

УДК 009.934+534

ОЦЕНКА РАЗБОРЧИВОСТИ РЕЧИ В ЗАШУМЛЕННОМ ПОМЕЩЕНИИ

Г.В. ДАВЫДОВ, Д.М. КАВАН, Ю.В. ШАМГИН

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь*

Поступила в редакцию 16 мая 2012

Проведены исследования оценки разборчивости речевого сигнала в зашумленном помещении методом корреляционного анализа. Показано существование сильной связи между разборчивостью речи и силой голоса с коэффициентом корреляции свыше 0,7. Последнее позволило считать эти переменные линейно коррелированными и построить линии регрессии, значительно упрощающие оценку разборчивости речи. Было исследовано также влияние скорости произнесения речи на разборчивость речи.

Ключевые слова: речевые сигналы, разборчивость речи, влияние на разборчивость речи окружающего шума.

Введение

Необходимость в оценке разборчивости речи (РР) возникает во многих случаях: таких как определение качества трактов передачи речи, измерения качества акустики залов, оценки качества устройств, обеспечения защиты речи от несанкционированного прослушивания.

Известно, что РР зависит не только от громкости ее произнесения, но и от других параметров речи. По стандарту [1], под РР понимается относительное количество (в процентах) правильно принятых элементов речи (слогов, слов, фраз) артикуляционных таблиц, распространяемых по тракту передачи.

Окружающий человека шум ухудшает восприятие речи. В этом случае происходит маскирование шумами речевого сигнала и снижение его разборчивости. В связи с этим разрабатываются различные расчетные методы оценки, которые, к сожалению, не дают достаточной уверенности в надежности результатов такого расчета. Кроме того, обработку зашумленной речи предлагается проводить на основе методов, используемых для оценки качества передачи речи в каналах связи. Поэтому стоит задача определения влияния на РР громкости и скорости ее произнесения, а также особенностей речи диктора.

Для решения указанной задачи были проведены исследования разборчивости зашумленных речевых сигналов на достаточно большом количестве реализаций, воспроизводимых 42-мя дикторами (мужчинами и женщинами).

Теоретические положения оценки разборчивости речи

Произносимые речевые сигналы имеют спектры с рядом максимумов и минимумов. Максимумы спектра называются формантами, а нулевые значения – антиформантами. Огибающая спектра для каждого звука речи имеет индивидуальную и вполне определенную форму. Таким образом, форманта – это область концентрации энергии в спектре звука речи [2]. Методы оценки РР делятся на формантные, модуляционные и эмпирические [1–7].

Формантные методы оценки РР основаны на разбиении спектра речи на полосы частот, в общем случае, произвольной ширины (октавных, третьоктавных, равноартикуляционных и т.п.), в пределах каждой из которых плотность вероятностей формант можно считать неизменными, а РР оценивается для разных соотношений сигнал/шум.

Модуляционные методы оценки РР основаны на представлении речевого сигнала в виде свертки импульсной характеристики звукового сигнала источника (для гласных звуков это импульсный сигнал с почти гармоническим спектром, для согласных это шумовой сигнал различной формы) с импульсной характеристикой голосового тракта. Речевой сигнал предлагается рассматривать как некоторый широкополосный сигнал с полосой от 125 до 8000 Гц, модулированный другим сигналом с низкой частотой. Частота модуляции определяется скоростью, с которой человек произносит форманты (скоростью артикуляции). Эксперименты показали, что частоты модуляции в обычной речи находятся в диапазоне от 0,63 до 16 Гц, причем наиболее вероятные частоты модуляции находятся в области от 0,5 до 7 Гц. Уменьшение глубины модуляции делает речевой сигнал более похожим на шумовой и, следовательно, приводит к уменьшению его разборчивости.

Среди эмпирических методов наиболее популярен метод % Alcons – метод измерения величины потери артикуляции согласных, выраженной в процентах. Метод % Alcons широко используется, особенно в США, для приближенной оценки РР и отображает потерю вокализованных согласных, вызванных реверберацией и поглощением звука в помещении.

Для оценки РР может использоваться модифицированный метод М.А. Сапожкова, разработанный для оценки разборчивости формант [2]. Как известно, формантный метод определения разборчивости речи является наиболее объективным и широко используется для количественной оценки разборчивости.

В указанном методе разборчивость формант (A) может быть определена с помощью простой зависимости

$$A = DF/K, \quad (1)$$

где D – величина эффективного динамического диапазона речи, дБ; F – ширина частотного диапазона речи, Гц; $k=1,05 \cdot 10^5$ (дБ·Гц) – нормирующий коэффициент.

В свою очередь динамический диапазон D речевого сигнала, в соответствии с [7], может быть определен как суммарный уровень этого сигнала на n октавных полос, шириной Δf_k , в пределах которых спектральный уровень сигнала B_k примерно постоянен

$$D = 10 \lg \sum_{k=1}^n 10^{0,1B_k} \frac{\Delta f_{\text{окт}}}{f_1}, \quad (2)$$

где $f_{\text{окт}}$ – ширина октавной полосы, Гц; f_1 – частота равная 1 Гц.

Значение спектрального уровня сигнала речи в данной октавной полосе можно вычислить как

$$B_{\text{окт}} = L_c - 10 \lg \left(\frac{\Delta f_{\text{окт}}}{f_1} \right). \quad (3)$$

Найденное по (1) значение формантной разборчивости (A) может быть выражено через словесную разборчивость (W), как более понятную в практическом смысле.

Если прослушивать речь в условиях шумов и помех, то ее разборчивость, согласно (1), снижается. Это обусловлено тем, что форманты имеют разные уровни интенсивности: у громких звуков речи выше, чем у глухих. Поэтому при увеличении уровня шумов сначала маскируются форманты с низкими уровнями, а затем с более высокими. Вследствие этого по мере увеличения уровня шумов и помех вероятность восприятия формант (а, следовательно, и РР) постепенно уменьшается. На практике это обусловлено снижением динамического диапазона речи D и уменьшением ширины частотного диапазона речи. Величина F определяется суммой октавных полос, в пределах которых речевой сигнал может быть услышан:

$$F = \sum_{k=1}^n (L_c \Delta f_{\text{окт}k}). \quad (4)$$

Дальнейшие вычислительные действия по этой методике оценки РР показали их значительную трудоемкость, поскольку требовалось проводить такие расчеты в каждой октавной полосе.

Нами был опробован метод непосредственного прослушивания зашумленной речи дикторов с последующей статистической обработкой результатов такого прослушивания и нахождения корреляционной связи между параметрами речевых сигналов.

Для оценки степени связи между громкостью голоса и РР проводился расчет коэффициента корреляции между суммарным уровнем речевого сигнала и разборчивостью для всех случаев наблюдения. Коэффициент корреляции определялся по формуле

$$r_{xy} = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{(n-1) s_x s_y}, \quad (5)$$

где за x_i принимались измеренные значения суммарного уровня речевого сигнала, а за y_i – измеренные значения разборчивости речи для каждого i -го случая измерений. Те же значения с чертой сверху обозначают средние значения измеряемых величин. s_x и s_y – величины среднего квадратичного отклонения (СКО), рассчитанные для этих переменных; n – количество измерений (в данном случае $n=80$).

Расчет величин СКО проводился по формулам (6) и (7)

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}} \quad (6)$$

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{(n-1)}} \quad (7)$$

Определив эмпирический коэффициент регрессии по формуле (8)

$$b_{y/x} = r \frac{s_y}{s_x}, \quad (8)$$

можно построить эмпирическую линию регрессии в виде прямой, отражающей зависимость РР от громкости голоса диктора для заданного отношения сигнал/шум:

$$y = \bar{y} + b_{y/x} (x - \bar{x}). \quad (9)$$

Описание эксперимента и его результаты

В качестве рабочего материала для записи речевых сигналов использовались артикуляционные таблицы фраз из стандарта [1]. Каждый из 42-х дикторов зачитывал по две таблицы из указанного стандарта в привычной для него манере, т.е. громкости голоса и скорости чтения. Зачитываемые таблицы содержали по 50 фраз и общее количество слов не менее двухсот в каждой таблице. Запись речи дикторов проводилась с помощью программы Sound Forge 9, позволяющей производить измерение параметров записываемого сигнала, а также накладывать шум на записываемый сигнал. Для облегчения восприятия фраз при непосредственном прослушивании, каждая запись речевого сигнала разбивалась маркером на отдельные фразы. Такая маркировка позволяла прослушивать каждую фразу несколько раз, что увеличивало точность оценки РР.

При анализе записанных речевых сигналов по громкости их звучания были выделены дикторы с тихими голосами, средний уровень звукового давления акустического сигнала у которых не превышал 65 дБ, и дикторы с громкими голосами, у которых средний уровень сигнала

был больше 68,1 дБ. Промежуточную группу составляли дикторы с голосом, средний уровень, которого лежал в пределах 65,1÷68,0 дБ.

При наложении шума определенного уровня на речевой сигнал разборчивость этого сигнала ухудшалась с увеличением уровня шума. Однако, поскольку интонация речи диктора менялась в процессе произнесения им фраз, происходило изменение соотношения сигнал/шум. В результате этого отдельные слова или фразы могли превышать уровень шума и быть услышанными при непосредственном прослушивании. Для исследования нами использовались следующие уровни отношения сигнал/шум: –6 дБ; –12 дБ; –14 дБ; –17 дБ и –20 дБ.

В процессе исследования была замечена зависимость РР от громкости голоса диктора и скорости чтения им артикуляционных таблиц. Громкость голоса диктора определялась суммарным уровнем речевого сигнала, измеренного в пяти октавных полосах, а скорость чтения – скоростью произношения диктором слов в секунду. Средние значения этих величин были соответственно равны: средний уровень речевого сигнала – 66,3 дБ, а скорость произношения – 1,66 сл./с.

Максимальный измеренный уровень громкости голоса диктора определялся средним уровнем речевого сигнала в 72,7 дБ, а минимальный уровень громкости голоса диктора в 58,5 дБ. Соответственно, скорость произношения слов была в пределах от 1,09 слов в секунду до 2,94 слов в секунду.

Расчитанные значения СКО для отношения сигнал/шум равное –12 дБ были соответственно равны: $s_x=3,62$; $s_y=26,5$, что позволило определить коэффициент корреляции равным $r_{xy} \cong 0,73$. Такая величина коэффициента корреляции указывает на существование достаточно сильной связи между суммарным уровнем речевого сигнала отдельного диктора и разборчивостью произносимой им речи.

Для оценки влияния на разборчивость речевых сигналов скорости произношения слов дикторами были проведены аналогичные расчеты. В результате были получены следующие статистические данные:

- СКО скорости произношения слов $S_x=0,33$;
- СКО разборчивости речи $S_y=26,5$;
- коэффициент корреляции $r_{xy}=-0,15$;
- эмпирический коэффициент регрессии $b_{x/y}=12$;
- эмпирическая линия регрессии $y=65-12x$.

Поскольку в данном случае коэффициент корреляции отрицательный – это показывает, что при меньшей скорости произношения обеспечивается большая РР дикторов. Но влияние громкости голоса превалирует.

Обобщенная линия регрессии, описываемая (6), была получена экспериментально для оценки разборчивости речи при отношении сигнал/шум –20 дБ. Она предоставляет возможность проводить расчетные оценки РР с учетом корреляционных зависимостей для громкости голоса, выраженных через уровень звукового давления x_1 и скорости произнесения речи x_2 :

$$y = (0,35x_1 - 3x_2 - 12) \pm s_y, \quad (10)$$

где x_1 – уровень звукового давления, характеризующий громкость голоса диктора, дБ_{ак}; x_2 – скорость произношения слов, слов в секунду; $\pm s_y$ – интервал оценки РР для отношения сигнал/шум –20 дБ, %.

Подстановка в (6) соответствующих значений переменных, определенных экспериментально для одного из дикторов, дает следующую оценку РР при условиях: соотношение сигнал/шум равно –20 дБ; $x_1=69,5$ дБ_{ак}; $x_2=1,46$ слов в секунду; $\pm s_y=4,7\%$; разборчивость речи лежит в интервале $y=(7,9 \pm 4,7)\% = 3,2 \div 12,6$.

Измеренное значение РР для этого же диктора при тех же условиях зашумления составило 8,5%, что находится в указанном интервале.

Как можно видеть, использование выражений для обобщенных эмпирических линий регрессии позволяет существенно сократить трудоемкость расчетной оценки РР по сравнению с существующими методами.

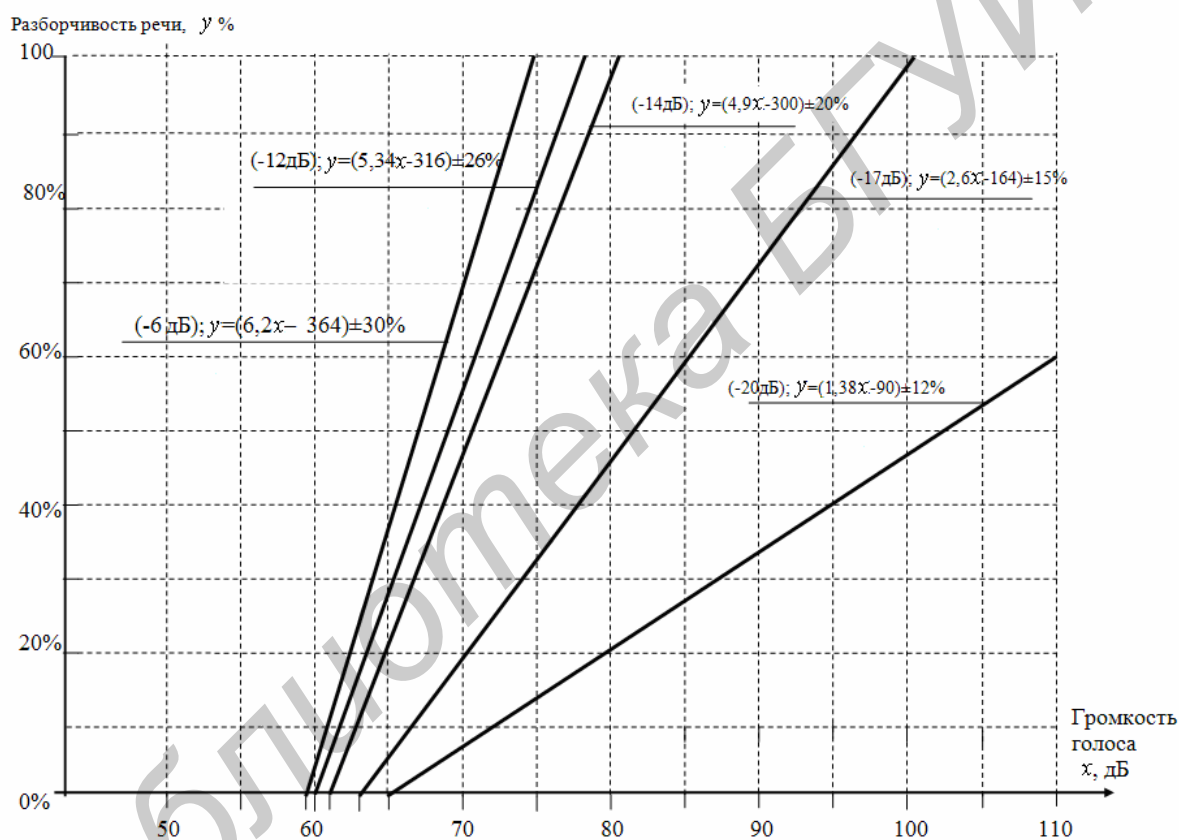
Оценка РР была проведена при условиях зашумления речи, обеспечивающих уровни отношения сигнал/шум, изменяющихся в пределах от –6 дБ до –20 дБ. Суммарный уровень ре-

чего сигнала для различных дикторов изменялся от 55 до 73 дБ. В результате были получены выражения для линий регрессии, приведенные в таблице и позволяющие проводить расчетную оценку РР при допустимых уровнях погрешности таких расчетов.

Эмпирические линии регрессии для различных уровней сигнал/шум

Отношение сигнал/шум	Эмпирическая линия регрессии	Расчетная погрешность оценки РР
-6 дБ	$y=6,2x-364$	$\pm 30\%$
-12 дБ	$y=5,34x-316$	$\pm 26\%$
-14 дБ	$y=4,9x-300$	$\pm 20\%$
-17 дБ	$y=2,6x-164$	$\pm 15\%$
-20 дБ	$y=1,38x-90$	$\pm 12\%$

Эмпирические линии регрессии для указанных выше пределов отношения сигнал/шум приведены на рисунке. Как можно видеть из рисунка, чтобы превысить уровень шума для улучшения РР, необходимо усиливать звучание голоса с помощью усилительных устройств начиная с 75 дБ.



Зависимость разборчивости речи от силы голоса диктора при различных отношениях сигнал/шум (эмпирические линии регрессии)

Заключение

Использование корреляционного анализа позволило проводить более объективную оценку РР, маскируемой шумами речевых сигналов. Результатом такого анализа стало построение линий регрессии, позволяющих значительно упростить расчеты РР при заданном значении громкости голоса, а также определить допустимые уровни шумового сигнала.

EVALUATION OF SPEECH INTELLIGIBILITY IN THE PREMISES PROTECTED BY ACOUSTIC NOISE GENERATOR

H.V. DAVYDAU, D.M. KAVAN, Y.V. SHAMGIN

Abstract

The evaluation of the speech intelligibility in the premises protected by acoustic noise generator was performed by the correlation analysis method. The existence of dependence between the speech intelligibility and the signal power was proofed with a correlation coefficient over 0.7. The latter finding makes possible to suggest that these variables are linearly correlated, and construct the regression line, greatly simplifying the evaluation of speech intelligibility. The influence of speech temp on the speech intelligibility was investigated.

Список литературы

1. СТБ ГОСТ Р 50840-2000. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. Минск, 2000.
2. Сапожков М.А. Речевой сигнал в кибернетике и связи. Преобразование речи применительно к задачам техники связи и кибернетики. М., 1963.
3. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. М., 1987.
4. Покровский Н.Б. Расчет и измерение разборчивости речи. М., 1962.
5. Рашевский Я.И., Каргашин В.Л. // Специальная техника. 2002. №3–6.
6. Рашевский Я.И., Каргашин В.Л. // Специальная техника. 2003. №1.
7. Сапожков М.А. Звукофикация помещений: проектирование и расчет. М., 1979.