

МЕТОДИКА ОЦЕНКИ ИНФОРМАЦИОННОЙ ЭФФЕКТИВНОСТИ
НЕЗАВИСИМЫХ ПАРАМЕТРОВ РЕЧЕВОГО СИГНАЛА

Н.Г. Загоруйко

В работе [1] было показано, что критерием правильности выбора системы параметров для опознавания речевых сигналов является величина её информационной эффективности. Указывалось также, что иногда имеет практический смысл оценка информационной эффективности отдельного независимого параметра. Действительно, если нет другого выхода, то по эффективности отдельного параметра можно составить себе некоторое представление о его вкладе в эффективность системы зависимых параметров. Далее, в настоящее время из-за ограниченных возможностей опознающей аппаратуры заведомо зависимые параметры часто используются как независимые. Наконец, для опознавания сигнала используется ряд действительно независимых параметров (например, длительность слова, наличие взрывного звука в начале или конце слова, средний уровень сигнала за время звучания слова и т.д.).

Рассмотрим методику оценки информационной эффективности независимых параметров, а также вопрос о выборе рационального числа градаций параметра и рационального размещения границ между заданным числом градаций. Покажем, что выбранная методика позволяет оценить также эффективность некоторого преобразования сигнала, называемого "нормализацией".

1. Если звуковой образ есть локальная область n -мерного пространства, то под параметром будем понимать любую из осей этого пространства.

В том случае, когда для опознавания речевого сигнала используется информация о его энергии в частотных полосах, параметром, по определению, является не спектр сигнала в целом, а каждая полоса частот в отдельности.

Под информационной эффективностью параметра понимается отношение $R_x = \frac{I_x}{n_x}$, где I_x - количество информации о звуковых образах, получаемое с помощью параметра x , а n_x - число операций ЭВМ, затрачиваемое на извлечение этого параметра из речевого сигнала.

Методика оценки количества информации I_x , содержащегося в одном независимом параметре, сводится к следующему. Пусть априорная вероятность появления каждого из N звуковых образов, подлежащих опознаванию, равна $q_1, q_2, q_3, \dots, q_N$, соответственно.

Используемый для опознавания параметр может принимать m различных значений, т.е. разбит на m уровней (рис. I).

Если распределение вероятности появления k -го образа при различных значениях параметра подчиняется нормальному закону с математическим ожиданием $M[x_k]$ и среднеквадратичным отклонением σ_k , то вероятность появления этого образа при i -ом значении параметра (т.е. при $x_{i-1} \leq x < x_i$) будет

$$P_{k/i} = \frac{q_k}{2} \left[\Phi\left(\frac{x_i - M[x_k]}{\sigma_k \sqrt{2}}\right) - \Phi\left(\frac{x_{i-1} - M[x_k]}{\sigma_k \sqrt{2}}\right) \right],$$

где $\Phi(x)$ - табличный интеграл вероятности.

Если известны числовые характеристики законов распределения всех образов, то энтропия системы образов при i -ом значении параметра будет определяться по формуле

$$H_i = - \left(\frac{P_{1/i}}{\sum_{k=1}^N P_{k/i}} \cdot \log \frac{P_{1/i}}{\sum_{k=1}^N P_{k/i}} + \dots + \frac{P_{N/i}}{\sum_{k=1}^N P_{k/i}} \cdot \log \frac{P_{N/i}}{\sum_{k=1}^N P_{k/i}} \right) =$$

$$= \frac{1}{\sum_{k=1}^N P_{k/i}} \left[\sum_{k=1}^N P_{k/i} \cdot \log \sum_{k=1}^N P_{k/i} - \sum_{k=1}^N (P_{k/i} \cdot \log P_{k/i}) \right].$$

Из принципа аддитивности энтропии следует, что общая неопределенность H_x в выборе образа по параметру $x_0 \leq x < x_m$ будет равна сумме энтропий H_i (с учетом того, что вероятность появления i -го значения параметра равна $\sum_{k=1}^N P_{k/i}$):

$$H_x = \sum_{i=1}^m \left[H_i \cdot \sum_{k=1}^N P_{k/i} \right] \dots$$

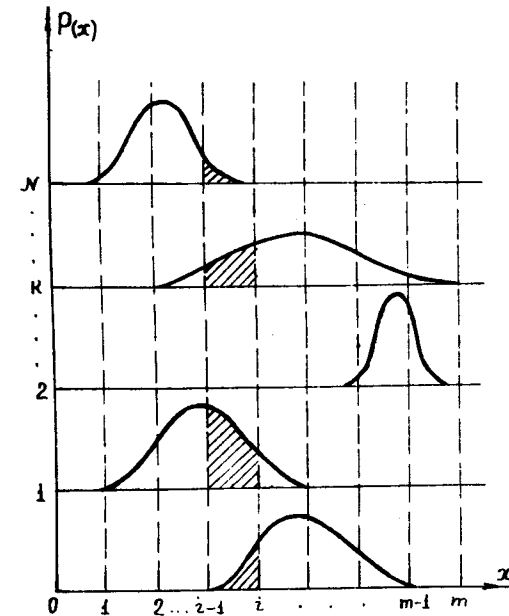


Рис. I.

Так как энтропия H_0 , имевшая место до использования параметра x , зависит только от априорных вероятностей появления образов, то

$$H_0 = - \sum_{k=1}^N q_k \cdot \log q_k.$$

Информация, извлекаемая при использовании рассмотренного параметра, определяется как степень уменьшения энтропии, т.е.

$$I_x = H_0 - H_x.$$

Изложенный метод справедлив для любого вида распределения вероятности, если под $P_{k/i}$ понимать суммарную плотность вероятности $P(x)$ на участке $x_{i-1} \leq x < x_i$ (затрихованная часть рис. I).

2. Чем точнее измеряется параметр (т.е. чем на большее число диапазонов m разбиты его значения), тем большую информацию можно из него извлечь. Увеличение числа диапазонов (градаций) параметра приводит к увеличению числа машинных операций (или усложнению аппаратуры). Поэтому важно знать, каким числом градаций параметра целесообразно ограничиваться и на какие участки нужно разбивать параметр (где ставить границы между участками).

Ответим на первый вопрос. При увеличении числа градаций возрастает как числитель, так и знаменатель отношения $R = \frac{\mathcal{I}}{n}$, причем знаменатель растет обычно быстрее числителя. Таким образом, с увеличением числа градаций абсолютная величина извлекаемой информации возрастает, а информационная эффективность параметра R уменьшается. Следовательно, если число машинных операций значения не имеет, то число градаций можно увеличивать до тех пор, пока количество извлекаемой информации $\mathcal{I}$ не перестанет практически возрастать. Если же числом операций (затратами машинного времени) пренебрегать нельзя, то число градаций будет ограничено минимально допустимым значением коэффициента R или максимально возможным числом машинных операций n .

Ответ на второй вопрос более сложен. Задачу о рациональном размещении $m-1$ границ между m градациями параметра можно решить методом полного перебора. Однако при большом числе границ и при высокой точности фиксации их положения такой метод приводит к очень громоздким вычислениям.

Решение задачи на минимум величины H_x в общем виде неизвестно. Эксперименты показали, что удовлетворительные результаты по минимизации H_x могут быть получены следующим образом. Вначале вычисляются значения H_i при максимально возможном числе градаций параметра. Прямоугольники, образованные сторонами, равными $\sum_{k=1}^N P_{k/i}$ и H_i , можно объединить в общую фигуру, площадь которой равна H_x (рис. 2). Вопрос о том, какие границы можно убрать, объединив соседние градации в одну без существенного уменьшения величины извлекаемой из параметра информации $\mathcal{I}_x$, решается исходя из вида этой фигуры. Оказалось, что в первую очередь можно убирать те границы, для которых величина $P_i \Delta H_i$ минимальна.

Здесь

$$P_i = \sum_{k=1}^N (P_{k/i} + P_{k/i+1}), \text{ а } \Delta H_i = |H_i - H_{i+1}|.$$

(Затрихованная часть рис. 2).

Значения величин H_i , $\sum_{k=1}^N P_{k/i}$ и $P_i \Delta H_i$, полученные в результате оценки информационной эффективности параметра, в качестве которого использовалась энергия речевого сигнала в полосе частот от 600 до 830 гц, приведены в таблице. Предварительно статистика набиралась для 15 градаций энергии в этой полосе. Оказалось, что информация $\mathcal{I}_x$ о 50 фонемах русской речи, которая содержится в этом параметре, измеряемом с данной точ-

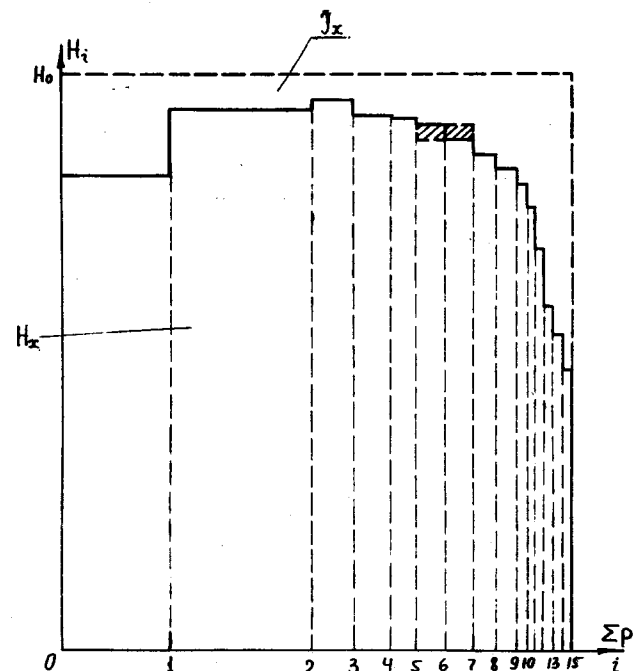


Рис. 2.

ностью, равна 0,694 бит. ($H_0 = 5,644$; $H_x = 4,950$). В 5-м столбце таблицы I приводятся величины информации $\Delta \mathcal{I}_i$, которые будут теряться, если убирать i -ую границу (при условии, что остальные границы останутся). Из этой таблицы видно, что объединение, например, 4 с 5, 8 с 9, 10 с 11, 12 с 13 и 14 с 15 градаций уменьшит их число с 15 до 10, практически не уменьшив количества информации, извлекаемой из параметра (теряется $\approx 1\% \mathcal{I}_x$). Анализ зависимости между информацией $\Delta \mathcal{I}_i$, теряемой при ликвидации i -ой границы, и произведением $P_i \Delta H_i$ показал, что коэффициент взаимной корреляции между ними равен + 0,95. Поэтому при решении вопроса о том, какие границы можно опустить в первую очередь, на величину $P_i \Delta H_i$ можно ориентироваться с достаточным основанием.

3. Для того, чтобы результат опознавания звукового образа не зависел от уровня громкости, часто прибегают к автоматическому регулированию уровня громкости с целью поддержания его постоянным на входе анализирующей аппаратуры. Эта процедура, получившая название "нормализации", применяется иногда и по отношению к другим параметрам речевого сигнала (нормали-

Таблица I

i	H_i	$\sum_{k=1}^N P_{k/i} \cdot 50$	$P_i \cdot \Delta H_i$	$\Delta \mathcal{I}_i$
I	4,666	47,551	14,403	0,094
2	5,313	64,349	1,358	0,028
3	5,387	33,603	1,206	0,007
4	5,262	17,961	0,062	0,002
5	5,250	14,594	0,491	0,003
6	5,172	18,201	1,062	0,004
7	5,008	14,807	0,640	0,002
8	4,881	10,187	0,564	0,002
9	4,723	9,438	0,420	0,002
10	4,590	5,333	0,490	0,001
11	4,336	3,307	0,596	0,001
12	3,958	3,219	0,844	0,001
13	3,392	2,304	0,310	0,001
14	3,098	1,161	0,291	0,001
15	2,778	1,481		

зация по длительности, нормализация по средней форме спектра и т.д.).

Улучшение результатов опознавания при использовании нормализации объясняется тем, что параметры сигнала защищаются от некоторой части помех. Так, после нормализации уровня громкости звука величина энергии в определенных частотных полосах будет зависеть от характера звука, а не от чувствительности микрофона, расстояния между диктором и микрофоном и т.д. Распределение нормализованного параметра будет характеризоваться меньшей величиной дисперсии. Изменение закона распределения параметра, как видно из предыдущего, приводит к изменению величины извлекаемой с его помощью информации, причем уменьшение среднеквадратичного отклонения приводит к увеличению информации $\mathcal{I}_x$.

Пусть m — число градаций параметра. Расставив границы между градациями оптимальным способом, мы до нормализации извлекаем информацию $\mathcal{I}_x$, а после нормализации — $\mathcal{I}_{xH}$. Тогда $R_H = \frac{\mathcal{I}_{xH} - \mathcal{I}_x}{n_H}$, где n_H — число машинных операций, затрачиваемых на нормализацию, будет характеризовать эффективность нормализации.

В наших опытах было установлено, например, что отклонения в длительности произнесения одного и того же слова одним диктором не превышают $\pm 15\%$ от математического ожидания длительности. Отклонения же в длительности произнесения того же слова разными дикторами достигают $\pm 45\%$.

Под нормализацией по длительности можно понимать подстройку эталонов распределения под диктора, т.е. процедуру, включающую в себя произнесение новым диктором определенного набора слов и совмещение математических ожиданий эталонов длительности с длительностью слов, реально произнесенных при подстройке. Если подстройку делать не часто (т.е. если один диктор работает с машиной долгое время), то затратами машинных операций на нормализацию по длительности можно пренебречь. Расчеты показали, что при 15 различных словах количество информации, извлекаемой из длительности звучания слов до нормализации $\mathcal{I}_x$, было равно 1,057 бит, а после нормализации $\mathcal{I}_{xH}$ возрастало до 1,862 бит.

Ясно, что если нет возможности нормализовать все параметры, то нужно нормализовать те из них, которые характеризуются большей величиной R_H .

В заключение отметим, что выбранная методика дает возможность оценивать не только информационную эффективность параметра, но и правильность выбора числа градаций параметра и размещения границ между ними, а также эффективность нормализации речевого сигнала.

Л и т е р а т у р а

И. Загоруйко Н.Г. Об обмене звуковой информацией между человеком и вычислительными системами (данный сборник).