УДК 681.327.12

Н. М. Быков, к. т. н., проф.; В. В. Ковтун, к. т. н., доц.; Н. Г.Савинова

ОЦЕНИВАНИЕ ВЛИЯНИЯ ПОМЕХ НА ДОСТОВЕРНОСТЬ РАБОТЫ ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНОЙ СИСТЕМЫ РАСПОЗНАВАНИЯ ГОЛОСА

Оценено влияние помех в речевом сигнале на достоверность работы информационно-измерительной системы распознавания голоса. Получены аналитические выражения для оценки ошибки распознавания, приведены результаты вычисления вероятности ошибки классификации двух классов голосов при разных уровнях шумов в сигнале.

Ключевые слова: *распознавание голоса, информационно-измерительная система, достоверность, шум, помехи, математическая модель, передаточный канал, распределительная гиперплоскость, нормальный закон распределения, правило минимума расстояния.*

Передаточный канал, через который происходит распространение речевых сигналов, которые будут использоваться для распознавания голоса, находится под воздействием разного рода шумов [1], основными среди которых являются шумы аппаратуры и окружающей среды. Указанное влияние приводит к уменьшению достоверности работы информационно-измерительной системы для распознавания голоса. Следовательно, оценивание этого влияния с учетом конструктивных особенностей информационно-измерительной системы данного типа является актуальной задачей.

Воспользовавшись линейной моделью [2, 3, 4], речевой сигнал можно рассматривать как квазидетерминированный процесс для одних и тех же классов на часовом отрезке $\tau = 10 \div 30$ мс. При наличии аддитивной помехи речевой сигнал будет иметь такой вид:

$$y^*(t,\tau) = y(t,\tau) + \xi(t,\tau). \tag{1}$$

В большинстве случаев помеха $\xi(t,\tau)$ имеет нулевое среднее значение и некоррелируемая с речевым сигналом $y(t,\tau)$. Следовательно, вектор сигнала можно представить как:

$$X = X_y + X_\xi, \tag{2}$$

где $X_y = \left(x_{y_1}, x_{y_2}, \ldots, x_{y_n}\right)^T$, $X_\xi = \left(x_{\xi_1}, x_{\xi_2}, \ldots, x_{\xi_n}\right)^T$ – векторы описания речевого сигнала и помехи соответственно;

T – символ транспонирования;

$x_i = x_{y_i}^2 + x_{\zeta_i}^2$ – спектральная мощность на i-й частотной полосе.

Для нивелирования влияния на результаты распознавания громкости речевого сигнала, используем нормированный вектор описания

$$\|X\| = \sum_{i=1}^{n} x_i = \sigma_c^2, \tag{3}$$

где σ_c^2 – дисперсия на входе устройства спектрального анализа сигнала.

В результате нормирования вектора (2) соответственно из (3) получим:

$$\widetilde{X} = \frac{r^2}{r^2+1}\widetilde{X}_y + \frac{1}{r^2+1}\widetilde{X}_\xi,$$

где $\widetilde{X}, \widetilde{X}_y, \widetilde{X}_\xi$ – нормированные векторы X, X_y, X_ξ соответственно;

$r = \sqrt{\dfrac{\sigma_y^2}{\sigma_\xi^2}}$ – отношение «сигнал/шум» на входе системы распознавания.

Учитывая, что на практике $r^2 >> 1$ уравнение будет иметь вид:

$$\widetilde{X} = \widetilde{X}_y + \frac{1}{r^2}\widetilde{X}_\xi . \tag{4}$$

Достоверность работы классификатора информационно-измерительной системы для распознавания голоса определим на примере двух классов дикторов Ω_1 и Ω_2, что потом не трудно обобщить для случая большего количества классов. Типичным классификатором, который используется в информационно-измерительной системе для распознавания голоса, является классификатор «за минимумом расстояния» [5]. Следовательно, для данного типа классификатора правило классификации имеет вид:

$$d_e\left(\widetilde{X}, \mu_i\right) = \min d_e\left(\widetilde{X}, \mu_j\right) \Rightarrow X \in \Omega_i, \qquad i, j = 1,2, \tag{5}$$

где $\mu_i = E^0\left\{\widetilde{X} \mid \widetilde{X} \in \Omega_i\right\}$ – среднее значение векторов $\widetilde{X}$, которые относятся к классу Ω_i (Ω_i – эталон класса);

$d_e\left(\widetilde{X}, \mu_i\right) = \left[\left(\widetilde{X} - \mu_i\right)^T \left(\widetilde{X} - \mu_i\right)\right]$ – эвклидово расстояние от вектора $\widetilde{X}$ к вектору μ_i.

В терминах гиперплоскости, которая разделяет классы дикторов Ω_1 и Ω_2, правило (5) для сигнала без помех будет иметь вид:

$$H_{12}\left(\widetilde{X}\right) = \left(\widetilde{X}_y - \mu_1\right)^T \left(\widetilde{X}_y - \mu_1\right) - \left(\widetilde{X}_y - \mu_2\right)^T \left(\widetilde{X}_y - \mu_2\right) = 0, \tag{6}$$

или

$$H_{12}\left(\widetilde{X}\right) = 2\widetilde{X}_y^T \left(\mu_2 - \mu_1\right) + \mu_1^T \mu_1 - \mu_2^T \mu_2 = 0. \tag{7}$$

Подставив (4) в (7), сформируем уравнение гиперплоскости при наличии помех, считая, что обучение классификатора проводилось на сигнале, в котором помехи отсутствовали

$$H_{12}^r\left(\widetilde{X}\right) = 2\widetilde{X}_y^T \left(\mu_2 - \mu_1\right) + \frac{2}{r^2}\widetilde{X}_\xi^T \left(\mu_2 - \mu_1\right) + \mu_1^T \mu_1 - \mu_2^T \mu_2 = 0, \tag{8}$$

или

$$H_{12}^r\left(\widetilde{X}\right) = H_{12}\left(\widetilde{X}\right) + \frac{2}{r^2}\widetilde{X}_\xi^T \left(\mu_2 - \mu_1\right) = 0.$$

Из уравнения (8) видно, что влияние помехи приводит к смещению гиперплоскости к одному из классов Ω_1 или Ω_2 в зависимости от знака вектора $\left(\mu_2 - \mu_1\right)$. Поскольку для квазидетерминированного речевого сигнала мощность мгновенного спектра в i-й частотной полосе является случайной функцией [6], то спектральный вектор описания $\widetilde{X}$ является случайным вектором, который характеризуется многомерным нормальным распределением. Тогда (7), (8) определяются плотностью одномерного нормального распределения. Среднее значение для дискриминантной функции (7) будет иметь вид:

$$E^0\left\{H_{12}\left(\widetilde{X}\right)\right\} = 2E^0\left\{\widetilde{X}_y\right\}\left(\mu_2 - \mu_1\right)^T + \mu_1^T \mu_1 - \mu_2^T \mu_2. \tag{9}$$

А для дискриминантной функции (8), принимая $\dfrac{2}{r^2}E^0\left\{\widetilde{X}_\xi^T\right\} \approx \dfrac{2}{r^2}\widetilde{X}_\xi^T$,

$$E^0\left\{H_{12}\left(\widetilde{X}\right)\right\}=2E^0\left\{\widetilde{X}_y\right\}\left(\mu_2-\mu_1\right)^T+\frac{2}{r^2}\widetilde{X}_\xi^T\left(\mu_2-\mu_1\right)+\mu_1^T\mu_1-\mu_2^T\mu_2, \qquad (10)$$

Дисперсия дискриминантной функции (7) будет иметь вид:

$$\sigma_н^2=E^0\left\{\left[2\widetilde{X}_y\left(\mu_2-\mu_1\right)^T+\mu_1^T\mu_1-\mu_2^T\mu_2\right]-\left[2E^0\left\{\widetilde{X}_y\right\}\left(\mu_2-\mu_1\right)^T+\mu_1^T\mu_1-\mu_2^T\mu_2\right]\right\}^2, \qquad (11)$$

упростив который получим

$$\sigma_н^2=2\left(\mu_2-\mu_1\right)^T\sum_i^0\left(\mu_2-\mu_1\right), \qquad i=1,2, \qquad (12)$$

где $\sum_i^0$ – ковариационная матрица i-го класса.

Для упрощения выражений принимаем $\sum_1^0=\sum_2^0=\sum^0$.

Подставив (8) в (11), получим уравнение дисперсии дискриминантной функции $H_{12}^r\left(\widetilde{X}\right)$, аналогичное уравнению (12).

Следовательно, при $\widetilde{X}\in\Omega_1$ дискриминантные функции $H_{12}\left(\widetilde{X}\right)$ и $H_{12}^r\left(\widetilde{X}\right)$ определяются нормальными законами $N_1\left(m_1^0,\sigma_н\right)$ и $N_1^r\left(m_1^0+\frac{2}{r^2}\widetilde{X}_\xi^T\left(\mu_2-\mu_1\right),\sigma_н\right)$ соответственно. Среднее значение m_1^0 получим, подставив в (11) значение $E^0\left\{\widetilde{X}_y\right\}=\mu_1$:

$$m_1^0=\mu_2^T\left(2\mu_1-\mu_2\right)-\mu_1^T\mu_1. \qquad (13)$$

Аналогично, при $\widetilde{X}\in\Omega_2$, дискриминантные функции $H_{12}\left(\widetilde{X}\right)$ и $H_{12}^r\left(\widetilde{X}\right)$ определяются нормальными законами $N_2\left(m_2^0,\sigma_н\right)$ и $N_2^r\left(m_2^0+\frac{2}{r^2}\widetilde{X}_\xi^T\left(\mu_2-\mu_1\right),\sigma_н\right)$ соответственно. Среднее значение m_2^0 получим, подставив в (9) значение $E^0\left\{\widetilde{X}_y\right\}=\mu_2$:

$$m_2^0=\mu_1^T\left(\mu_1-2\mu_2\right)-\mu_2^T\mu_2. \qquad (14)$$

Таким образом, вероятность возникновения ошибок первого и второго рода для дискриминантной функции (17) при условии отсутствия помех определяются уравнением

$$P(e)=P(\Omega_1)P\left(H_{12}\left(\widetilde{X}\right)<\theta\,|_{\widetilde{X}\in\Omega_1}\right)+P(\Omega_2)P\left(H_{12}\left(\widetilde{X}\right)>\theta\,|_{\widetilde{X}\in\Omega_2}\right), \qquad (15)$$

где θ – порог распознавания.

Учитывая (13) и (14), получим:

$$P\left(H_{12}\left(\widetilde{X}\right)<\theta\,|_{\widetilde{X}\in\Omega_1}\right)=\int_{-\infty}^{\theta}\frac{1}{\sigma_н\sqrt{2\pi}}e^{\frac{-\left[H_{12}\left(\widetilde{X}\right)-m_1^0\right]}{2\sigma_н^2}}dH_{12}\left(\widetilde{X}\right), \qquad (16)$$

$$P\left(H_{12}\left(\widetilde{X}\right)>\theta\,|_{\widetilde{X}\in\Omega_2}\right)=\int_{\theta}^{\infty}\frac{1}{\sigma_н\sqrt{2\pi}}e^{\frac{-\left[H_{12}\left(\widetilde{X}\right)-m_2^0\right]}{2\sigma_н^2}}dH_{12}\left(\widetilde{X}\right). \qquad (17)$$

Подставив (16), (17) в (15), получим:

$$P(e) = P(\Omega_1)\Phi^*\left(\frac{\theta - m_1^0}{\sigma_н}\right) + P(\Omega_2)\Phi^*\left(\frac{m_2^0 - \theta}{\sigma_н}\right), \tag{18}$$

где $\Phi^*(a)$ – функция Лапласа.

Выбирая бинарную функцию потерь (0 – правильное распознавание, 1 – ошибка), пороговое значение θ будет определяться отношением:

$$\frac{P\left(H_{12}(\tilde{X}) = \theta \,|_{\tilde{X}\in\Omega_1}\right)}{P\left(H_{12}(\tilde{X}) = \theta \,|_{\tilde{X}\in\Omega_2}\right)} = \frac{P(\Omega_1)}{P(\Omega_2)} = \theta_0. \tag{19}$$

Подставив (15) и (16) в (19), получим:

$$\theta = \frac{m_1^0 + m_2^0}{2} + \frac{\sigma_н^2 \ln\theta_0}{m_1^0 - m_2^0}. \tag{20}$$

Заменяя в (20) величины m_1^0, m_2^0 и $\sigma_н^2$ их значениями, получаем пороговое значение θ

$$\theta = \sum^0 \ln\theta_0. \tag{21}$$

В случае присутствия помехи пороговое значение θ запишем как:

$$\theta_\xi = \frac{4}{r^2} X_\xi^T(\mu_2 - \mu_1) + \sum^0 \ln\theta_0 = \Delta_\xi + \theta. \tag{22}$$

Учитывая (22), вероятность возникновения ошибок первого и второго рода для дискриминантной функции (18), при условии присутствия помехи, будет иметь вид:

$$P(e_\xi) = P(\Omega_1)\Phi^*\left(\frac{\theta + \Delta_\xi - m_1^0}{\sigma_н}\right) + P(\Omega_2)\Phi^*\left(\frac{m_2^0 - \Delta_\xi - \theta}{\sigma_н}\right). \tag{23}$$

Уравнение (23) определяет зависимость погрешности распознавания от присутствия в речевом сигнале шума. К преимуществу предложенного способа учета влияния присутствующих в речевом сигнале шумов на достоверность работы информационно-измерительной системы для распознавания голоса можно отнести то, что нет потребности вычислять интеграл плотности распределения вероятностей информативных признаков в n-мерном пространстве.

Используя формулы (18) и (23), проведено исследование зависимости ошибки распознавания голоса от присутствия в речевом сигнале шума. Результаты исследования представлены на рис. 1.

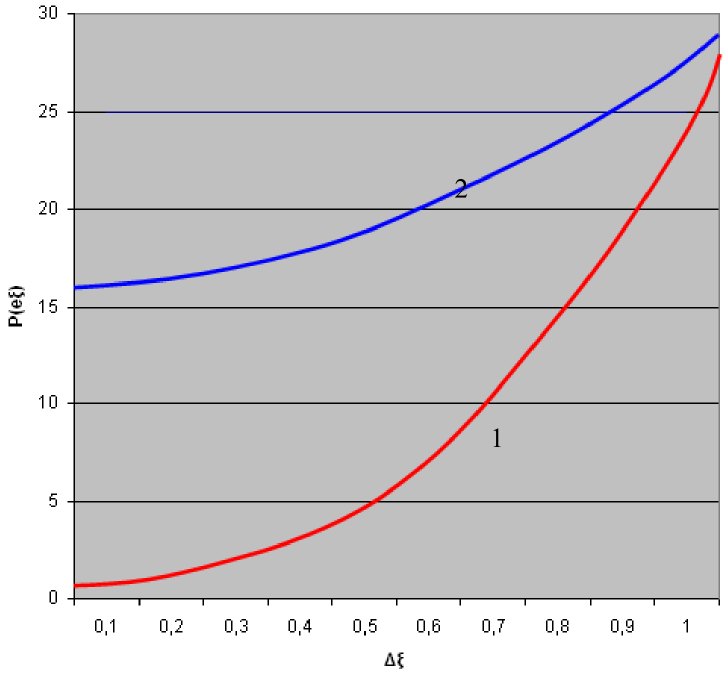

Рис. 1. Зависимость ошибки распознавания от величины шума присутствующего в речевом сигнале: 1: $m_1^0 = 1,\ \sigma_н = 1;$ 2: $m_1^0 = 1,\ \sigma_н = 0{,}5$

ВЫВОДЫ

В работе получена математическая модель влияния помех на достоверность распознавания голоса, которая устраняет необходимость интегрирования области ошибочных решений во многомерном признаковом пространстве. Проведенные аналитические изложения и экспериментальные результаты показали:

– уменьшение влияния помех на достоверность работы информационно-измерительной системы для распознавания голоса методом их фильтрации является эффективным для информативных признаков с высокой раздельностью классов дикторов, то есть при условии $\sigma_н \le m_1^0 / 2$;

– фильтрация помех с большой амплитудой дает больший эффект при распознавании голосов, чем фильтрация помех с малой амплитудой.

СПИСОК ЛИТЕРАТУРЫ

1. Давенпорт В. Б., Рут В. Л. Введение в теорию случайных сигналов и шумов. – М.: ИЛ, 1970. – 498 с.
2. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с.
3. Фант Г. Акустическая теория речеобразования. – М.: Наука, 1964. – 284 с.
4. Рабинер Л. Р. Шафер Р. В. Цифровая обработка речевых сигналов: Пер. с англ. – М.: Радио и связь, 1981. – 496 с.
5. Ротштейн А. П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Винница: УНИВЕРСУМ-Вінниця, 1999. – 320 с.
6. Харкевич А. А. Спектры и анализ. – М.: Физматиз, 1962. – 320 с.

Быков Николай Максимович – к. т. н., проф., кафедры компьютерных систем управления.

Ковтун Вячеслав Васильевич – к. т. н., доц., кафедры компьютерных систем управления.

Савинова Наталия Геннадьевна – аспирант кафедры компьютерных систем управления, e-mail: savinova1987@mail.ru.

Винницкий национальный технический университет.