О НАДЕЖНОСТИ НЕКОТОРЫХ ПРОСТЫХ ДЛЯ ВЫЧИСЛЕНИЙ МЕР КАЧЕСТВА РЕЧЕВЫХ И МУЗЫКАЛЬНЫХ СИГНАЛОВ

А.Н. ПРОДЕУС, И.В. КОТВИЦКИЙ

Национальный технический университет Украины «КПИ», Киев

Исследованы особенности простых, с вычислительной точки зрения, мер качества речевых и музыкальных сигналов виде сегментного отношения сигнал-шум и лог-спектральных искажений. Уточнены предложения по повышению достоверности оценок указанных мер.

ВВЕДЕНИЕ

Задача оценки качества речевых и музыкальных сигналов в коммуникационных каналах является актуальной. Принципиальным преимуществом объективного (инструментального) подхода к оцениванию качества акустических сигналов является его быстрота и относительная дешевизна [1-4]. Наиболее достоверные перцептуальные меры качества, основанные на моделях слуховой системы человека, весьма сложны, поэтому понятно желание использовать более простые для вычислений меры качества. При этом, однако, с особой остротой возникает вопрос о достоверности используемых мер качества.

Наиболее привлекательными в плане простоты вычислений являются такие меры как сегментное отношение сигнал-шум (Segmental Signal to Noise Ratio – SSNR) и лог-спектральные искажения (Log-Spectral Distortion – LSD). Практическое использование этих мер показало, что они обладают рядом особенностей, которые, к сожалению, недостаточно освещены в литературе. Этот пробел частично восполнен в работах [4-7], а в данной работе предлагается как уточнение прежних, так и ряд новых результатов.

1. ОСОБЕННОСТИ МЕРЫ SSNR

Одной из простейших, в вычислительном плане, мер качества речевых сигналов является сегментное отношение сигнал-шум

$$SSNR = \frac{1}{L}\sum_{l=1}^{L} 10\lg\left[\frac{\sum_{n=Rl}^{Rl+N-1} x^2(l,n)}{\sum_{n=Rl}^{Rl+N-1}[x(l,n)-y(l,n)]^2}\right], \qquad (1)$$

где $x(l,n)$ и $y(l,n)$ – n-я выборка l-го фрейма эталонного и искаженного сигналов $x(n)$ и $y(n)$, соответственно.

Существенная зависимость результатов вычислений SSNR от ошибки выравнивания во времени сравниваемых сигналов отмечена в [3], хотя количественная оценка степени такого влияния не приведена. В [5] этот пробел восполнен, и для простейшей модели гармонического сигнала такая количественная оценка получена. Действительно, полагая

$$x(t) = \cos 2\pi f_0 t\,, \quad y(t) = \cos 2\pi f_0 (t-\tau)\,, \qquad (2)$$

приходим к простому соотношению

$$SSNR = SNR = 10\lg\int_0^T x^2(t)\,dt \Big/ \int_0^T [x(t)-y(t)]^2\,dt = 10\lg\frac{0{,}5}{1-\cos 2\pi f_0\tau}, \qquad (3)$$

где f_0 – частота сигнала, τ – ошибка выравнивания во времени, T – интервал наблюдения. Конкретные числовые значения $SSNR(\tau)$ для различных соотношений F_s/f_0, где F_s – частота дискретизации, приведены в табл. 1.

Таблица 1. Значения $SSNR(\tau)$

F_s/f_0 \ τF_s	**1**	**0,5**
22,05	10,93 дБ	16,93 дБ
2	-6,02 дБ	-3,01 дБ

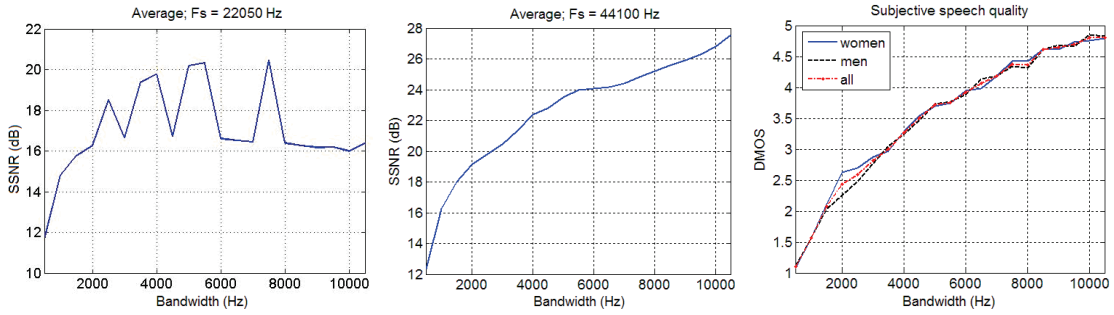
Поскольку на практике значения SSNR обычно ограничивают сверху величиной 35 дБ, заключаем, что ошибка выравнивания гармонических сигналов всего в один период дискретизации (τF_s=1) приводит к ошибке оценивания SSNR, близкой 24 дБ для сочетания значений f_0=1 кГц и F_s=22050 Гц. Для $f_0 \approx 11$ кГц, при той же частоте дискретизации, ошибка достигает 41 дБ. Столь быстрый рост легко объяснить, если учесть, что при $f_0 \approx 11$ кГц и $F_s = 22050$ Гц на одном периоде гармонического сигнала оказывается всего две выборки этого сигнала. Поэтому простой и достаточно очевидный способ борьбы с отмеченным явлением состоит в интерполяции сигналов $x(n)$ и $y(n)$, направленной на улучшение «прорисовки» высокочастотных компонентов [5].

В табл. 1 также приведены значения SSNR для ошибки выравнивания, равной половине расстояния между отдельными выборками (τF_s = 0,5). Такая ошибка случается, если $y(n)$ является результатом фильтрации сигнала $x(n)$ нерекурсивным фильтром четного порядка. Отсюда следует вывод о необходимости контролировать порядок нерекурсивных фильтров, который должен быть нечетным [5].

Эффективность данных рекомендаций вначале продемонстрируем на примере оценки качества речевых сигналов, ограниченных по полосе частот Δf [5]. На рис. 1а показана зависимость $SSNR(\Delta f)$, полученная без учета приведенных выше рекомендаций, а на рис. 1б – с учетом таковых. В данном примере фигурируют зависимости $SSNR(\Delta f)$, полученные усреднением 8 оценок для мужской и женской речи. Как видим, оказалось достаточным повысить F_s (путем интерполяции) с 22050 Гц (рис. 1а) до 44100 Гц (рис. 1б), одновременно обеспечивая нечетный порядок нерекурсивных НЧ фильтров. На рис. 1в приведены результаты субъективного оценивания качества этих же речевых сигналов [6] с использованием шкалы Degradation Mean Opinion Score (DMOS) [3] (в эксперименте участвовало 17 слушателей, юношей и девушек 21-22 лет).

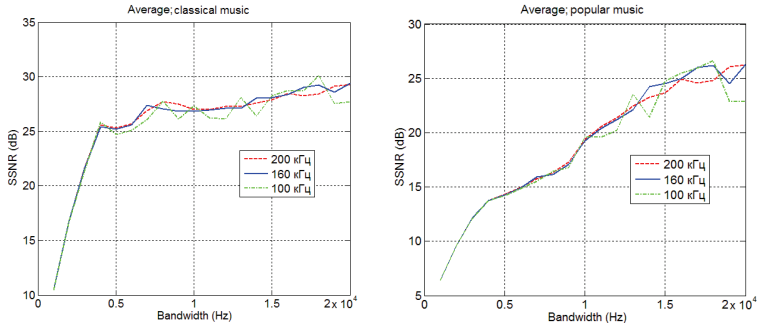
Анализ качества музыкальных сигналов, ограниченных по полосе частот, показал, что для получения монотонной зависимости $SSNR(\Delta f)$ частоту дискретизации следует увеличивать в большей степени – в 4-5 раз (рис. 2). Результаты субъективного оценивания (30 слушателей 20-25 лет) приведены на рис. 3 (пунктирными линиями указаны границы 95%-ного доверительного интервала). Как видим, среднее значение оценок DMOS монотонно и плавно повышается до значения Δf = 14 кГц, после чего

стабилизируется на уровне 4,5-4,7. Графики оценок $SSNR(\Delta f)$ заметно отличаются: до 4 кГц происходит быстрый рост, а после 4 кГц скорость роста заметно падает. Как и в случае речевых сигналов, можно говорить о достаточно хорошей согласованности результатов объективного и субъективного оценивания, свидетельствующей о действенности предложенных рекомендаций.



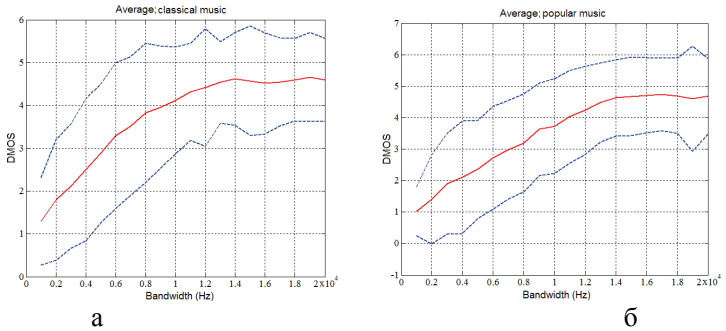
а б в

Рис. 2. Зависимости $SSNR(\Delta f)$ до учета особенностей SSNR (а) и после учета таковых (б) [5], а также субъективная оценка качества речевых сигналов [6]



а б

Рис. 2. Зависимости $SSNR(\Delta f)$ для музыки: классической (а) и эстрадной (б)



а б

Рис. 3. Субъективная оценка качества музыки: классической (а) и эстрадной (б)

2. ОСОБЕННОСТИ МЕРЫ LSD

Мера качества речевых сигналов, именуемая лог-спектральными искажениями (Log-Spectral Distortion – LSD), описывается соотношением:

$$LSD_{(p)} = \left(\frac{1}{F} \int_0^F \left| LX(f) - LY(f) \right|^p \right)^{1/p}, \quad LX(f) = 20\lg |X(f)|, \quad LY(f) = 20\lg |Y(f)|, \qquad (4)$$

где $|X(f)|$ и $|Y(f)|$ – спектры амплитуд эталонного процесса $x(t)$ и искаженного процесса $y(t)$, соответственно, F – полоса частот анализа, $p = 1,2,\infty$ – параметр, определяющий разновидность меры. В [7, 8] показано существование связей меры LSD_2 с кепстральными мерами. Несмотря на важность этих результатов, следует констатировать, что свойства меры LSD и ее оценок изучены недостаточно. Возможно, этим объясняется сравнительно редкое ее использование при экспериментальных исследованиях.

В ряде исследований мера LSD_1 показала себя как «субъективно значимая» [9-10], т.е. субъективно б*о*льшим искажениям сигнала соответствовали б*о*льшие значения меры LSD_1. Однако при оценивании качества речевых сигналов, ограниченных полосой частот Δf, оказалось, что зависимость $LSD_1(\Delta f)$ содержит локальные экстремумы (Рис. 4а), не исчезающие с увеличением времени анализа [11].

Для изучения данного феномена целесообразно использовать две модели сигналов: белый шум и узкополосный стационарный случайный процесс. Обоснованием такого выбора моделей может служить тот факт, что в речевом сигнале чередуются широкополосные (согласные) и узкополосные (гласные) звуки.

Представленные в [11] результаты исследований модели белого шума показали, что зависимость $LSD_1(\Delta f)$ является немонотонной и описывается соотношением

$$LSD(kF) = LSD^*(k) = 10\left[k\left|\lg k\right| + \left|\lg k + 2\lg M\right| \cdot (1-k)\right], \qquad (5)$$

где $k = \Delta f / F$ $(0 \le k \le 1)$ – нормированная полоса пропускания НЧ фильтра с прямоугольной АЧХ, $\lg M = 2{,}5$. Важно заметить, что графики зависимости (5), показанные на рис. 4б, получены для нормированных по стандартному отклонению сигналов. В отсутствие такого нормирования получаем линейную зависимость $LSD*(k) = 50 \cdot (1-k), \quad 0 \le k \le 1$, график которой представлен на рис. 4в.

Переходя к модели узкополосного сигнала, представим ее в виде отклика рекурсивного полосового фильтра второго порядка на воздействие в виде белого шума:

$$y_n = a_0 x_n - b_1 y_{n-1} - b_2 y_{n-2}, \qquad (6)$$

$a_0 = 1 + b_1 + b_2$, $b_1 = -2\rho\cos(\theta)$, $b_2 = \rho^2$, $\theta = 2\pi(f_c/F_s)\left[1 - 0.25(F_s/f_w)^2\right]^{1/2}$, $\rho = \exp(-\pi f_c / F_s)$, f_c и f_w – центральная частота и ширина полосы пропускания фильтра, соответственно; F_s – частота дискретизации. Форма АЧХ фильтра для f_c=700 Гц и f_w=50 Гц показана на рис. 5а. Эту АЧХ можно рассматривать также как модель форманты речевого сигнала.

Графики зависимости $LSD(\Delta f)$ (при нормировании сопоставляемых сигналов) для формант с частотами от 100 до 700 Гц приведены на рис. 5б. Как видим, в дополнение к «острому» локальному минимуму и следующему за ним «плавному» локальному максимуму, вблизи частоты форманты f_c появляется еще один сравнительно небольшой локальный максимум. В отсутствие нормирования сигналов зависимости $LSD(\Delta f)$ монотонно спадают (рис. 5в).

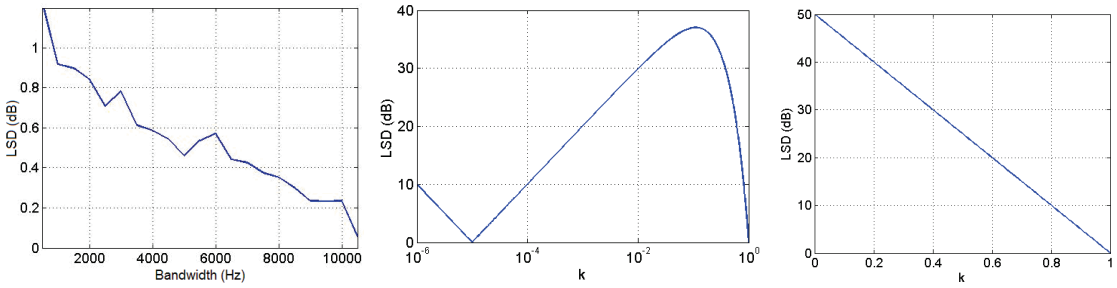


Рис. 4. Усредненная по 8 дикторам зависимость $LSD_1(\Delta f)$ [11] (а) и зависимость $LSD^*(k)$ для детерминированной версии модели белого шума: при нормировании по стандартному отклонению (а) и в отсутствие нормирования (б)

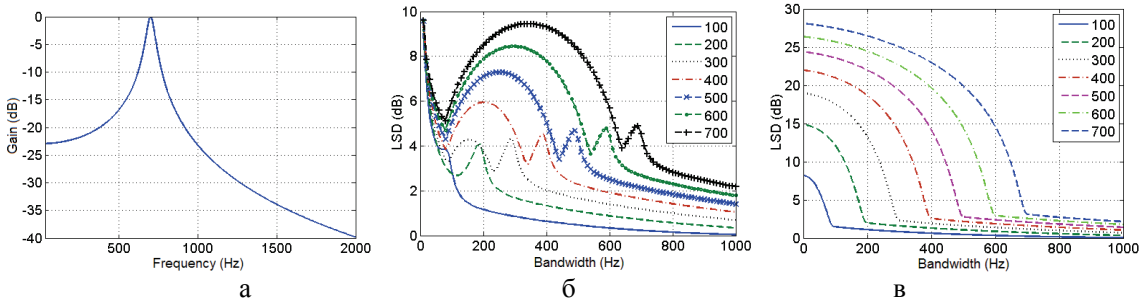


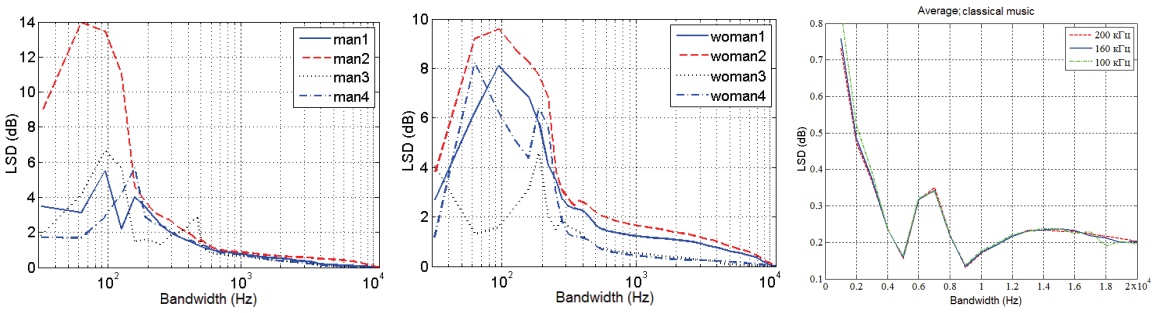
Рис. 5. Частотная характеристика фильтра для f_c =700 Гц , f_w =50 Гц (а) и зависимость $LSD(\Delta f)$ для модели узкополосного сигнала: при нормировании по стандартному отклонению (б) и в отсутствие такового (в)

Исследования реальных речевых сигналов показали, что оценки LSD весьма чувствительны к форме частотных характеристик НЧ фильтров, реализующих ограничение полосы частот. Чтобы избавиться от артефактов, обусловленных этой причиной, ограничение полосы частот сигналов производилась с помощью НЧ фильтров с идеально прямоугольными АЧХ (фильтрация в частотной области). Соответствующие оценки $LSD(\Delta f)$ приведены на рис. 6 для сигналов, нормированных по стандартному отклонению.

Как видим, нарушение монотонности зависимости $LSD(\Delta f)$ наблюдается лишь при сильных искажениях речевых сигналов (полоса частот 500 Гц и менее), что можно объяснить средоточием формант вблизи указанной полосы частот. Для музыкальных сигналов зависимости $LSD(\Delta f)$ не являются монотонными в значительно более широком диапазоне частот (рис. 6в), что можно пояснить большей сложностью спектров сигналов.

ЗАКЛЮЧЕНИЕ

Проведенный анализ особенностей мер SSNR и LSD качества речевых и музыкальных сигналов на примере сигналов, ограниченных по полосе, позволил дать объяснение обнаруженным артефактам и выработать ряд рекомендаций, направленных на улучшение согласованности результатов объективной и субъективной экспертизы.



а б в

Рис. 6. Оценки $LSD(\Delta f)$ для речи: мужчины (а), женщины (б) и для классической музыки (в)

ЛИТЕРАТУРА

1. *Quackenbush S., Barnwell T., and M. Clements*, Objective Measures of Speech Quality. – Englewood Cliffs, NJ: Prentice-Hall, 1988.
2. *Loizou P.* Speech enhancement: Theory and Practice. Second Edition. – Boca Raton: CRC Press, Taylor & Francis Group, 2013.
3. *Cote N.* Integral and diagnostic intrusive prediction of speech. – Berlin/Heidelberg: Springer-Verlag, 2011.
4. *Voran S.* Estimation of Speech Intelligibility and Quality, in Handbook of Signal Processing in Acoustics /Ed. by D. Havelock, S. Kuwano, M. Vorlander. – Springer Science+Business Media, LLC, 2008.
5. *Prodeus A.* Reducing Sensitivity of Segmental Signal-to-Noise Ratio Estimator to Time-Alignment Error, // International Journal of Electrical and Electronic Science. –2015. – Vol. 2(2). – P. 31–36.
6. *Замша К.С., Лозинский Б.В., Митяй Ю.А., Степановская Е.С., Продеус А.Н.* Объективное и субъективное оценивание качества речевых сигналов с ограниченной полосой частот // Electronics and Communications.– 2016.– Vol. 21, № 1(90).– P. 18–26.
7. *Gray R.M., Buzo A., Gray A.H., Matsuyama Y.* Distortion measures for speech processing, // IEEE Transactions on Acoustics Speech and Signal Processing.– 1980.– Vol. 28, No. 4.– P. 367–376.
8. *Gray A. H., Jr., and Markel J. D.* Distance measures for speech processing, // IEEE Trans. Acoust., Speech, Signal Processing.– 1976.– Vol. ASSP-24.– P. 380–391.
9. *Prodeus A., Didkovskyi V., Didkovska M., Kotvytskyi I.* On Peculiarities of Evaluating the Quality of Speech and Music Signals Subjected to Phase Distortion, // Proceedings of IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO).– Kyiv, Ukraine, 2017.– P. 455–460.
10. *Prodeus A., Didkovskyi V.S.* Objective estimation of the quality of radical noise suppression algorithms, // Radioelectronics and Communications Systems.– 2016.– Vol. 59, Issue 11.– P. 502–509 [doi:10.3103/S0735272716110042].
11. *Prodeus A.* On Some Features of Log-Spectral Distortion as Speech Quality Measure // Automation, Software Development & Engineering.– 2016.– Vol. 1. – P.1–9.