

О некоторых особенностях развития объективных методов измерений разборчивости речи

Сопоставлены три объективных метода оценки разборчивости речи: метод парциальных отношений сигнал-шум, формантный метод и модуляционный метод. Показана возможность объединения достоинств формантного и модуляционного методов.

Three objective methods of estimation of speech intelligibility had been confronted: a method of partial signal-noise relations, formant method and modulation method. The possibility of advantages join of formant and modulation methods is proved.

Введение

Известен ряд показателей и методов оценки разборчивости речи, применяемых для оценки качества акустики помещений, линий связи, а также защищенности речевой информации от утечки по акустическим каналам [1]. Одни методы изначально «приспособлены» к условиям шумовой помехи – к таковым можно отнести метод парциальных отношений сигнал-шум и формантный метод. Другие методы и показатели ориентированы на условия преобладающего влияния реверберационной помехи – среди таких можно назвать метод %Alcons и коэффициенты четкости C50 и C80 [2-4]. Значительный интерес вызывают «универсальные» методы, позволяющие одновременно учитывать влияние как шумовых, так и реверберационных, помех – это группа модуляционных методов STI, RASTI, STITEL и STIPA [5, 6].

В данной работе сопоставляются три объективных метода оценки разборчивости речи: метод парциальных отношений сигнал-шум, формантный метод и модуляционный метод. Цель такого рассмотрения – проследить, как и почему, с развитием объективных методов, изменились подходы к выбору показателя разборчивости. Результаты такого анализа могут быть весьма полезными, помогая четче увидеть сильные и слабые стороны методов и, возможно, спрогнозировать дальнейшее их развитие.

1. Метод парциальных отношений сигнал-шум

Метод парциальных отношений сигнал-шум, в силу своей сравнительной простоты, весьма популярен у специалистов в области защиты информации от утечки по акустическим каналам [2].

Сущность метода состоит в измерении отношений сигнал-шум (в точках возможного прослушивания) в смежных полосах частот, с последующим принятием решения о защищенности помещения.

Если требуется обеспечить полную неразборчивости речи в точке наблюдения, вплоть до исчезновения признаков речи на фоне шума, легко показать, что данный метод действительно позволяет решить задачу кратчайшим путем. Так, из представленных на рис. 1 графиков коэффициента восприятия $P(E)$ [1] для различных октавных полос частот, где E – отношение сигнал-шум, следует, например, что для полосы с центральной частотой $f_0 = 500 \text{ Гц}$ вероятность восприятия формант равна нулю при $E_{\min} \approx -14 \text{ дБ}$, а для полосы с $f_0 = 1000 \text{ Гц}$ – при $E_{\min} \approx -17 \text{ дБ}$. Аналогично находят минимальные значения отношения сигнал-шум E_{\min} для остальных полос частот. Набор значений E_{\min} и есть тот набор показателей, которым должно удовлетворять защищенное помещение.

Сложнее обстоит дело, если допустимым является меньший уровень защиты, при котором разрешается, например, словесная разборчивость 5% или 10%. Можно показать [2], что в этом случае одной и той же словесной разборчивости соответствует несчетное множество различных комбинаций парциальных отношений сигнал-шум.

Ослабить напряженность ситуации можно, переходя от непрерывной шкалы отношений сигнал-шум к дискретной шкале – очевидно, с прикладных позиций дискретность с шагом 1 дБ может считаться приемлемой. Однако даже в этом случае не исчезает проблема полноты и компактности описания множества предельно допустимых комбинаций отношений сигнал-шум [2]. Если множество предельно допустимых отношений сигнал-шум описывать таблично, такой способ, в силу своей громоздкости, вряд ли будет пригоден для практических приложений. Если прибегнуть к аналитическому описанию данного множества, с помощью аппроксимации его элементов по методу множественной регрессии, тогда описание множества становится более приемлемым для практического использования. Хотя и этот способ неудобен в прило-

жениях, поскольку показателем качества служит не одна величина, а совокупность пяти-шести парциальных отношений сигнал-шум величин для пяти-шести октавных полос частот. Если же возникает необходимость в проведении акустической экспертизы с использованием 1/3-октавных фильтров, из-за существенного возрастания количества анализируемых полос задача становится практически неразрешимой.

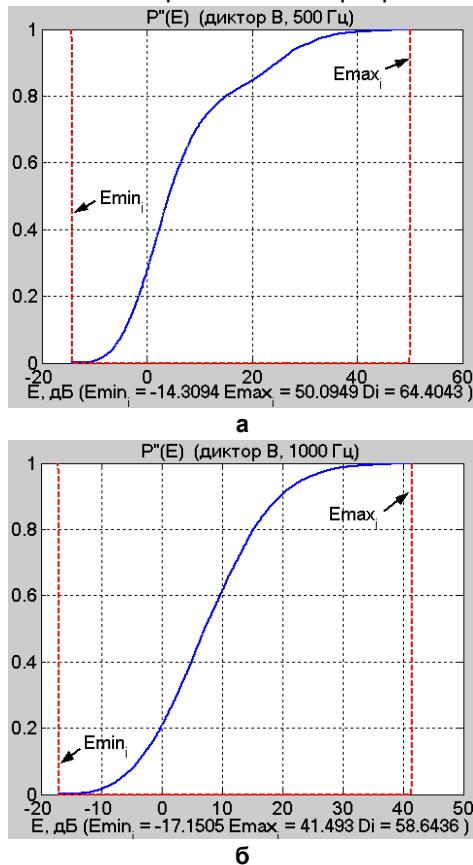


Рис. 1. Коэффициенты восприятия для $f_0 = 500$ Гц (а) и $f_0 = 1000$ Гц (б)

2. Формантный метод

Формантный метод выгодно отличается от метода парциальных отношений сигнал-шум тем, что в роли показателя качества используется скалярная величина - формантная разборчивость, представляющая собой среднюю вероятность восприятия формант на фоне шума:

$$A = \sum_{k=1}^K p_k \cdot w_k(E_k), \quad (1)$$

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

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

$F_1(f)$ - функция распределения вероятностей формант по частоте; $w_k(E_k)$ - коэффициент

восприятия (вероятность восприятия формант на фоне шума в k -той полосе частот); E_k - уровень восприятия (отношение сигнал-шум) в k -той частотной полосе.

Сопоставляя формантный метод с методом парциальных отношений сигнал-шум, нетрудно видеть известную аналогию с задачей синтеза оптимального обнаружителя сигналов на фоне помех [7], где вместо распределения вероятностей

$$P(a_i, b_j), \quad i, j = 0, 1 \quad (3)$$

исчерпывающим образом характеризующего качество решения задачи обнаружения сигнала на фоне помех (a_0 - сигнала нет; a_1 - сигнал есть; b_0 - решение об отсутствии сигнала; b_1 - решение о наличии сигнала) используют единственный показатель, именуемый «средним риском»:

$$R_c = \sum_{i=0}^1 \sum_{j=0}^1 R_{ij} P(a_i, b_j), \quad (4)$$

где R_{ij} - величина (риска), характеризующая ущерб от неправильного решения [7].

Еще одним примером - задача оптимального кодирования сообщения, представленного элементами s_i [8]. Ее решают, используя показатель в виде средней длины кода:

$$H = - \sum_i p(s_i) \cdot \log_2 p(s_i), \quad (5)$$

где $p(s_i)$ - вероятность появления элемента s_i .

Сопоставляя соотношения (1), (4) и (5), нетрудно увидеть принципиальное сходство подходов к эффективному решению различных задач, а именно: показателем качества служит среднее значение некоей скалярной величины.

2.1. Вероятность как показатель качества

Важной особенностью формантного метода является то, что показателем качества является вероятность. Используя информацию о статистических свойствах речи, среднюю вероятность восприятия формант A затем можно пересчитать в вероятность правильного распознавания звуков, слогов, слов и даже фраз.

Это чрезвычайно удобно, поскольку при использовании субъективных методов оценки разборчивости речи показателем качества также служит вероятность правильного распознавания звукосочетаний, слогов, слов или фраз. Следовательно, для сопоставления качества решения одной и той же задачи объективным и субъективным методами достаточно сравнить две оценки вероятностей.

2.2. Проблемы формантного метода

К концу 50-х годов прошлого века в СССР сформировались три «школы» в теории разборчивости речи, представители которых – Ю. С. Быков, Н. Б. Покровский и М. А. Сапожков – отстаивали авторские версии формантного метода [9-11].

Различие версий состояло, во-первых, в различии трактовок понятия «формантный спектр». Во-вторых, отличалась форма коэффициентов восприятия. Наконец, не совпадали мнения авторов версий и по поводу формы функции распределения вероятностей формант по частоте $F_1(f)$.

Ю.К. Калинцев [12] предпринял попытку со-поставления коэффициентов восприятия разных авторов на одном графике. При этом он совершил ряд ошибок, которые были исправлены в работах [13-15]. На рис. 2 показаны результаты корректного сопоставления коэффициентов восприятия Н.Б. Покровского $P(E'_k)$, М.А. Сапожкова $w(E)$, Ю.С. Быкова $P_i(\Delta\lambda_i)$, а также их американского аналога $P_{AI}(\Delta L)$ – коэффициента восприятия американской версии формантного метода, именуемой индексом артикуляции (Articulation Index – AI).

Сопоставление коэффициентов восприятия на одном графике свидетельствует о большой проблеме советских версий формантного метода, а именно: неоднозначность определения коэффициента восприятия для русской речи.

Анализ причин неоднозначности свидетельствует, что виной тому различие методик формирования коэффициента восприятия. В работах [1, 13-15] показано, что наиболее корректна методика М.А. Сапожкова, согласно которой коэффициент восприятия $w(x)$ определяется соотношением:

$$w(x) = 1 - F(-x), \quad (6)$$

где $F(x)$ – функция распределения уровней речевого сигнала. В работе [1] представлены результаты оценивания коэффициентов восприятия русской речи с использованием соотношения (6), а именно: семейство новых коэффициентов восприятия для пяти октавных полос частот, с центральными частотами 250, 500, 1000, 2000, 4000 Гц. Некоторые коэффициенты этого семейства показаны на рис. 3, где также для сравнения приведены коэффициенты восприятия М.А. Сапожкова и Н.Б. Покровского.

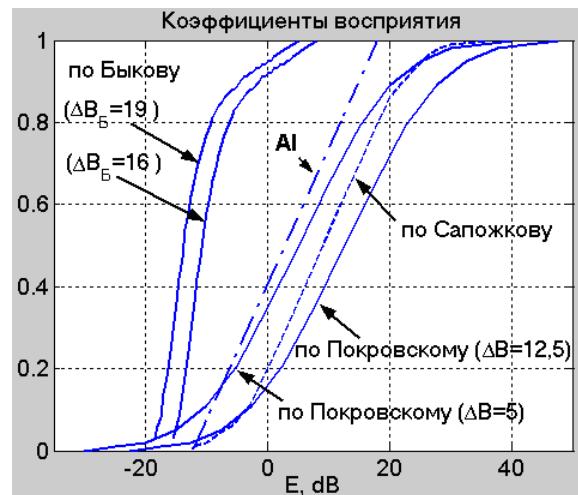


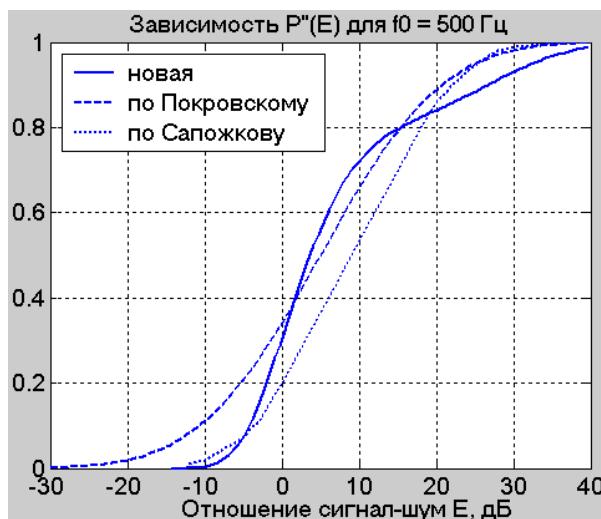
Рис. 2. Коэффициенты восприятия $P_{AI}(\Delta L)$, $w(E)$, $P_i(\Delta\lambda_i)$ и $P(E'_k)$ [15]

Интересным представляется тот факт, что для полосы частот с центральной частотой 1000 Гц новый коэффициент восприятия практически совпал с коэффициентом восприятия М.А. Сапожкова (рис. 3,б). На наш взгляд, такое совпадение можно трактовать как признак достоверности новых коэффициентов восприятия. В остальных полосах частот (рис. 3,а) новые коэффициенты восприятия, отличающиеся от коэффициента восприятия М.А. Сапожкова, следует рассматривать как уточнение результатов, полученных М.А. Сапожковым.

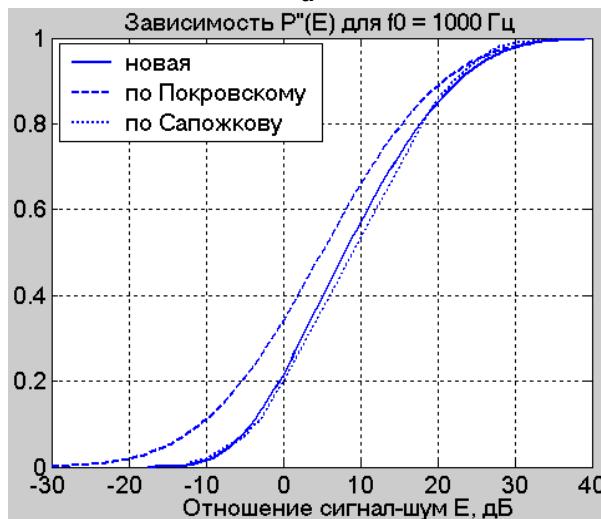
Заметим, что современные российские исследователи [16-18] используют коэффициент восприятия, предложенный Н.Б. Покровским, что, на наш взгляд, неприемлемо, поскольку методика формирования коэффициента восприятия Н.Б. Покровского содержит ряд недостаточно обоснованных предположений и даже ошибок [1, 13-15, 19].

2.3. Коэффициенты восприятия для украинской речи

Отсутствие в литературе сведений о формантных свойствах украинской речи означало, в сущности, невозможность получения корректных оценок разборчивости украинской речи. Этот пробел был восполнен в работах [19, 20], где показано, что функции распределения формант по уровню для украинской и русской речи при больших уровнях речевого сигнала (выше среднего уровня речи) практически одинаковы. При малых уровнях речевого сигнала функции распределения украинской речи почти везде превышают таковые для русской речи (рис. 4).



а



б

Рис. 3. Сопоставление коэффициентов восприятия

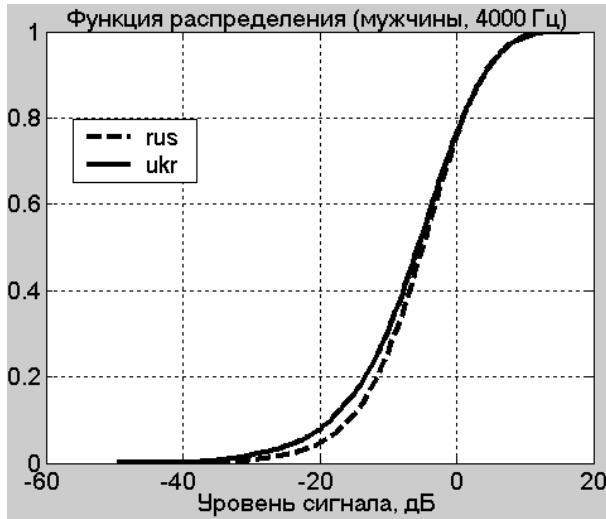
Данный факт, с учетом соотношения (6), позволяет прийти к выводу, что при малых отношениях сигнал-шум разборчивость русской и украинской речи практически одинакова, а при больших отношениях сигнал-шум разборчивость украинской речи может быть несколько ниже.

2.4. Кусочно-линейная аппроксимация коэффициента восприятия

М.А. Сапожков отмечал практически линейный характер коэффициента восприятия в пределах некоторого интервала значений отношения сигнал-шум [10]. Идея линеаризации коэффициента восприятия доведена до логического завершения в современной американской версии формантного метода, именуемой «индексом артикуляции» (AI – articulation index), где вместо «гладкого» коэффициента восприятия используют его кусочно-линейную аппроксимацию [1, 16]. Закономерен вопрос: какой положительный эффект достигается путем такой аппроксимации?



а



б

Рис. 4. Коэффициенты восприятия украинской и русской речи

По нашему мнению, положительный эффект состоит в следующем:

- удается избежать неоднозначности результатов, вызванной различием методик оценивания коэффициента восприятия (рис. 2);
- удается избежать погрешностей результатов, обусловленных различием условий измерения функции $F(x)$.

Отрицательный эффект состоит в том, что теперь мерой качества является не вероятность, а некий «индекс». Как следствие, становится труднее сопоставлять результаты объективных и субъективных оценок. Кроме того, из-за изломов кусочно-линейной характеристики могут возникнуть проблемы с трактовкой результатов измерений при малых отношениях сигнал-шум, типичных для задачи защиты акустических каналов от утечки информации.

2.5. Учет влияния реверберационной помехи

При оценивании разборчивости речи в соответствии с соотношением (1), предполагается, что речь воспринимается на фоне шумовой помехи. Однако, в принципе, влияние реверберационной помехи также несложно учесть. Действительно, поскольку реверберационная и шумовая помехи статистически независимы, общая мощность помех равна сумме их мощностей, и результирующее отношение сигнал-шум равно:

$$SNR = SNR_N + SNR_R - 10 \lg(10^{0,1 \cdot SNR_N} + 10^{0,1 \cdot SNR_R}), \quad (7)$$

где SNR_N и SNR_R – отношения сигнал-шум для шумовой и реверберационной помех, соответственно.

Таким образом, при совместном влиянии шумовой и реверберационной помехи в соотношении (1) в качестве E_k следует использовать отношение сигнал-шум, определяемое выражением (7).

Значение SNR_R можно вычислить по известному времени реверберации T_{60} :

$$SNR_R = 10 \lg(e^{\frac{0,69}{T_{60}}} - 1). \quad (8)$$

Соотношение (8) получено в предположении $SNR_R = C_{50}$, где C_{50} – индекс четкости, характеризующий отношение энергий прямого и отраженного звука [18]:

$$C_{50} = 10 \lg \frac{\int_0^{0,05} p^2(t) dt}{\int_{\infty}^{0,05} p^2(t) dt}$$

где $p^2(t) = p^2 \cdot e^{-\alpha t}$ – квадрат импульсной характеристики помещения; $\alpha = \frac{13,8}{T_{60}}$; $0,05$ – момент времени, выраженный в секундах.

К сожалению, весьма спорным является выбор значения $t = 0,05$ с в качестве границы между участками импульсной характеристики помещения, характеризующими прямой и рассеянный звук. В работе [4], например, указывается несколько иное значение $t = 0,095$ с, т.е. почти вдвое выше.

В этом случае $SNR_R = 10 \lg(e^{\frac{1,31}{T_{60}}} - 1)$. Например, при $T_{60} = 1$ с в первом случае получаем $SNR_R \approx 0$ дБ, а во втором $SNR_R \approx 4,3$ дБ. Как видим, разница весьма ощутимая.

3. Модуляционный метод

Модуляционный метод [5, 6] базируется на идее использования тестового шумового сигнала, мощность которого изменяется по гармоническому закону. Показателем качества при этом служит глубина модуляции мощности сигнала в точке наблюдения. Эта идея оказалась весьма плодотворной, поскольку позволила с единых позиций учитывать влияние как шумовой, так и реверберационной помех.

В случае шумовой помехи, зависимость глубины модуляции m_N от отношения сигнал-шум SNR , выраженного в децибелах, имеет вид:

$$m_N = (1 + 10^{-0,1 \cdot SNR})^{-1}. \quad (9)$$

В случае реверберационной помехи глубина модуляции зависит от частоты модуляции F и определяется соотношением:

$$m_R(F) = \frac{1}{\sqrt{1 + \left(\frac{2\pi FT}{13,8}\right)^2}}, \quad (10)$$

где T – время реверберации.

Объединяя соотношения (9) и (10), получают окончательное выражение для глубины модуляции («функции передачи модуляции»):

$$m(F) = \frac{1}{\sqrt{1 + \left(\frac{2\pi FT}{13,8}\right)^2}} \cdot \frac{1}{1 + 10^{-0,1 \cdot SNR}} = m_R(F) \cdot m_N(F). \quad (11)$$

Соотношение (11) является «теоретической основой» модуляционного метода. При практических измерениях множество значений $m(F)$ усредняют по частотам F и по полосам частот k . В конечном счете, после такого усреднения, множество значений $m(F)$ преобразуется в единственный числовой показатель, именуемый индексом передачи речи STI (Speech Transmission Index).

Согласно «полной» (не ревизованной) версии модуляционного метода, преобразование множества значений $m(F)$ в индекс STI выполняют поэтапно.

Этап 1. Оценивают 98 значений коэффициента $m_k(F_i)$, получаемых для 14 значений частот модуляции F_i из диапазона значений 0,63...12,5 Гц в каждой из семи октавных полос с центральными частотами от 125 Гц до 8 кГц (k – номер полосы частот).

Этап 2. Каждую из 98 оценок коэффициента $m_k(F_i)$ «конвертируют» в эффективное отношение сигнал-шум:

$$(S/N)_{\text{эф}k,i} = 10 \log \frac{m_k(F_i)}{1 - m_k(F_i)}.$$

Если $(S/N)_{dB\text{эф}k,i} < -15 \text{ дБ}$, принимают

$$(S/N)_{dB\text{эф}k,i} = -15 \text{ дБ},$$

а если $(S/N)_{dB\text{эф}k,i} > +15 \text{ дБ}$, принимают

$$(S/N)_{dB\text{эф}k,i} = +15 \text{ дБ}.$$

Этап 3. Производят усреднение оценок эффективных отношений сигнал-шум для каждой из октавных полос:

$$(S/N)_{\text{эф}k} = \frac{1}{14} \sum_{i=1}^{14} (S/N)_{\text{эф}k,i}$$

Этап 4. Вычисляют взвешенное среднее:

$$(\overline{S/N})_{\text{эф}} = \frac{1}{14} \sum_{k=1}^7 (S/N)_{\text{эф}k} \cdot w_k,$$

где w_k - весовые коэффициенты ($\sum_k w_k = 1$).

Этап 5. Вычисляют индекс STI, используя соотношение (6) либо рис. 5.

$$STI = \frac{(\overline{S/N})_{\text{эф}} + 15}{30}$$

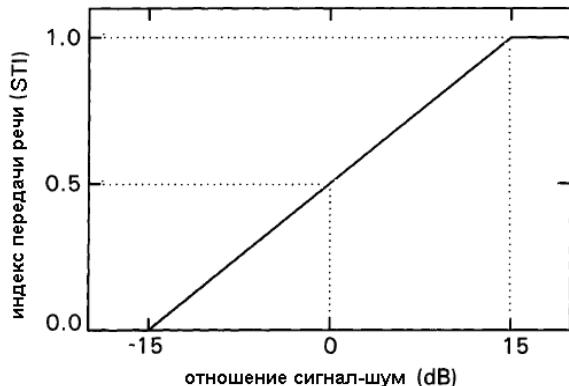


Рис. 5. Зависимость между STI и эффективным отношением сигнал-шум [17]

Нетрудно заметить существенное сходство кусочно-линейных характеристик, используемых в методах AI и STI (см. рис. 2 и 5).

3.1. Учет влияния реверберационной помехи

Очевидным достоинством модуляционного метода, по сравнению с формантным, является «легкость» учета влияния реверберационной помехи. Экспериментальные исследования свидетельствуют, что, по точности экспертизы помещений с преобладанием реверберационной помехи, модуляционный метод превосходит как коэффициент четкости C50, так и метод %Alcons [19].

3.2. О кусочно-линейной аппроксимации зависимости $m = m(SNR)$

Первый шаг к такой аппроксимации совершен на **Этапе 2**, когда при $(S/N)_{dB\text{эф}k,i} < -15 \text{ дБ}$ принималось

$$(S/N)_{dB\text{эф}k,i} = -15 \text{ дБ},$$

а при $(S/N)_{dB\text{эф}k,i} > +15 \text{ дБ}$ принималось

$$(S/N)_{dB\text{эф}k,i} = +15 \text{ дБ}.$$

В результате область возможных значений $(S/N)_{dB\text{эф}k,i}$ оказывалась ограниченной интервалом $[-15 \text{ дБ}; 15 \text{ дБ}]$.

Второй и завершающий шаг совершен на **Этапе 5**, где вместо плавной кривой зависимости $m = m(SNR)$, описываемой соотношением (9) (штриховая линия на рис. 6), предложено использовать линейную зависимость (сплошная линия на рис. 6).

Как видим, авторы метода STI избегают вычислений по формуле (9), что, при современном уровне вычислительной техники, вызывает недоумение: сегодня такие вычисления доступны любому калькулятору.

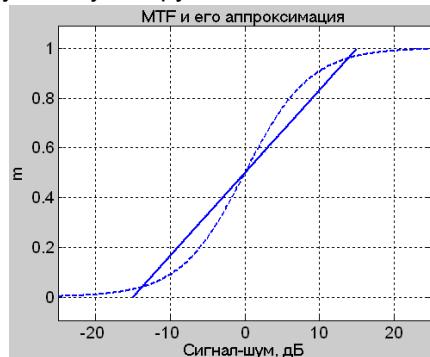


Рис. 6. Зависимость между m и SNR

4. О возможности слияния формантного и модуляционного методов

Учитывая, что сильной стороной формантного метода является вычисление вероятности восприятия формант, а сильной стороной модуляционного метода является эффективный учет влияния не только шумовой, но и реверберационной помехи, можно создать новый метод, объединяющий эти достоинства.

В результате приходим к следующему порядку оценивания формантной разборчивости.

Шаг 1. Используя модуляционный метод, оценивают эффективные отношения сигнал-шум $(S/N)_{\text{эф}k}$, не производя сужение области зна-

чений $(S/N)_{dB\text{эфф},i}$ до интервала [-15 дБ; 15 дБ] на Этапе 2.

Шаг 2. Формантную разборчивость оценивают в соответствии с соотношением (1), полагая $E_k = (S/N)_{\text{эфф},k}$.

Выводы

Весьма популярный, благодаря своей простоте, метод парциальных отношений сигнал-шум весьма неудобен в ситуации, когда допускается некоторая, пусть и весьма малая, разборчивость. В этом случае одному и тому же значению разборчивости соответствует практически необозримое множество комбинаций предельно допустимых парциальных отношений сигнал-шум.

Значительно более удобными для практического использования являются формантный и модуляционный методы, где показателем качества служит единственное число. Оба эти метода - и формантный, и модуляционный, - обладают сильными и слабыми сторонами. Сильной стороной формантного метода является использование вероятности в качестве меры разборчивости, а сильной стороной модуляционного метода является легкость учета влияния reverberационной помехи.

Поэтому представляется целесообразным создание нового метода оценивания разборчивости речи, объединяющего достоинства формантного и модуляционного методов.

Литература

1. Дидковский В.С., Дидковская М.В., Продеус А.Н. Акустическая экспертиза каналов речевой коммуникации. Монография. – К: Имэкс-ЛТД, 2008. – 420 с.
2. Гавриленко А.В., Дидковский В.С., Продеус А.Н. Парциальные отношения сигнал-шум в роли критерия, альтернативного разборчивости речи. // Электроника и связь, №4, 2006. С.48-53.
3. Дідковський В.С., Луньова С.А. Основи архітектурної та фізіологічної акустики. – К.: 2001. – 422 с.
4. Jacob K.D. Correlation of Speech Intelligibility Tests in Reverberant Rooms with Tree Predictive Algorithmes. // J. Audio Eng. Soc. - 1989. - Vol. 37. – P. 1019-1030.
5. Houtgast T. and Steeneken H. J. M. Evaluation of Speech Transmission Channels by using Artificial Signals. Acustica. - 1971. -Vol. 25. - P. 355–367.
6. Steeneken H. J. M., and Houtgast T. A physical method for measuring speech-transmission quality. // J. Acoust. Soc. Am. - 1980. - Vol. 67. – P. 318-326.
7. Гуткин Л.С. Теория оптимальных методов радиоприема при флуктуационных помехах. – М.: Сов.радио. – 1972. – 448 с.
8. Ватолин Д., Ратушняк А., Смирнов М., Юткин В. Методы сжатия данных. – М.: Диалог-МИФИ. – 2002. – 381 с.
9. Покровский Н.Б. Расчет и измерение разборчивости речи. – М.: Связьиздат, 1962. – 390 с.
10. Сапожков М.А. Речевой сигнал в кибернетике и связи. – М.: Связьиздат, 1963. – 472 с.
11. Быков Ю. С. Теория разборчивости речи и повышение эффективности радиотелефонной связи. – М.-Л.: Госэнергоиздат, 1959. – 350 с.
12. Калинцев Ю. К. Разборчивость речи в цифровых вокодерах. – М.: Радио и связь, 1991. – 219 с.
13. Гавриленко А.В., Дидковский В.С., Продеус А.Н. Сравнительный анализ некоторых методов оценки разборчивости речи. // Сб.трудов Акустического симпозиума "Консонанс-2007", К., 2007. С.273-278.
14. Гавриленко А.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.227-231.
15. 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.
16. Рашевский Я.И., Каргашин В.Л. Обзор зарубежных методов определения разборчивости речи. // Специальная техника, №№ 3 - 6 за 2002 год, № 1 за 2003 год (http://www.detektor.ru/publication/d2003_1.htm).
17. 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.
18. Макаров Ю.К., Хорев А.А. К оценке эффективности защиты акустической (речевой) информации // Специальная техника, 2000, № 5, с. 46 – 56.
19. Продеус А.Н. Об измерениях вероятностных характеристик формант украинской и русской речи. // Сб.трудов Акустического симпозиума "Консонанс-2009", К., 2009. С.285-291.
20. Дидковский В.С., Продеус А.Н. Сопоставление формантных свойств украинской и русской речи. - // Электроника и связь, тематический выпуск "Электроника и нанотехнологии", ч.2, 2009. №4-5. - С.88-94.