

ОБ ИЗМЕРЕНИЯХ ВЕРОЯТНОСТНЫХ ХАРАКТЕРИСТИК ФОРМАНТ УКРАИНСКОЙ И РУССКОЙ РЕЧИ

А.Н. ПРОДЕУС

Национальный технический университет Украины "КПИ", Киев

Приведена дополнительная аргументация предпочтительности формирования коэффициентов восприятия по методике М.А. Сапожкова, а не по методике Н.Б. Покровского. Показано, что экспериментально полученные, с использованием методики М.А. Сапожкова, коэффициенты восприятия для украинской и русской речи отличаются весьма незначительно.

ВВЕДЕНИЕ

Среди множества методов оценки разборчивости речи наибольшей глубиной теоретической и экспериментальной проработки отличаются формантный и модуляционный методы [1]. Наиболее удачной идеей модуляционного метода является использование тестового сигнала в виде модулированного по уровню шума. С другой стороны, существенным достоинством формантного метода является вероятностный подход к описанию свойств речевого сигнала. Поэтому в перспективе может оказаться целесообразным объединение достоинств обоих методов и создание единого формантно-модуляционного метода. Однако прежде чем говорить о таком объединении, необходимо предварительно решить некоторые «внутренние проблемы» формантного метода.

Дело в том, что сегодня ситуация с применением формантного метода для оценивания разборчивости русской речи выглядит весьма неоднозначно. Еще в конце 50-х годов прошлого века в СССР сформировались три научные школы, представители которых – Ю. С. Быков, Н. Б. Покровский и М. А. Сапожков – отстаивали каждый свою версию формантного метода, не слишком заботясь о корректном их сопоставлении [1-4]. Как следствие, современные инженеры неизбежно сталкиваются с проблемой выбора «лучшей» версии. В ряде последних работ российских исследователей [5-7] в качестве базовой принята версия Н. Б. Покровского. Между тем, такое решение небесспорно – и в данной работе приведены дополнительные аргументы в пользу версии М. А. Сапожкова, как заслуживающей значительно более пристального внимания.

Еще более неопределенной выглядит ситуация с оцениванием разборчивости украинской речи – ведь до последнего времени вероятностные характеристики украинской речи оставались неизвестными.

Для восполнения указанного пробела в данной работе по достаточно большой выборке речевых сигналов (17 дикторов-мужчин и 17 дикторов-женщин) оценены и сопоставлены функции распределения речевых сигналов по уровню для украинской и русской речи. На их основе, с использованием методики М. А. Сапожкова, сформированы соответствующие коэффициенты восприятия.

1. СОПОСТАВЛЕНИЕ ВЕРСИЙ ФОРМАНТНОГО МЕТОДА

В каждой из трех версий формантного метода – Ю. С. Быкова, Н. Б. Покровского и М. А. Сапожкова – формантную разборчивость A определяют как среднюю вероятность отсутствия маскировки речи шумом:

$$A = \sum_{k=1}^K p_k \cdot w(E'_k), \quad (1)$$

где p_k - вероятность пребывания формант в k -той полосе частот с граничными частотами f_{nk} и f_{ek} , $k = 1, \dots, K$:

$$p_k = F_1(f_{ek}) - F_1(f_{nk}), \quad (2)$$

$F_1(f)$ - функция распределения вероятностей формант по частоте; $w(E'_k)$ - коэффициент восприятия речи (вероятность отсутствия маскировки речи шумом в k -той полосе частот); E'_k - разность уровней формантного спектра и спектра помехи (отношение сигнал-шум) в k -той частотной полосе.

Различие версий состоит, прежде всего, в различии трактовок понятия «формантный спектр»: Н.Б. Покровский и Ю.С. Быков определяют формантный спектр как спектр шума, на фоне которого достигается соответственно 50%-ная и 0%-ная разборчивость речи, тогда как М.А. Сапожков фактически отождествляет формантный спектр с долговременным спектром речи. В работах [8,9] показано, что если учесть это различие определений формантного спектра, коэффициенты восприятия можно корректно совместить на одном графике (рис. 1), где E - разность уровней спектра речевого сигнала и спектра маскирующей его помехи, ΔB - разница уровней спектров речи и формант. На этом же рисунке приведен график индекса артикуляции AI - коэффициента восприятия американской версии формантного метода [1,11].

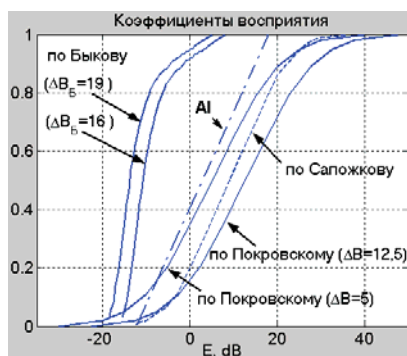


Рис. 1. Совмещение коэффициентов восприятия Ю. С. Быкова, Н. Б. Покровского, М. А. Сапожкова и индекса артикуляции AI [7,8]

Как следует из рис. 1, налицо существенное отличие формы коэффициентов восприятия в советских версиях формантного метода. Причина такого различия - различие методик формирования коэффициентов восприятия.

В работах [8-11] показано, что версия Н.Б. Покровского основана на недостаточно корректной методике формирования коэффициента восприятия $P(E'_k)$, в рамках которой:

- не учтена зависимость $P(E'_k)$ от частоты;
- выдвинута ничем не подкрепленная гипотеза о центральной симметричности, относительно $E'_k = 0$, коэффициента $P(E'_k)$;
- «динамический диапазон» значений аргумента E'_k коэффициента $P(E'_k)$ составил 70 дБ, что почти в два раза больше «общепринятого» значения динамического диапазона уровней речевого сигнала 30...40 дБ.

Как ни странно, но по поводу последнего обстоятельства лишь в работе Ю. С. Быкова [3, стр. 93] было выражено некоторое недоумение. Остальные советские и российские исследователи просто проигнорировали указанное несоответствие, хотя степень его столь значительна, что заставляет серьезно усомниться в правильности полученных Н.Б. Покровским результатов.

Анализ работы [2] свидетельствует, что изначально Н.Б. Покровским был получен коэффициент восприятия $P_1(E'_k)$ «идеализированной артикуляционной бригады», с динамическим диапазоном значений аргумента E'_k , близким 40 дБ. Однако затем, для учета разброса уровней голосов дикторов и порогов восприятия слушателей, этот коэффициент восприятия был преобразован в коэффициент восприятия $P(E'_k)$ «типичной артикуляционной бригады» с использованием соотношения свертки:

$$P(E'_k) = P_1(E'_k) \otimes h(E'_k), \quad (3)$$

где \otimes - символ свертки; $h(E'_k)$ - нормальный закон распределения с нулевым средним значением и стандартным отклонением $\sigma \approx 4$ дБ. Это и послужило причиной расширения «динамического диапазона» значений аргумента E'_k коэффициента восприятия с 40 до 70 дБ.

На наш взгляд, некорректность описываемого соотношением (3) приема состоит в том, что процедуре свертки подвергается функция, аргумент которой *нормирован по среднему уровню речи*. Именно благодаря такому нормированию *уже учтен* разброс уровней речи дикторов, обусловленный различием громкости речи либо различным расстоянием между микрофоном и губами различных дикторов. Разумеется, небольшие флуктуации динамического диапазона, обусловленные сменой дикторов, имеют место, однако «общепринятость» значения динамического диапазона уровней речевого сигнала 30...40 дБ позволяет заключить, что стандартное отклонение таких флуктуаций существенно меньше значения $\sigma \approx 4$ дБ.

Что касается учета разброса порогов восприятия слушателей, сам Н.Б. Покровский [2, стр. 131] убедительно показывает, что переменная E'_k практически не зависит от порога восприятия при уровнях шума, превышающих порог восприятия не менее чем на 10 дБ. Очевидно, в соотношении (3) данное обстоятельство также не учтено.

Значительно корректнее выглядит методика М. А. Сапожкова, единственным недостатком которой является предположение о независимости коэффициента восприятия от частоты. В работе [10] этот недостаток был устранен, а также было показано, что принципиальным достоинством методики М. А. Сапожкова является обоснованность аналитического описания связи между функцией распределения вероятностей формант $F(Y)$ по уровню речевого сигнала Y и коэффициентом восприятия $w(E)$ для любого типа помех. В частном случае флуктуационной (шумовой) помехи эта связь, например, определяется соотношением:

$$w(E) = \begin{cases} 1, & E > D - \Pi_{p \max}; \\ 1 - F(-E), & -\Pi_{p \max} \leq E \leq D - \Pi_{p \max}; \\ 0, & E < -\Pi_{p \max}, \end{cases} \quad (4)$$

где $D = Y_{p \max} - Y_{p \min}$ - динамический диапазон речевого сигнала; $Y_{p \max}$ - максимальный уровень речевого сигнала; $Y_{p \min}$ - минимальный уровень речевого

сигнала; $P_{r\max}$ – максимальный пикфактор речевого сигнала. Иллюстрация соотношения (4) представлена на рис. 2.

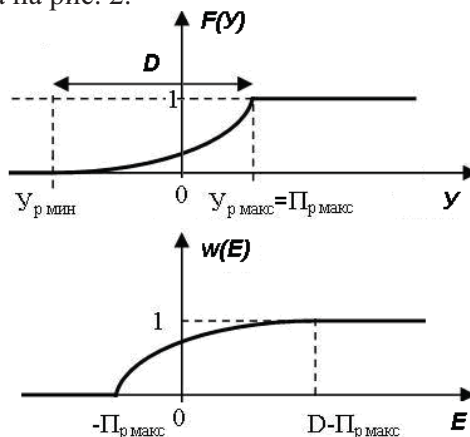


Рис. 2. Связь между функциями $w(E)$ и $F(Y)$

2. МЕТОДИКА И РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Запись речевых сигналов производилась в заглушенной комнате, обеспечивающей существенное подавление реверберационной помехи, для двух групп дикторов: семнадцати мужчин и семнадцати женщин, одинаково хорошо владеющих русским и украинским языками. Зачитывались фонетически сбалансированные украинский и русский тексты, состоящие из фрагментов художественных произведений, технических и юридических публикаций. Дикторы читали текст по возможности слитно, не делая заметных пауз между словами и предложениями.

Для записи речевых сигналов использовались конденсаторный микрофон Groove Tubes_GT57, специализированная звуковая плата M-Audio Audiophile 2496, компьютерная программа Sound Forge 7.0. Параметры записи речевых сигналов: длительность сигнала 17 мин; частота дискретизации 22050 Гц; квантование 16 бит.

Этапы обработки записанных речевых сигналов:

- 1) стационаризация речевого сигнала $s(t)$ для подавления произвольного изменения уровня сигнала во времени из-за усталости диктора или иных причин;
- 2) полосовая фильтрация речевого сигнала $s(t)$ для получения сигналов $s_k(t)$ в каждой k -той полосе частот;

- 3) вычисление функций распределения $F_k(Y)$ сигналов $A_k(t) = \int_{t-T}^t s_k^2(t) dt / T$,

где $T = 0.125$ с.

Полосовая фильтрация осуществлялась гребенкой из семи октавных фильтров Ремеза с неравномерностью 1 дБ в полосе пропускания, уровнем боковых лепестков –60 дБ, средними частотами 125; 250; 500; 1000; 2000; 4000; 8000 Гц.

Поскольку уровень Y речевых сигналов $s_k(t)$, $k = 1, \dots, 7$, удобно представлять в логарифмическом виде, вычисление функций распределения $F_i(Y)$ осуществлялось путем предварительного нелинейного квантования (логарифмирование с последующим

10-битним лінійним квантуванням) сигналів $A_k(t)$, нормованих по своєму середньому значенню.

Листинги відповідальних вичислювальних програм (язык програмування середовища Matlab) приведені в роботах [1,10].

Результати вимірювань функцій розподілення $F_i(V)$ для октавної полоси частот с центральною частотою 500 Гц представлені на рис. 3-6 (слева сопоставлены средние значения оценок $F_i(V)$, справа – границы 95% доверительных интервалов). Аналогічні результати отримані для остальных шести октавних полос частот.

Приведенные на рис. 3-6 графики свидетельствуют, что законы распределения уровней речи для русского и украинского языков отличаются незначительно, а при уровнях речи более 0 дБ практически совпадают. Более того, при уровнях речи более 0 дБ практически совпадают и функции распределения дикторов-мужчин и дикторов-женщин.

Учитывая соотношение (4), приходим к выводу, что коэффициенты восприятия для украинской речи несущественно отличаются от таковых для русской речи, а при малых отношениях сигнал-шум (менее 0 дБ) коэффициенты восприятия практически одинаковы не только для русского и украинского языков, но и для дикторов-мужчин и дикторов-женщин.

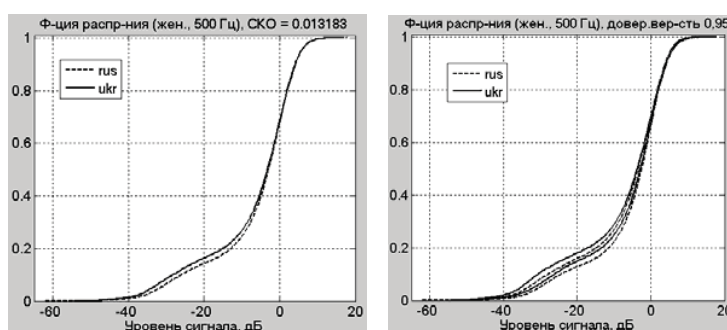


Рис. 3. Функции распределения уровней речи для дикторов-женщин

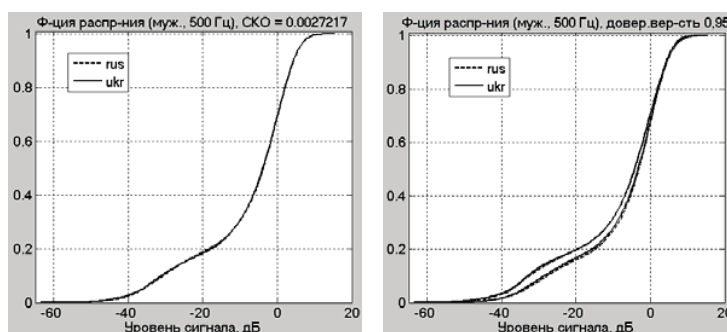


Рис. 4. Функции распределения уровней речи для дикторов-мужчин

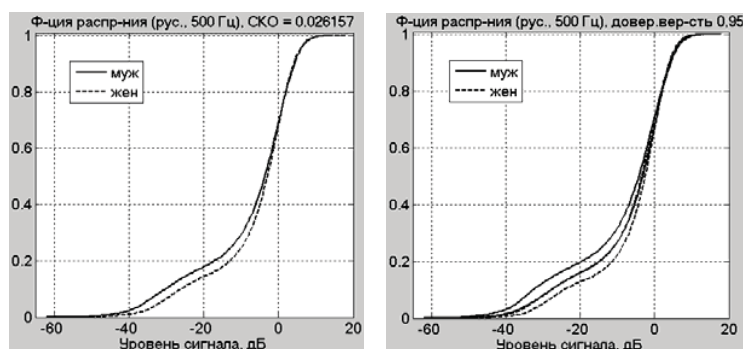


Рис. 5. Функции распределения уровней речи для русского языка

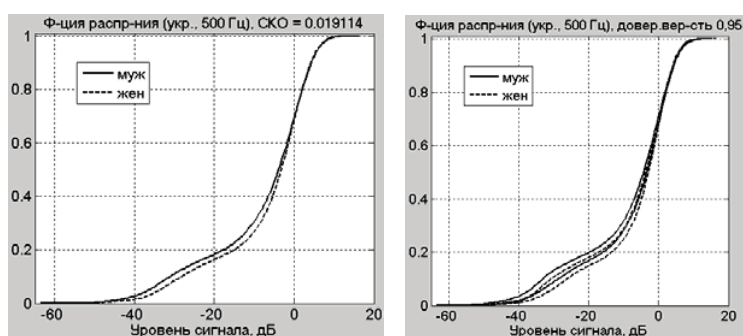


Рис. 6. Функция распределения уровней речи для украинского языка

ЗАКЛЮЧЕНИЕ

Экспериментально получены, с использованием методики М.А. Сапожкова, функции распределения уровней речевого сигнала для украинской и русской речи, для семи октавных частотных полос. Полученные результаты свидетельствуют, что различие коэффициентов восприятия для русского и украинского языков невелико и меньше гендерного. Более того, при малых отношениях сигнал-шум (менее 0 дБ) коэффициенты восприятия практически одинаковы не только для русского и украинского языков, но и для дикторов-мужчин и дикторов-женщин.

ЛИТЕРАТУРА

1. Гавриленко О.В., Дидковский В.С., Продеус А.Н. Сравнительный анализ некоторых методов оценки разборчивости речи. // Труды международной конференции CONSONANS-2007, с. 54-65.
2. Покровский Н.Б. Расчет и измерение разборчивости речи. –М.: Связьиздат, 1962.–390 с.
3. Быков Ю.С. Теория разборчивости речи и повышение эффективности радиотелефонной связи. – М.-Л.: Госэнергоиздат, 1959. – 350 с.
4. Сапожков М.А. Речевой сигнал в кибернетике и связи. – М.: Связьиздат, 1963. –472 с.

5. Железняк В.К., Макаров Ю.К., Хорев А.А. Некоторые методические подходы к оценке эффективности защиты речевой информации // Специальная техника. – М.: 2000, № 4. – С. 39-45.
6. Макаров Ю.К., Хорев А.А. К оценке эффективности защиты акустической (речевой) информации // Специальная техника, 2000, № 5, с. 46 – 56.
7. Каргашин В. Л. Совершенствование методических принципов оценки защищенности помещений от утечки речевой информации. – Специальная техника, №6, 2001 (<http://www.bnti.ru/showart.asp?lvl=&aid=579>).
8. Гавриленко А.В., Дидковский В.С., Продеус А.Н., Самойленко Н.А. Формантный метод оценки разборчивости речи. // Электроника и связь, №5, 2007. С.63-70.
9. Гавриленко О. В., Дидковский В. С., Продеус А. Н. Сопоставление версий формантного метода оценки разборчивости речи // Электроника и связь, Тематический выпуск «Проблемы электроники», ч.1. - 2008. – С. 227-230.
10. Гавриленко О.В., Дидковский В.С., Продеус А.Н. Расчет и измерение разборчивости речи при малых отношениях сигнал-шум. Часть 1. Корректное измерение функции распределения речевого сигнала // Электроника и связь, Тематический выпуск «Проблемы электроники», ч.1. - 2007. – С. 137-141.
11. Дидковский В.С., Дидковская М.В., Продеус А.Н. Акустическая экспертиза каналов речевой коммуникации. Монография. – К: Имэкс-ЛТД, 2008. – 420 с.