УДК 621.391

В. Ю. ПЕЛЕВИН

(Ленинград)

ИНФОРМАЦИОННЫЕ ПОТЕРИ РАЗЛИЧНЫХ СПОСОБОВ ПРЕДСТАВЛЕНИЯ СОНОГРАММ

Большое число систем распознавания речевых сигналов основывается на обработке речи в спектральном представлении [1]. Представление речи в координатах частота — время — интенсивность (яркость) обычно называют видимой речью или сонограммой речи. Конкретные реализации указанного представления речи могут заметно различаться. В настоящей работе рассматриваются способы представления сонограмм, обработка которых осуществляется в автоматическом режиме в соответствии с концепциями векторного квантования [2].

Отличия в представлении сонограмм связаны как с рассмотрением различных длительностей промежутка анализа речевого сигнала при вычислении мгновенных спектров сигналов, так и с использованием комплексного мгновенного спектра речевых сигналов или мгновенного энергетического спектра. Другим представлением сонограмм является использование логарифмического масштаба по переменным «интенсивность» и (или) «частота». В ряде спектроанализаторов происходит усреднение спектров речи в четвертьоктавных полосах.

Ясно, что перечисленные особенности различных способов представления видимой речи не могут не сказаться при ее автоматической обработке и экспертном изучении.

В [3] для сравнения систем автоматической обработки речи было предложено применять критерий информационных потерь по аналогии с каналами передачи информации с потерями в системах связи. Этот же подход может быть использован и при сопоставлении различных способов представления сонограмм речи. Действительно, если в качестве инициирующего звуковые колебания сигнала на входе речеобразующего тракта, представленного линейной системой с передаточной характеристикой $K(i\omega)$, взять смесь квазипериодической последовательности импульсов известной формы и гауссового белого шума для гласных и звонких согласных либо только гауссового белого шума для глухих согласных, то учет точностных ограничений реальной аппаратуры позволяет рассматривать сонограммы речи как проквантованный по переменным частота и время m-мерный случайный процесс с ограниченным последействием, частотные отсчеты которого для одного и того же момента времени взаимно независимы. Ясно, что в описанной математической модели сонограмм различные представления последних порождают различные вероятности потерь при обработке элементов входного алфавита — фонем в автоматической системе обработки речи. Оценка канальных матриц систем обработки речи, использующих различные представления сонограмм, может быть получена в рамках описанной вероятностной модели сонограмм в соответствии с изложенными в [4, 5] методами вычислений вероятностей принятия конкурирующих гипотез. Так, если система распознавания речи преобразует алфавит фонем $B = \{b_j\}_{j=1}^{r}$ в алфавит параметрических эталонов $A = \{a_i\}_{i=1}^{s}$, то информационные потери, описываемые канальной матрицей системы обработки $\{p(a_i/b_j)\}_{i,j=1}^{s,r}$, задаются значениями условных вероятностей появления на выходе системы обработки элемента алфавита параметрических эталонов a_i при условии реализации диктором фонемы b_j. По канальной матрице системы обработки речи могут быть вычислены размер эквивалентного алфавита фонем $k(A/B) = 2^{H(A/B)}$, где $H(A/B)$ — условная энтропия для канальной матрицы $\{p(a_i/b_j)\}_{i,j=1}^{s,r}$, ... [3], и средняя разборчивость алфавита параметрических эталонов $\lambda(A/B)$:

$$\lambda(A/B) = \sum_{i=1}^{s} p(b_i)\, p(a_i/b_i).$$

При этом предполагается использование в системе автоматического распознавания сонограмм одной и той же процедуры обработки, основанной на вычислении «расстояния» между эталоном, соответствующим элементу a_i выходного алфавита A, и предъявляемой сонограммой, соответствующей фонеме $b_j \in B$, определяемой метрикой Евклида на сонограммах как

$$\mu(U(\omega, t), V(\omega, t)) = \int_{t}^{t+T} \int_{0}^{\Omega} |U(\omega, t) - V(\omega, t)|^2 \, d\omega \, dt.$$

Здесь $U(\omega, t)$ — рассматриваемое представление сонограммы; $V(\omega, t)$ — эталон.

По результатам вычисления меры близости сонограммы с эталоном системой автоматического распознавания принимается решение о соответствующем значении для вектора признаков, в качестве которого в данном случае выступает сонограмма речевого сигнала. Для этого осуществляется проверка попадания вычисленного зна-

чения меры близости с эталоном a_i в критическую область проверяемой гипотезы. При этом имеется в виду следующее построение алгоритма классификации: предъявляемая сонограмма используется для параллельного формирования мер близости со всеми параметризованными эталонами $a_i \in A$, после чего последовательно происходит проверка статистических гипотез о совпадении с j-м параметрическим эталоном против альтернативы о наличии совпадения с одним из эталонов $a_{j+1}, \ldots, a_s$. Работа классификатора заканчивается, как только принимается гипотеза.

Соответственно при использовании комплексных спектров речевых сигналов рассматриваемая мера близости переходит в выражение

$$\mu_1(F, \Phi) = \int_t^{t+T} \int_0^{\Omega} G(\omega) \mid F(\omega, t) - \Phi(\omega, t) \mid^2 d\omega\, dt,$$

где $G(\omega)$ — множитель, компенсирующий спад спектра 6—12 дБ на октаву; $F(\omega, t) = \underset{t+\Delta t}{} = \frac{1}{\sqrt{2\pi}} \int_t^{t+\Delta t} e^{i\omega t} x(t)\, dt$; $\Phi(\omega, t) = \frac{1}{\sqrt{2\pi}} \int_t^{t+\Delta t} e^{i\omega t} f(t)\, dt$; $x(t)$ — речевой сигнал; $f(t)$ — эталон.

Заметим, что оптимальная в смысле теории Неймана — Пирсона процедура обработки речевых сигналов требует вычисления экстремального значения на параметрическом множестве эталонов достаточно близкого к рассмотренному выражению функционала

$$\mu_0(F, \Phi) = \langle \mathbf{C}^{-1}(F(\omega_i, t_k) - \Phi(\omega_i, t_k)),\ (F(\omega_i, t_k) - \Phi(\omega_i, t_k)) \rangle.$$

Здесь $\mathbf{C}$ — матрица корреляции мгновенных спектров. При неперекрывающихся отрезках — промежутках анализа мгновенного спектра речевых сигналов — оптимальный в смысле теории Неймана — Пирсона функционал является также функционалом, определяющим согласованный фильтр:

$$\mu_0(F, \Phi) = \sum_t \int_0^{\Omega} \frac{\mid F(\omega, t) - \Phi(\omega, t) \mid^2}{\mid K(i\omega, t) \mid^2} d\omega.$$

При использовании энергетических спектров сигналов метрика Евклида примет вид

$$\mu_2(F, \Phi) = \int_t^{t+T} \int_0^{\Omega} (\mid F(\omega, t) \mid^2 - \mid \Phi(\omega, t) \mid^2)^2 G^2(\omega)\, d\omega\, dt.$$

Если используется логарифмический масштаб по параметру «интенсивность», то метрика Евклида переходит в меру близости вида

$$\mu_3(F, \Phi) = \int_t^{t+T} \int_0^{\Omega} \left| \log \left| \frac{F(\omega, t)}{\Phi(\omega, t)} \right| \right|^2 d\omega\, dt.$$

При использовании логарифмического масштаба по переменной «частота» в любом из рассмотренных представлений сонограмм метрика Евклида преобразуется к виду

$$\mu(U(\omega, t), V(\omega, t)) = \int_t^{t+T} \int_0^{\Omega} \mid U(\omega, t) - V(\omega, t) \mid^2 d \ln \omega\, dt =$$

$$= \int_t^{t+T} \int_0^{\Omega} \frac{\mid U(\omega, t) - V(\omega, t) \mid^2}{\omega} d\omega\, dt.$$

Если не принимать специальных мер, устранятся компенсирующее воздействие функции $G(\omega)$.

Отметим, что при использовании регрессионных моделей речевых сигналов приведенные выше меры близости с эталоном имеют аналогами соответственно для μ_0 — метрику отношения вероятностей, а для μ_3 — кепстральную метрику [6].

Последовательное проведение вычислений в соответствии с [4, 5] вероятностей правильного и ошибочного распознаваний для перечисленных мер близости с параметрическими эталонами гласных позволяет провести оценку канальной матрицы системы распознавания для каждого из перечисленных вариантов представления сонограммы. При отсутствии ошибок и погрешностей обработки, обусловленных погрешностями работы аппаратуры, канальные матрицы систем обработки, распознающих гласные звуки, представлены в табл. 1. Эффективный размер алфавита для таких систем равен соответственно 1,161; 1,164; 1,357; 1,363.

Таблица 1

Мера близости	у	о	а	э	и	ы	$k(A/B)$; $\lambda(A/B)$
μ_0	0,975	0,005	0,005	0,006	0,005	0,004	$k = 1{,}161$
	0,005	0,977	0,004	0,005	0,004	0,005	$\lambda(\mu_0) = 0{,}976$
	0,004	0,004	0,977	0,005	0,005	0,005	
	0,005	0,005	0,006	0,975	0,005	0,004	
	0,004	0,004	0,005	0,005	0,977	0,005	
	0,005	0,005	0,005	0,004	0,004	0,977	
μ_1	0,975	0,005	0,005	0,005	0,005	0,005	$k = 1{,}164$
	0,005	0,975	0,005	0,005	0,005	0,005	$\lambda(\mu_1) = 0{,}975$
	0,005	0,005	0,975	0,005	0,005	0,005	
	0,005	0,005	0,005	0,975	0,005	0,005	
	0,005	0,005	0,005	0,005	0,975	0,005	
	0,005	0,005	0,005	0,005	0,005	0,975	
μ_2	0,949	0,011	0,012	0,011	0,009	0,008	$k = 1{,}357$
	0,012	0,959	0,011	0,007	0,007	0,004	$\lambda(\mu_2) = 0{,}941$
	0,014	0,009	0,942	0,012	0,011	0,012	
	0,017	0,008	0,013	0,934	0,012	0,016	
	0,008	0,007	0,012	0,012	0,906	0,055	
	0,005	0,006	0,010	0,008	0,012	0,959	
μ_3	0,975	0,005	0,005	0,005	0,005	0,005	$k = 1{,}363$
	0,091	0,889	0,005	0,005	0,005	0,005	$\lambda(\mu_3) = 0{,}926$
	0,005	0,005	0,975	0,005	0,005	0,005	
	0,005	0,005	0,005	0,886	0,005	0,094	
	0,005	0,005	0,005	0,042	0,848	0,095	
	0,005	0,005	0,005	0,005	0,005	0,975	

Таблица 2

Мера близости	у	о	а	э	и	ы	$k(A/B)$; $\lambda(A/B)$
μ_0	0,943	0,016	0,019	0,012	0,005	0,005	$k = 1{,}664$
	0,086	0,616	0,057	0,092	0,090	0,059	$\lambda(\mu_0) = 0{,}875$
	0,007	0,031	0,865	0,043	0,023	0,031	
	0,008	0,005	0,005	0,923	0,013	0,046	
	0,005	0,005	0,005	0,027	0,953	0,005	
	0,005	0,005	0,024	0,008	0,010	0,948	
μ_1	0,974	0,005	0,005	0,006	0,005	0,005	$k = 1{,}165$
	0,006	0,973	0,006	0,005	0,005	0,005	$\lambda(\mu_1) = 0{,}974$
	0,005	0,005	0,974	0,006	0,005	0,005	
	0,006	0,005	0,006	0,973	0,005	0,005	
	0,005	0,005	0,005	0,005	0,975	0,005	
	0,005	0,005	0,005	0,005	0,005	0,975	
μ_2	0,948	0,012	0,012	0,011	0,009	0,008	$k = 1{,}362$
	0,012	0,959	0,010	0,007	0,007	0,005	$\lambda(\mu_2) = 0{,}940$
	0,014	0,009	0,940	0,013	0,011	0,013	
	0,018	0,008	0,013	0,933	0,012	0,016	
	0,008	0,007	0,012	0,012	0,904	0,057	
	0,005	0,006	0,010	0,008	0,011	0,960	
μ_3	0,865	0,095	0,022	0,005	0,005	0,008	$k = 1{,}845$
	0,095	0,779	0,095	0,010	0,005	0,016	$\lambda(\mu_3) = 0{,}834$
	0,071	0,087	0,823	0,005	0,005	0,009	
	0,005	0,005	0,005	0,817	0,073	0,095	
	0,037	0,006	0,025	0,095	0,742	0,095	
	0,005	0,005	0,005	0,005	0,005	0,975	

Учет ограничений динамического диапазона и базы обрабатываемого сигнала даже при отсутствии фазовых искажений приводит к падению средней разборчивости алфавита параметрических эталонов при распознавании. Для базы 100 и динамического диапазона обрабатываемых сигналов 50 дБ в отсутствие искажений фазы в табл. 2 приведены оценки канальной матрицы систем распознавания гласных по различным представлениям сонограмм. В описанном случае эффективный размер алфавита гласных равен соответственно 1,664; 1,165; 1,362; 1,845.

Результаты оценивания канальных матриц дают основание для следующих выводов:

1. Использование оптимальных с точки зрения обработки представлений сонограмм комплексными мгновенными спектрами во времени при динамическом диапазоне сигнала менее 50 дБ и базе меньшей 100 в шумах наблюдений при отклонениях от идеализированной модели речеобразования не дает сколько-либо значительного выигрыша в эффективности работы системы распознавания.

2. Логарифмический масштаб по шкале интенсивности в представлении сонограмм отсчетами энергетического спектра делает процедуру обработки помехонеустойчивой, что сильно снижает эффективность автоматической обработки.

3. Использование изменяющегося во времени мгновенного энергетического спектра для представления сонограмм при удовлетворительной эффективности распознавания приводит к повышению помехоустойчивости системы автоматической обработки речи.

4. Логарифмический масштаб по переменной «частота» в представлении сонограмм снижает эффективность работы системы распознавания.

ЛИТЕРАТУРА

1. **Автоматическое** распознавание слуховых образов // Тез. докл. 14 Всесоюз. семинара (АРСО 14) 26—28 августа 1986 г.— Каунас: КПИ, 1986.
2. **ТИИЭР.**— 1985.— **73**, вып. 11.
3. **Обжелян Н. К., Трунин-Донской В. Н.** Речевое общение в системах «человек — ЭВМ».— Кишинев: Штиинца, 1985.
4. **Пелевин В. Ю.** Сравнение мер акустического сходства речевых сигналов в задачах предварительной обработки речи // Теория передачи информации по каналам связи.— Л.: ЛЭИС, ТУИС, 1984.— Вып. 117.
5. **Пелевин В. Ю.** Требования к точностным характеристикам аппаратуры предварительной обработки речи // Оптимизация систем передачи информации по каналам связи.— Л.: ЛЭИС, ТУИС, 1986.— Вып. 126.
6. **Gray A. H., Markel J. D.** Distance measures for speech processing // IEEE Trans. Acoust. Speech, Sign. Proc.— 1975.— ASSP-24.— P. 380.

Поступило в редакцию 11 ноября 1987 г.

УДК 621.391.274

В. П. КОРЯЧКО, К. В. МЕРОВ, В. Н. ПЕРЕПЕЛКИНА, С. И. СИДЕЛЬНИКОВ

(Рязань)

МЕТОДИКА УПЛОТНЕНИЯ ЦИФРОВЫХ ДАННЫХ

Информационно-измерительные системы (ИИС) находят широкое применение во многих отраслях народного хозяйства и предназначены для сбора, обработки и регистрации больших объемов цифровой информации. Одним из важных параметров, влияющим на характеристики всей ИИС, является пропускная способность информационных каналов — передачи данных и регистрации. Увеличение пропускной способности достигается как аппаратным (использование высокочастотных кабелей, оптических линий связи, совершенствование магнитных головок регистраторов и т. д.), так и алгоритмическими способами (сжатие и уплотнение данных).

Сжатие данных осуществляется путем устранения избыточных отсчетов за счет аппроксимации измеряемого сигнала известной функцией [1], отклоняющейся от аппроксимируемого сигнала не более чем на допустимую погрешность.

Уплотнение данных позволяет за счет перекодирования [2] уменьшить объем исходного сообщения без внесения в него погрешности.

В настоящей статье рассматривается методика уплотнения цифровых данных в аналоговом канале связи, позволяющая для заданного потока цифровых данных уменьшить требуемую полосу пропускания.

При передаче цифровой информации телеграфным кодом требуемая ширина полосы пропускания канала F определяется как [3]

$$F = m \times Q, \qquad (1)$$