

## **Анализ возможностей корреляционного метода оценки эффективности маскирования речи белым шумом**

Д.т.н. С.В. Дворянкин, С.Б. Козлачков, к.т.н. А.М. Бонч-Бруевич, МГТУ им. Н.Э. Баумана

Методы объективной оценки показателей защищенности речевой информации (РИ) от утечки по техническим каналам (ТКУРИ) регулярно обсуждаются в печати [1- 4], предоставляя потребителю достаточно мощный инструмент для научно обоснованного проведения мероприятий по защите РИ. Использование разборчивости речи (РР) как показателя опасности ТКУРИ, адекватного информационному содержанию переговоров, может быть полезным и перспективным при реализации защиты РИ [4].

Исследователями Национального технического университета Украины «КПИ» апробирован корреляционный метод оценки эффективности маскирования речи белым шумом [5]. Предполагается что корреляционный метод, в отличие от инструментально-расчетных методов определения РР, использующих усредненные критерии, позволит адекватно оценивать объективные показатели РР и, нивелируя недостатки предшествующих методов, может адаптировать уровень маскирующего сигнала к текущим значениям уровня речи.

Исследование функции эффективности маскирования речи (отношения сигнал/помеха  $SN(\Delta\omega, t)$ ) выполнялось по временным значениям функции речевого  $si(\Delta\omega, t)$  и маскированного  $su(\Delta\omega, t)$  сигналов. Для оценки тесноты связи между исследуемыми сигналами применялся коэффициент корреляционного отношения Пирсона. Временной интервал анализа (15-30 мс) выбирался в соответствии с психофизиологическими критериями стационарности речевого сигнала РС [5].

При этом в процессе такого моделирования не принимались во внимание спектральные характеристики исследуемых сигналов.

Для корректного учета влияния спектральных параметров в МГТУ им. Н.Э. Баумана были проведены исследования адекватности корреляционного

метода оценки эффективности маскирования речи при анализе спектральных реализаций РС и маскирующего сигнала в виде белого шума (БШ).

Особое внимание обращалось на соответствие временной модели РС современным представлениям акустического сигнала как сумме узкополосных спектральных составляющих [6]. В первую очередь это касается различий между временным и спектральным представлениями РС. Тональный сигнал во временной области может иметь различные реализации в виде  $\sin(\omega t)$  или  $\cos(\omega t)$ , а в спектральной области он будет представлен в виде гармоники основного тона, что соответствует модели слухового аппарата, не анализирующего фазовые компоненты РС.

Другим исследуемым фактором, в соответствии с положениями формантной теории, являлся учет весовых коэффициентов формант в соответствующих спектральных полосах.

Речевой сигнал, записанный через микрофон речью диктора длительностью 8 сек., оцифровывался с разрядностью 16 бит и частотой дискретизации 22050 Гц. Маскирующий сигнал (БШ) формировался с аналогичными параметрами в программе MatLab. В этой же программе формировалась смесь сигнала и маскирующей помехи с заданным соотношением сигнал/шум ( $L_{SN}$ ).

Для оценки тесноты связи между исследуемыми сигналами также, как и в Национальном техническом университете Украины, применялся коэффициент корреляционного отношения Пирсона.

Спектр сигналов формировался с помощью быстрого преобразования Фурье в окне Хемминга, база преобразования Фурье составляла 512 отчетов.

Коэффициент корреляции между амплитудными спектрами сигнала и смеси вычислялся по формуле:

$$K = \frac{M(y_1 y_2) - M(y_1) \cdot M(y_2)}{\sqrt{M(y_1^2) - M(y_1)^2} \cdot \sqrt{M(y_2^2) - M(y_2)^2}}, \text{ где}$$

$M$  – символ математического ожидания,  $y_1, y_2$  – отчеты амплитудного спектра сигнала и помехи соответственно.

При исследованиях за основу были приняты положения формантной теории [1, 2], в соответствии с которыми спектр РС разделяется на 20 полос равной разборчивости с равными весовыми коэффициентами (5%). Некоторые статистические данные о полосах равной разборчивости приведены в табл. 1.

Таблица 1. Уровни речевого сигнала  $L_{si}$  в равноартикуляционных полосах

Номер полосы речевого сигнала	Частотные границы полосы $f_n \dots f_b$ , Гц	Типовые интегральные уровни речи $L_s$ , измеренные на расстоянии 1 м от источника сигнала, дБ		
		$L_s=64$ (тихая речь)	$L_s=70$ (речь со средним уровнем)	$L_s=76$ (громкая речь)
1.	100 ... 420	62	68	74
2.	420 ... 570	56	62	68
3.	570 ... 710	53	59	65
4.	710 ... 865	51	57	63
5.	865 ... 1030	49	55	61
6.	1030 ... 1220	48	54	60
7.	1220 ... 1410	46	52	58
8.	1410 ... 1600	44	50	56
9.	1600 ... 1780	43	49	55
10.	1780 ... 1960	42	48	54
11.	1960 ... 2140	41	47	53
12.	2140 ... 2320	40	46	52
13.	2320 ... 2550	40	46	52
14.	2550 ... 2900	41	47	53
15.	2900 ... 3300	40	46	52
16.	3300 ... 3660	39	45	51
17.	3660 ... 4050	38	44	50
18.	4050 ... 5010	41	47	53
19.	5010 ... 7250	42	48	54
20.	7250 ... 10000	41	47	53

В соответствии с методикой Покровского Н. Б. [1] в каждой полосе равной разборчивости определяется коэффициент восприятия формант, зависящий от соотношения сигнал/шум, и затем вычисляется интегральный индекс артикуляции речи.

Поэтому, вначале в процессе моделирования выполнялась фильтрация – разделение сигналов на 20 полос равной разборчивости. Характеристики некоторых из таких фильтров (420 - 570 Гц и 1960 - 2140 Гц) представлены на рис. 2, 3. Синтезированные фильтры имеют линейную ФЧХ в полосе пропускания и вносят ослабление свыше -50 дБ в полосе задерживания.

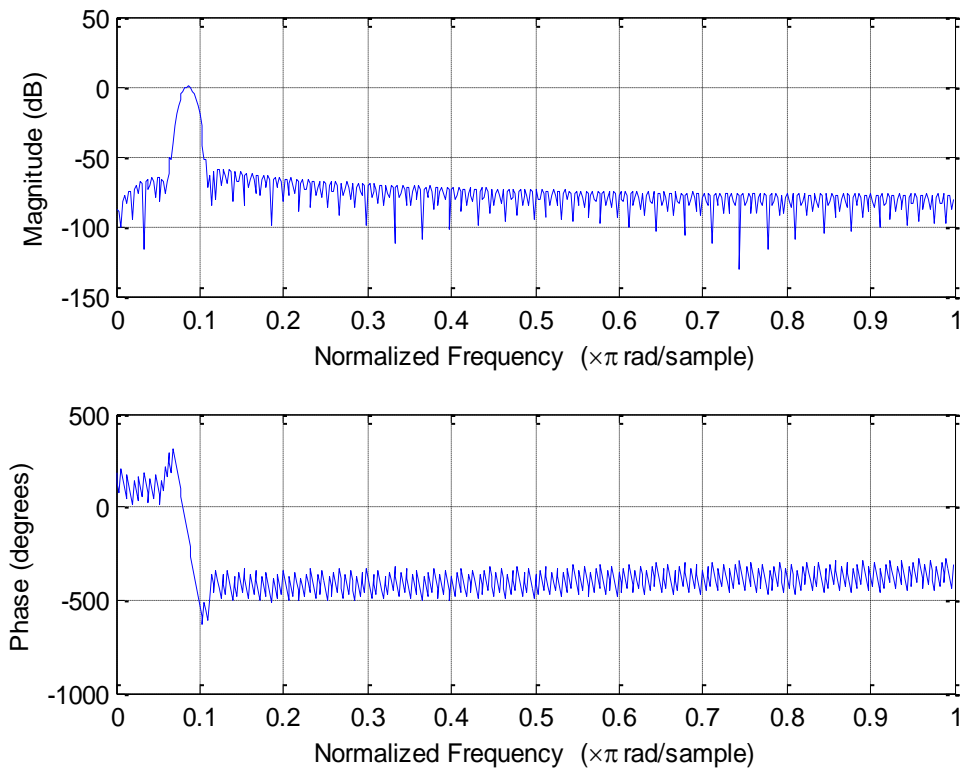


Рис. 2. АЧХ и ФЧХ фильтра 420 - 570 Гц (для 2-й полосы)

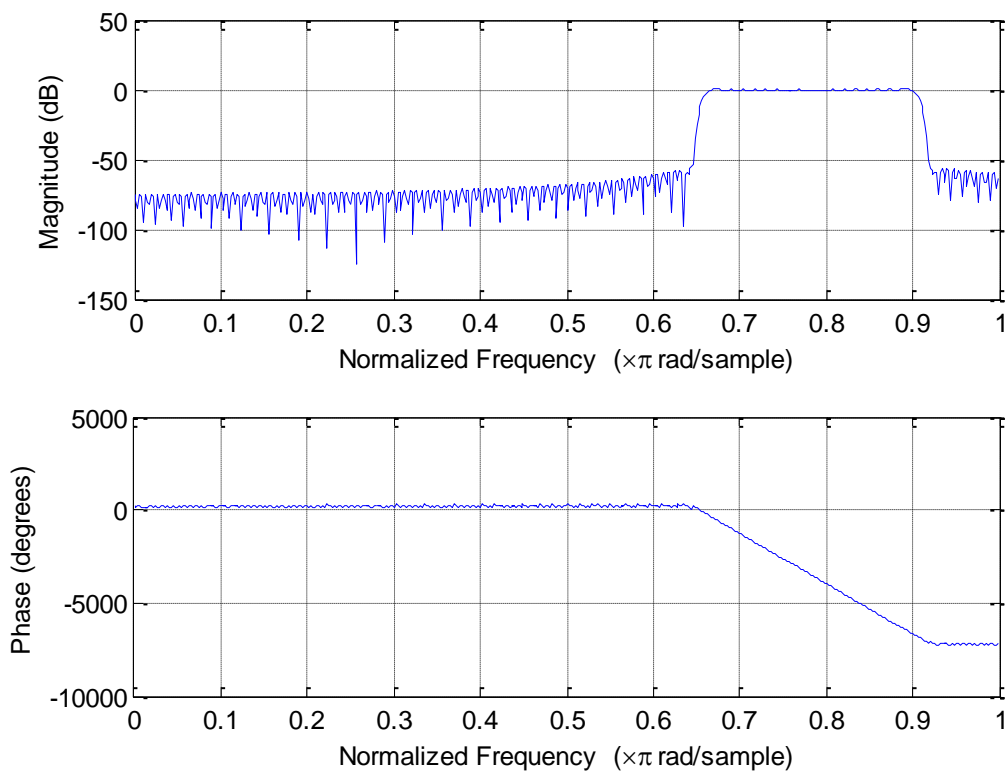


Рис. 3 . АЧХ и ФЧХ фильтра 1960 - 2140 Гц (для 11-й полосы)

Затем сигналы разбивались на временные интервалы длительностью по 30 мс и выполнялась операция сглаживания с помощью окна Хэмминга.

После этого с помощью быстрого преобразования Фурье сигналы трансформировались из временного в спектральный вид.

Далее в пределах каждой  $i$ -ой полосы равной разборчивости определялся коэффициент корреляции  $K_i$  смеси сигнал-шум с сигналом, а также

интегральный коэффициент корреляции  $K_{инт} = \frac{1}{20} \sum_{i=1}^{20} K_i$  при заданном

соотношении сигнал/шум ( $SN$ ). Затем проводилось усреднение  $K_i$  и суммы

$\sum_{i=1}^{20} K_i$  в течение временного интервала 8 сек. При изменении соотношения  $SN$  с

шагом 3 дБ для заданного РС были построены зависимости  $K_i$  и  $K_{инт}$  от соотношения уровней сигнал/шум ( $L_{SN}$ ). Схема алгоритма моделирования приведена на рис.4.

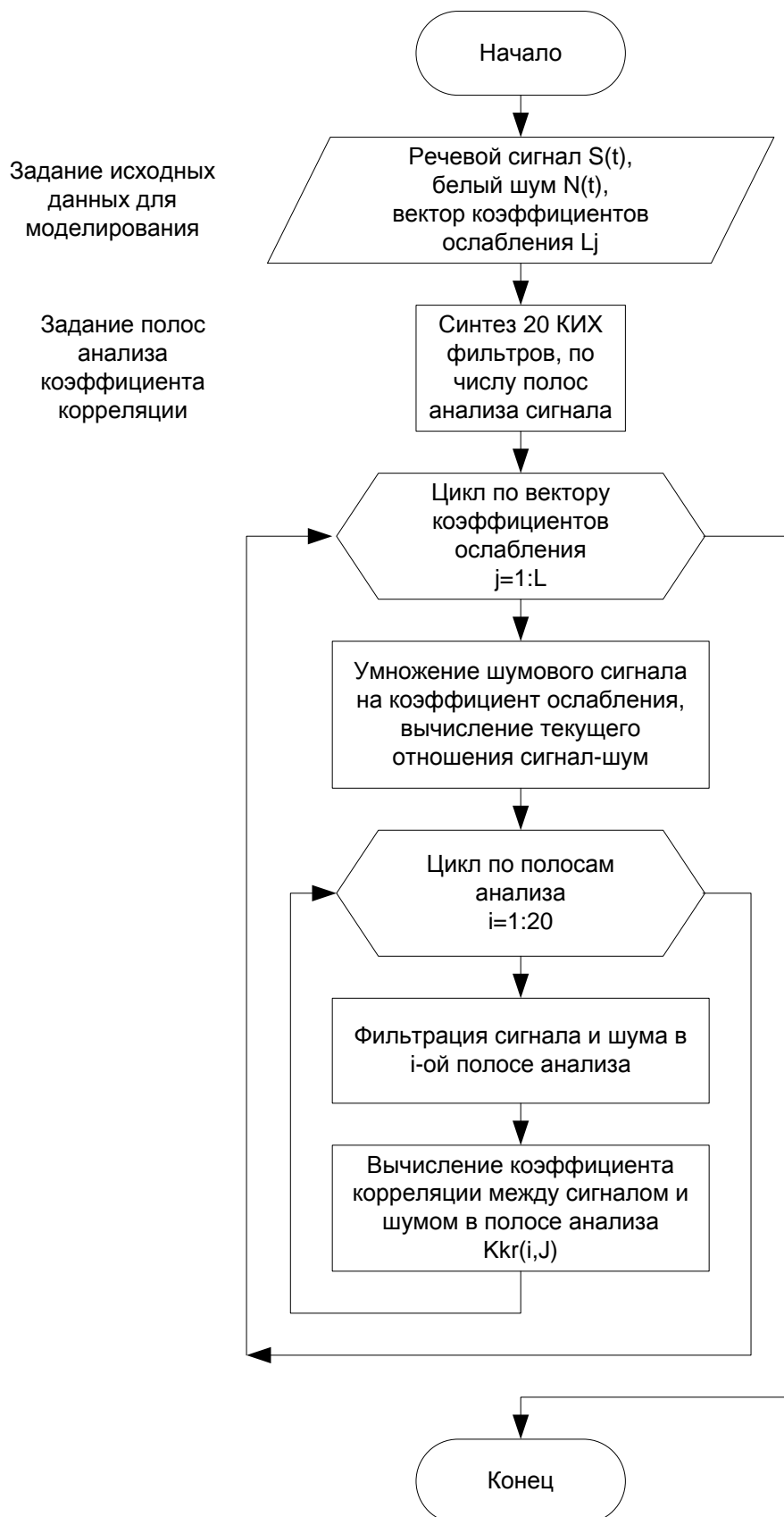


Рис. 4. Схема алгоритма проведения моделирования.

Результаты моделирования в программе MatLab показаны на рис. 5, 6.

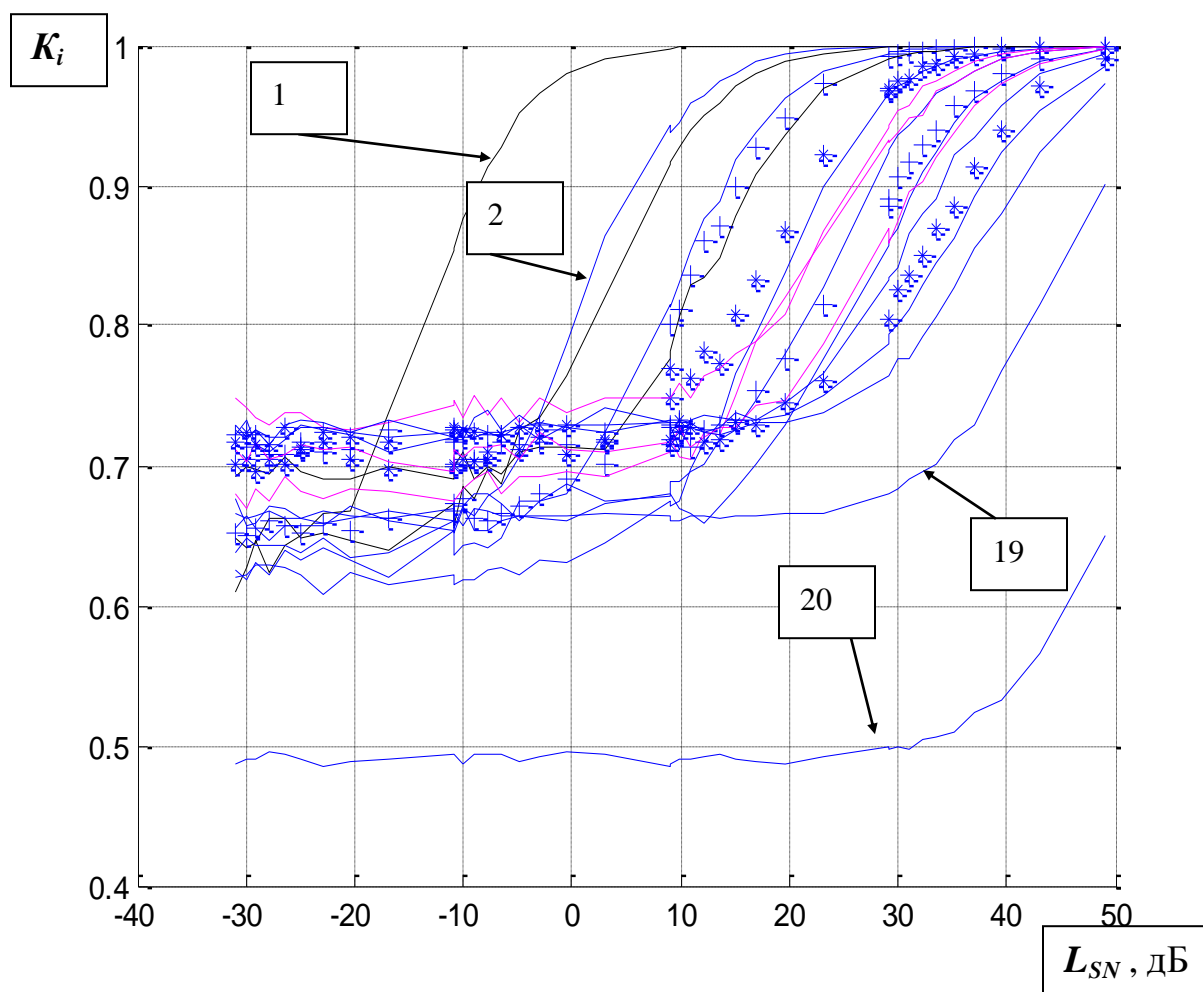


Рис. 5. Графики зависимости  $K_i$  от соотношения сигнал/шум ( $L_{SN}$ ).  
 1 – 1-я полоса 100 - 420 Гц, 2 – 2-я полоса 420 - 570 Гц, и т.д., 19 – 19-я  
 полоса 5010 – 7250 Гц, 20 – 20-я полоса 7250 – 10000 Гц (см. табл.1).

Анализ результатов, представленных на рис.5 позволяет сделать следующие выводы.

1. В полосах равной разборчивости минимальное значение коэффициента корреляции не составляет менее 0,5, т.е. коэффициент корреляции сигнала со смесью сигнал/шум никогда не достигает нулевых значений.

Это обусловлено тем, что при равных значениях мгновенных амплитуд сигнала и шума в смеси сигнал-шум, сигнал начинает коррелировать с шумом,

а не с сигналом. В случае дальнейшего снижения уровня сигнал-шум ( $L_{SN}$ ) сигнал будет коррелировать только с шумом, а не с речевым сигналом.

2. Следует обратить внимание, что на рис.5 насчитывается только 16 графиков, вместо 20 (для каждой полосы равной разборчивости).

Этот факт объясняется тем, что графики отдельных полос равной разборчивости соединились (слились) в одну линию, поскольку в отдельных полосах уровни речевого сигнала имеют одинаковые значения.

Некоторые из таких полос с равными значениями уровней сигнала в таблице выделены серым фоном. Например: 11-я и 14-я, 18-я и 20-я; а также 12-я, 13-я и 15-я. В данном эксперименте слияние графиков наблюдалось для 4-й и 5-й полос, а также других 3-х полос среднечастотного диапазона.

3. Анализ графика для 19-й полосы (5010–7250 Гц) показывает, что максимальное значение  $K_{19}$  не превышает 0,9 при  $L_{SN}$  более 30 дБ, а при значениях  $L_{SN}$  менее 25 дБ значение  $K_{19}=0,66$  и остается постоянным. Таким образом, корреляционным методом для 19-й полосы возможно определение соотношений сигнал/шум только в пределах от 20 до 50 дБ, т.е. чувствительность корреляционного метода ограничена указанным диапазоном – более 20 дБ. Аналогичные явления характерны и для других полос равной разборчивости. В целом чувствительность  $K_i$  для 16 полос из 20-ти ограничена  $L_{SN} = 0$ дБ, а чувствительность для 50% полос равной разборчивости ограничена ( $K_i$  не изменяется) при  $L_{SN}$  менее 15 дБ. В соответствии с положениями формантной теории [1] при дальнейшем снижении этого уровня теряется более 50% информации РС.

Это наблюдение согласуется с другими данными, согласно которым компьютерные системы автоматического распознавания речи (основанные на корреляционных методах) дают удовлетворительные результаты распознавания при  $L_{SN}$  свыше 15 дБ, в противном случае резко увеличивается количество ошибок.

4. Наиболее высокая чувствительность  $K_i$  отмечается в первых трех полосах (минимальное значение  $L_{SN} = -20$  дБ) и в 1-й полосе.



Эти данные хорошо согласуются с моделью речеобразования Фанта, поскольку в отмеченных полосах преобладают вокализованные форманты высокого уровня, полная маскировка которых (отсутствие признаков РС в смеси сигнал/шум) происходит при  $L_{SN}$  менее -20 дБ.

График интегрированного (в 20-ти полосах) коэффициента корреляции  $K_{инт}$  представлен на рис.6.

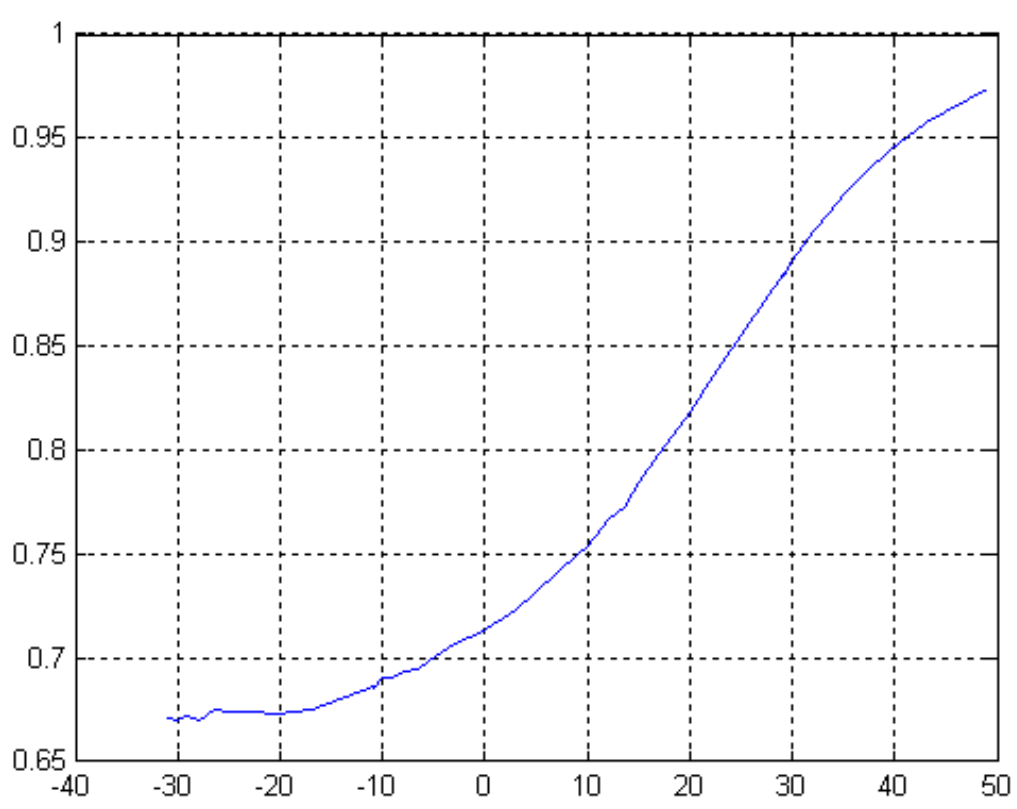


Рис. 6. График зависимости  $K_{инт}$  от отношения сигнал/шум ( $L_{SN}$ ).

Сравнивая зависимости  $K_i$  и  $K_{инт}$  от  $L_{SN}$  можно заметить существенную разницу в характере поведения графиков на рис. 5 и рис. 6. В частности обособленный анализ  $K_{инт}$  не позволяет определить границы применимости корреляционного метода для отрицательных значений  $L_{SN}$ .

Следовательно, интегрированный коэффициент корреляции  $K_{инт}$ , с учетом результатов 20-ти полосного анализа  $K_i$ , не позволяет с необходимой точностью определить отношение сигнал/шум при  $L_{SN}$  менее 0 дБ для смеси речевого сигнала с белым шумом для полос равной разборчивости с 7-й по 20-ю в диапазоне от 1220 до 10000 Гц.

Таким образом, можно сделать обобщающий вывод, что корреляционный метод оценки эффективности маскирования речи белым шумом не позволяет определить степень защищенности РС в октавных полосах со среднегеометрическими частотами свыше 2000 Гц.

### **Литература**

1. Железняк В. К., Макаров Ю. К., Хорев А. А., «Некоторые методические подходы к оценке эффективности защиты речевой информации». Специальная техника № 6. М-2000.
2. Хорев А. А., Макаров Ю. К., «К оценке эффективности защиты акустической (речевой) информации». Специальная техника № 5. М-2000.
3. Халяпин Д.Б. «Вас подслушивают? Защищайтесь!» – М.: «Мир безопасности», 2001.
4. Кученков Е. Б., «Подходы к разработке методики оценки защищенности ВТСС». Специальная техника №1. М-2002.
5. Метод экспериментального анализа функции эффективности маскирования речи. Журавлёв В.М., Архипова Е.А. Национальный технический университет Украины "КПИ". УДК 681.391. Вестник Винницкого политехнического института, 2009, № 1.
6. Дворянкин С.В. Речевая подпись. М.: РИО МГУСИ. 2003. - 184 с.