ИЗВЕСТИЯ
ОРДЕНА ОКТЯБРЬСКОЙ РЕВОЛЮЦИИ И ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ ПОЛИТЕХНИЧЕСКОГО ИНСТИТУТА
имени С. М. КИРОВА

Том 194 1972

К ОЦЕНКЕ ЭФФЕКТИВНОСТИ ПАРАМЕТРОВ РЕЧЕВЫХ СИГНАЛОВ

Б. Н. ЕПИФАНЦЕВ, В. П. ЕВМЕНОВ

(Представлена научным семинаром кафедры вычислительной техники)

Известно, что при переходе от абсолютного описания речевых сигналов к системе признаков последние выбираются методом «проб и ошибок» [1]. Но чтобы оценить, насколько удачно был выбран тот или иной признак, необходим критерий, дающий количественную оценку признака с точки зрения его последующего применения. До последнего времени в качестве такого критерия в большинстве случаев использовался процент ошибок классификации. Для этого в пространстве выбранного признака (признаков) строились собственные области образов, оценивалась их форма (выпуклость, вогнутость), степень пересечения областей, относящихся к разным классам, и т. д. После этого предлагалась методика классификации, ставился эксперимент и проверялось, какую ошибку следует ожидать при распознавании данных образов. Указанной величиной ошибки характеризовалась эффективность признака (признаков). То ли в силу огромного числа признаков и их коррелированности, то ли в результате применения не тех методик классификации, получить приемлемых результатов при ориентировке на рассматриваемый способ не удалось до сих пор. Нетрудно понять, что данная процедура оценки эффективности признаков не дает информации, в какой степени исследуемый признак важен для человека, распознающего речь. Поэтому мы не можем сказать, отражает ли процент ошибок классификации информацию, необходимую либо для понимания сказанного, либо для характеристики особенностей произношения. Этим, пожалуй, объясняется тот интерес, который сейчас проявляется к полуобъективному критерию, коэффициенту разборчивости и способам оценки эффективности признаков, отсюда вытекающим.

Одним из способов, основанным на коэффициенте разборчивости, является способ «анализа через синтез». Идея его состоит в том, что с помощью подбора параметров электрической модели речевого тракта синтезируются звуки и в случае получения приемлемого с точки зрения разборчивости сигнала производится математическое описание модели.

Другой способ может быть назван «способом искажений». Эффективность признаков по этому способу оценивается коэффициентом разборчивости речи, подвергнутой намеренным целенаправленным искажениям. Факт исчезающего малого влияния соотношения амплитуд в сигнале на разборчивость (клиппированная речь) получен в результате применения «способа искажения».

Несмотря на кажущуюся простоту, методика постановки рассмотренных методов (исключая первый) не разработана, чем, собственно, можно объяснить эпизодичность работ на их основе. С другой стороны, коэффициент разборчивости может быть использован для оценки отдельных признаков и ряда их совокупностей.

Ниже предлагается иной подход к оценке эффективности признаков. По существу он является продуктом объединения рассмотренных способов и позволяет получить ту информацию о признаке, которую не в состоянии дать существующее.

Анализ речи дал много фундаментальных фактов. Нас будут интересовать такие, как, например, увеличение разборчивости клиппированной речи, предварительно пропущенной через четырехполюсник, амплитудно-частотная характеристика которого имеет подъем в сторону высоких частот, т. е. такие преобразования, которые, бесспорно, затрагивают информационную структуру речевых сигналов. Тогда влияние тех или иных преобразований может быть описано вектором $\bar{\delta}$, каждая орта которого $\delta_i = \pm (A_i/A)$, где A_i — коэффициент разборчивости речи, подвергнутой i преобразованию, A — коэффициент разборчивости неискаженной речи, знак „+“ ставится, когда $A < A_i$, „—“, когда $A > A_i$.

Но нетрудно получить аналогичный вектор $\bar{\sigma}$. При этом $\sigma_i = \pm (d_i/d_0)$, где d_i — критерий различения по выбранным признакам между классифицируемыми речевыми сигналами, подвергнутыми i преобразованию, d_0 — тот же критерий для неискаженной речи, знак „+“ ставится при условии $d_0 < d_i$ „—“, когда $d_0 > d_i$. Поскольку размерность векторов $\bar{\sigma}$ и $\bar{\delta}$ одинакова, оценкой эффективности признака может служить величина

$$d(\bar{\sigma}, \bar{\delta}) = \frac{\sum_{i=1}^{n} (\sigma_i - \delta_i)^2}{\sum_{i=1}^{n} (|\sigma_i| + |\delta_i|)^2}.$$

Очевидно, $0 \leqslant d(\bar{\sigma}, \bar{\delta}) \leqslant 1$, при этом, когда $d(\bar{\sigma}, \bar{\delta}) = 0$, эффективность рассматриваемых признаков максимальна (100%), при $d(\bar{\sigma}, \bar{\delta}) = 1$ эффективность ($Э$) равна 0. Зависимость $Э = f[d(\bar{\sigma}, \bar{\delta})]$ является предметом специальных исследований. Величина $d(\bar{\sigma}, \bar{\delta})$ содержит информацию о том, насколько важен при распознавании исследуемый признак (признаки) для человека и насколько близка применяемая процедура опознавания к используемой человеком.

Развитием этого способа может служить привлечение информации о влиянии факторов (громкость, длительность звуков и т. д.) на значение выбранного параметра (параметров) и эвристических соображений о функциональной зависимости коэффициента разборчивости от рассматриваемых факторов.

Проиллюстрируем сказанное примерами.

1. Требуется оценить эффективность средневозвышенной частоты звуков (величины, пропорциональной плотности переходов функции через нуль $\rho_{T_з}$, найденной на интервале существования звука $T_з$).

Для решения этой задачи звуки последовательно через преобразователь аналог—код вводились в оперативную память ЭВМ БЭСМ-2М. Программным путем для каждого звука подсчитывалось число переходов функции через нуль $\rho_{T_з}$, которое затем делилось на время $T_з$, т. е. находилась величина

$$F_{T_з} = \frac{\rho_{T_з}}{T_з},$$

Контроль соответствия введенного в ЭВМ сигнала исходному осуществлялся прослушиванием записанной в машину информации путем вывода ее через синтезатор на громкоговоритель.

Каждый из гласных звуков j произносится одним диктором одинаковой громкостью по 40 раз и для каждой реализации i (1, 2..., ... 40) определялась величина $F^j_{T_{3i}}$. Затем по данным статистического ряда $F^j_{T_{3i}}$ вычислялись частоты многоугольника распределения $P^{j\,3}_{\delta F_{Ti}}$, попадающие в соответствующий ряд δF_{T_3}, и эвклидовы расстояния

$$d\,(P_{\delta F^j_{T_{3i}}},\ P_{\delta F^\kappa_{T_{3i}}}) = \sqrt{\sum_{\delta=1}^{N}(P_{\delta F^j_{T_{3i}}} - P_{\delta F^\kappa_{T_{3i}}})^2}.$$

По ряду значений $d\,(\cdots)$ находилось общее расстояние

$$d_{\text{ср}} = \frac{1}{M}\sum_{j\kappa} d\,(P_{\delta F^j_{T_{3i}}},\ P_{\delta F^\kappa_{T_{3i}}})$$

путем усреднения всех возможных расстояний $j\kappa\,(j \neq \kappa)$.

Следующий этап — получение величин $d^{(и\kappa)}_{\text{ср}}$, $d^{(п\Phi)}_{\text{ср}}$, которые отличаются от $d_{\text{ср}}$ тем, что перед вводом в машину речь пропускалась либо через частотный корректор, имеющий подъем частотной характеристики в сторону высоких частот на 6 *дб/окт*, либо через фильтр (250—5000 *гц*). Возможны и другие искажения речи [2], которые можно охарактеризовать величинами $d^{(\cdots)}_{\text{ср}}$. В нашем примере мы ограничимся двумя ($d^{(и\kappa)}_{\text{ср}}$, $d^{(п\Phi)}_{\text{ср}}$). Величины $d_{\text{ср}}$, $d^{(и\kappa)}_{\text{ср}}$, $d^{(п\Phi)}_{\text{ср}}$ приведены в табл. 1.

Из литературных источников известно [2, 3], что введение частотного корректора и полосового фильтра увеличивает разборчивость речи. Это значит, что указанные преобразования разносят собственные области образов, построенных в координатах признаков, по которым человек распознает речевые сигналы. Объективные же данные ($d^{(\cdots)}_{\text{ср}}$, табл. 1) дают иную картину. Вывод единственный, при распознавании звуков речи человек не использует такой параметр, как F_{T_3}.

Об этом же говорит величина введенного критерия $d\ (\bar{\sigma},\ \delta) = 1$.

Таблица 1

Элемент проверки	Объективный анализ	Субъективный анализ
Неискаженная речь (ограничитель)	$d_{\text{ср}} = 0{,}63$	A=92%
Частотный корректор-ограничитель	$d^{(и\kappa)}_{\text{ср}} = 0{,}52$	A=97%
Полосовой фильтр-ограничитель	$d^{(п\Phi)}_{\text{ср}} = 0{,}53$	A=95%

$$d\,(\bar{\sigma},\ \bar{\delta}) = 1,\quad \Im = 0$$

Кстати сказать, отрезок синусоиды частотой F_{T_3}, заданный на интервале T_3, никогда не напоминает какой-либо из звуков, а приводимые в литературе примеры более удачного решения некоторых проблем, нежели это делает природа, в случае распознавания речи вряд ли приемлемы. Следует, наконец, заметить, что если факторы громкости и диктора оказывают существенное влияние на значение F_{T_3}, то для пони-

мания сказанного человеку безразлично, кто из дикторов говорил и с незначительными допущениями, как громко произносились звуки.

2. В качестве второго примера оценим эффективность распределений плотности вероятностей длительности интервалов между нулями.

Поступая аналогично пункту 1, определим общие эвклидовы расстояния между усредненными по реализации распределениями j и κ звуков для случаев с частотным корректором, полосовым фильтром и без них. Результаты этой работы приведены в табл. 2. Там же представлены и результаты исследования влияния факторов громкости и диктора на параметры рассматриваемых распределений, заимствованные из [4].

Таблица 2

Элемент проверки	Объективный анализ	Субъективный анализ
Неискаженная речь (ограничитель)	$d_{ср} = 0{,}42$	$A=92\%$
Частотный корректор-ограничитель	$d_{ср}^{ик} = 0{,}41$	$A=97\%$
Полосовой фильтр-ограничитель	$d_{ср}^{пФ} = 0{,}31$	$A=95\%$
$d(з, б) = 1, \quad Э = 0$		
Фактор громкости	влияет	не влияет
Фактор диктора	влияет	не влияет

Нетрудно сделать заключение, что форма распределений плотности вероятностей длительности интервалов между нулями безразлична для человека, распознающего речевые сигналы.

ЛИТЕРАТУРА

1. А. А. Харкевич. О выборе признаков при машинном опознавании. Изв. АН СССР, ОТН, Техническая кибернетика, № 2, 1963.

2. Ю. Г. Ростовцев. О возможностях применения в системах связи предельного амплитудного ограничения речевых сигналов. «Электросвязь», № 6, 1958.

3. Ю. С. Быков. Теория разборчивости речи и повышение эффективности радиотелефонной связи. Госэнергоиздат, 1959.

4. В. П. Евменов, Б. Н. Епифанцев. О распределении временных интервалов между нулями речевых сигналов. Данный сборник.