четвертом графике совмещены 512 оценок спектральной плотности для тех же, что и на третьем графике, восьми соседних участков сигнала, искаженных восемью модельными вибропомехами различной частоты при восьми значениях фазы виброколеба - ний. Эти графики иллюстрируют суммарный разброс оценок от реализации к реализации при различных частотах и фазах виброколебаний, т.е. такой разброс, который имеет место в системах распознавания, не имеющих средств компенсации вибропомех.

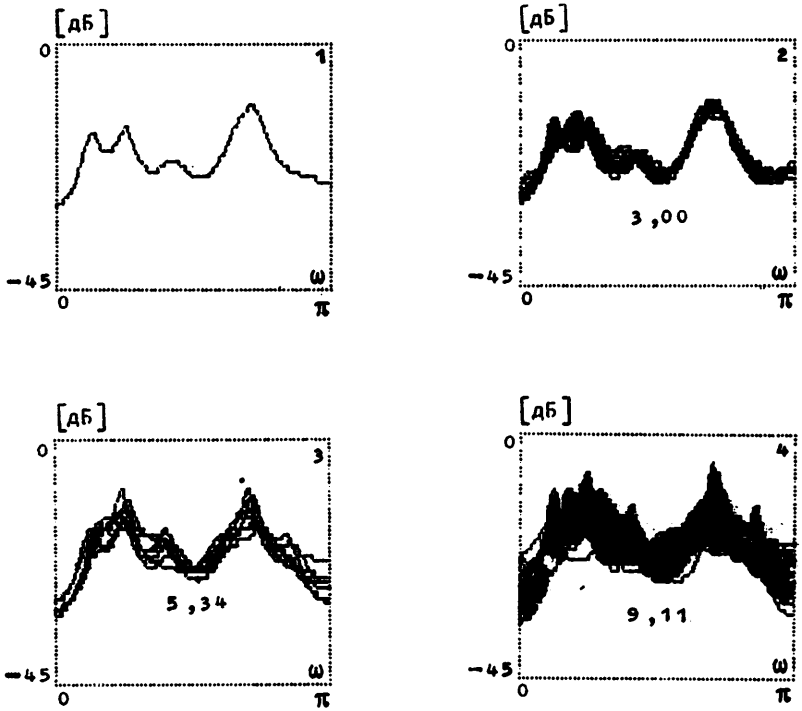


Рис. 3. Оценки авторегрессионной спектральной плотности шума вдоха

На каждом из графиков 2,3 и 4 приведены величины средне - квадратических отклонений от эталонной оценки, представленной на первом графике. Эти величины позволяют провести количественное сравнение дисперсии оценок при различных условиях обработки сигнала.Значения отклонений усреднены по мужскому и женскому голосам.

Приведенные значения свидетельствуют об увеличении дисперсии оценок в среднем. Максимальное же увеличение дисперсии оценок из-за вибраций может в 3-5 раз превышать указанные средние величины, что значительно превосходит разброс оце-

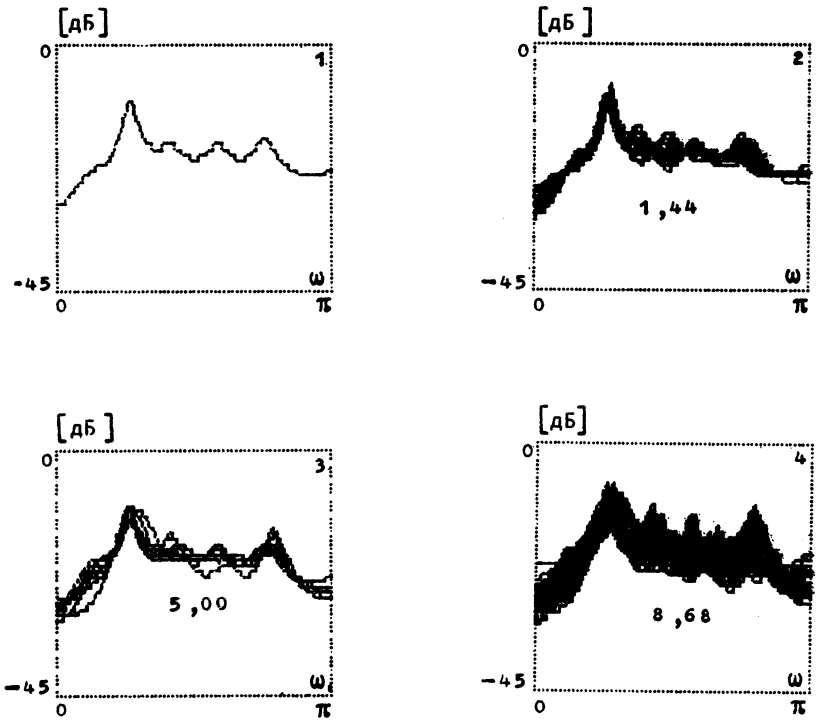


Рис. 4. Оценки авторегрессионной спектральной плотности шума выдоха

нок, который имеет место при обработке различных реализаций одной и той же фонемы или звука в отсутствие вибраций. Поэтому можно ожидать снижения надежности распознавания речи, если при выделении признаков не будут использованы какие-либо средства борьбы с вибропомехами.

Поскольку колебания оценок из-за вибраций происходят около эталонной оценки, в качестве одного из средств борьбы с вибропомехами или их компенсации можно использовать обычное сглаживание или усреднение оценок признаков.

## 3. Влияние вибраций на надежность распознавания

Как уже упоминалось выше, известные к настоящему времени системы распознавания речи не имеют средств компенсации помех типа вибраций. В связи с этим возникают вопросы: насколько устойчивы эти системы к вибропомехам и возможно ли применение этих систем в условиях реальных вибропомех без использования каких-либо средств компенсации.

Для ответа на эти вопросы был проведен эксперимент по распознаванию изолированных слов системой распознавания речи "Сибирь", разработанной в Институте математики СО АН СССР. В этой системе распознавания в качестве признаков используются оценки интенсивностей сигнала на выбранной сетке частот,которые с точностью до постоянной приближенно равны корню квадратному из оценок спектральной плотности сигнала на этой сетке частот.Граничные частоты фильтров спектроанализатора имели следующие значения: 300-500-1250-1900-3100-5000 Гц. Для распознавания использовался известный алгоритм динамического программирования (с некоторыми модификациями).

При проведении эксперимента каждая из трех реализаций команды, записанная на магнитофон в условиях реальных вибропомех для трех словарей и семи дикторов при восьми фиксированных частотах виброколебаний (см. п.1.1), предъявлялась на контроль. В качестве эталонов использовались записи тех же дикторов в отсутствие вибраций. Эксперименты показали, что при условии точного выделения границ полезного сигнала надежность распознавания лежит в пределах 98-100% при действии вибропомех частотой 4,5,6, 10,15 и 25 Гц (для всех дикторов, словарей и обоих типов микрофонов). При вибрациях 7 и 8 Гц надежность распознавания падает до 85-90%. Если же границы полезного сигнала в окружении модулированных шумов дыхания выделяются с ошибкой, то надежность распознавания для всех частот вибропомех может снизиться до 30-50%, что неприемлемо для практики.

Приведенные выше результаты по надежности распознавания получены при использовании сглаживания оценок по двум соседним участкам сигнала длительностью 16 мс. Если сглаживание как средство борьбы с помехами не использовать, то надежность распознавания снижается на 2-4%.

## 4. Обсуждение результатов. Выводы

Проведенные исследования выявили характерные особенности виброречи, которые, в свою очередь, позволили построить модель речевого сигнала в условиях вибропомех. Несмотря на простоту, модель достаточно точно отражает объективно существующую нелинейность эффекта виброискажений, а именно его мультипликативный характер. При этом мультипликативность вибропомех подтверждается экспертным визуальным анализом осциллограмм реальной виброречи, сходством реальных и модельных сигналов во временной области и отсутствием слуховых различий естественной и искусственной виброречи.

В плане поиска признаков, инвариантных к виброискажениям, отдельного рассмотрения требует вопрос, изменяют ли виброискажения моменты перехода речевого сигнала через нулевой уровень. Для ответа на этот вопрос напомним, что смещение этих точек перехода во времени, их пропажа или возникновение заметно снижают разборчивость речи. У реальной же виброречи подобного снижения разборчивости не наблюдается. Далее, следует учесть, что частоты виброколебаний, по крайней мере, на порядок ниже частоты переходов речевого сигнала через нулевой уровень. Так что "искаженных" моментов перехода (если считать, что они имеются) у виброречи по сравнению с неискаженной речью будет весьма немного. Поэтому можно утверждать, что виброискажения практически не затрагивают множества точек перехода сигнала через нулевой уровень. Это заключение подтверждается при визуальном сравнении осциллограмм речевых сигналов одного и того же диктора до

и после реальных виброискажений. Наконец, заметим, что и модельная искажающая функция имеет свойство не изменять, не устранять и не добавлять точек перехода сигнала через ноль. Это свойство - еще одно свидетельство в пользу адекватности предложенной модели.

Отдавая себе отчет в том, что любая модель какого-либо явления отражает не все реально существующие процессы, следует заметить, что построенная модель позволила выяснить степень влияния вибропомех на устойчивость оценок признаков речевого сигнала. Результаты моделирования свидетельствуют о том, что оценки авторегрессионной спектральной плотности речевого сигнала после виброискажений колеблются около эталонной оценки, сохраняя форму спектра сигнала до искажений. При этом дисперсия оценок может значительно увеличиться.

Если обучение системы распознавания проводить в идеальных условиях, т.е. в отсутствие вибраций, а распознавание - при их наличии, то, по сравнению с идеальными условиями, надежность распознавания уменьшится (если только не применять средства компенсации вибраций или не использовать признаков, устойчивых к виброколебаниям). Скорее всего, аналогичная ситуация должна иметь место и в случае, когда обучение и распознавание производится в условиях вибраций. Это предположение мотивировано тем, что фаза виброколебаний имеет случайный характер по отношению к процессу артикуляции.

Увеличение числа ошибок распознавания происходит,по крайней мере, по двум причинам. Первая состоит в увеличении дисперсии (уменьшении устойчивости) оценок признаков, что показано на примере спектральных оценок. Вторая причина - уменьшение точности выделения границ полезного сигнала в окружении модулированных шумов дыхания. Как известно, шумы дыхания являются весьма серьезной помехой для систем распознавания речи даже в том случае, когда другие помехи отсутствуют. Наличие вибраций

значительно затрудняет фильтрацию шумов дыхания (выделение границ полезного сигнала) из-за того, что модулированные шумы дыхания становятся похожими на речь. При этом сходство наблюдается как при слуховом восприятии, так и при статистическом оценивании признаков.

Существующие системы распознавания речи, а точнее, задействованные в них методы выделения признаков, можно считать вполне пригодными для работы в условиях вибропомех лишь при словарях объемом 10-20 слов, если частота виброколебаний разнится с частотой артикуляции. Поскольку надежность распознавания имеет тенденцию к снижению с ростом объема словарей, наличие вибропомех может существенно усилить этот эффект даже при частотах виброколебаний, отличных от частоты артикуляции. Поэтому необходимы средства компенсации помех, если их частота известна и интенсивность велика. Ввиду того, что малая интенсивность виброколебаний не ведет к сколько-нибудь заметному уменьшению надежности распознавания и увеличению дисперсии оценок признаков, а большая ухудшает указанные характеристики, следует выяснить зависимость этих характеристик от интенсивности вибраций. Эта зависимость необходима для того, чтобы знать, при какой пороговой интенсивности вибропомех следует воспользоваться компенсатором.

На практике частоты и интенсивности виброколебаний, как правило, неизвестны и к тому же случайны. В этих условиях единственным способом борьбы с искажениями являются методы выделения признаков, устойчивых или инвариантных к вибропомехам, частота и интенсивность которых меняется в широком диапазоне. Как показывают результаты экспериментов, в качестве одного из средств борьбы с вибропомехами, повышающих устойчивость оценок и увеличивающих надежность распознавания, можно использовать усреднение или сглаживание.

При применении сглаживающих процедур необходимо учитывать, что если обработка речи ведется без перекрытия соседних участков анализа, то усреднение оценок следует вести не более чем на двух-трех соседних участках сигнала. Это ограничение связано с тем, что,с одной стороны, усреднение оценок по большему числу окон анализа желательно для лучшей компенсации вибропомех, но с другой - усреднение по большему числу примыкающих участков снижает информативность самого первичного описания, что может привести к понижению надежности распознавания. Поэтому для разрешения указанного противоречия можно применять обработку речи с перекрывающимися окнами анализа, не выходя по длительности интервала усреднения за рамки двух-трех неперекрывающихся соседних участков сигнала. Если при этом окно анализа содержит $N$ отсчетов сигнала, то в рамках двух соседних окон можно получить $N$ оценок признаков, сдвигая окно анализа на один отсчет, и по этим оценкам провести сглаживание. Эта процедура нуждается в экспериментальной проверке.

Относительно методов обработки речевых сигналов, инвариантных к вибропомехам, необходимо отметить следующее. Эти методы как средство борьбы с вибрациями априори выглядят предпочтительнее методов повышения устойчивости путем сглаживания,поскольку эффективность последних зависит от интенсивности вибропомех (чем больше их интенсивность, тем больше разброс оценок и тем менее эффективно сглаживание). Инвариантность же по определению означает отсутствие подобной зависимости. Вопрос лишь в том, удастся ли построить такие методы и что должно лежать в их основе.

Очевидно, что эти методы должны базироваться на таких характеристиках или участках сигнала, которые не подвержены воздействию виброколебаний. Опыт обработки реальной виброречи показывает, что при вибропомехах, частоты и интенсивности которых изменяются в широких пределах, точки перехода сигнала че-

рез нулевой уровень остаются практически неподвижными. Как известно, эти точки позволяют восстанавливать разборчивую речь, поэтому можно попытаться построить инвариантные признаки на базе последовательности этих точек.

В заключение отметим, что проблема обработки речевых сигналов в условиях вибропомех в настоящее время еще далека от завершения. Многие эффекты, возникающие в речи из-за виброколебаний, остаются пока малоизученными, что сдерживает разработку адаптивных компенсаторов, настраивающихся на частоту и интенсивность вибраций. Вместе с тем, в результате исследований показано, что в качестве неадаптивного компенсатора виброискажений может использоваться обычная процедура сглаживания. Представляется, однако, самым главным, что полученные результаты демонстрируют определенную ограниченность традиционных методов борьбы с вибропомехами и заставляют более целенаправленно заняться созданием методов и алгоритмов выделения признаков, инвариантных к вибрациям.

Поступила в ред.-изд.отд.
13 декабря 1990 года