

Теория сигналов и систем

УДК 004.934

А.Н. Продеус, д.-р. техн. наук, Н.В. Богданова, канд. техн. наук

Национальный технический университет Украины “Киевский политехнический институт”,
ул. Политехническая, 16, г. Киев-56, 03056, Украина.

Оценка влияния нелинейности фазовой частотной характеристики тракта на качество музыкальных сигналов

Показано, что для слуховой системы человека приемлемыми являются фазовые искажения музыкальных сигналов, если максимальная разница групповых времен задержки тракта в области высоких и низких частот не превышает 70 мс. Анализ объективных показателей качества сигналов в виде сегментного отношения сигнал-шум (SSNR) и логарифмически-спектральных искажений (LSD), барк-спектральных искажений (BSD) и перцептуальной оценки качества речи (PESQ) свидетельствует, что все рассмотренные показатели адекватно реагируют на степень фазовых искажений музыкального сигнала. Библ. 7, рис. 6, табл. 1.

Ключевые слова: гребенка фильтров; фазовые искажения; качество музыкального сигнала; показатели качества.

Введение

В [1] рассмотрена модель возникновения

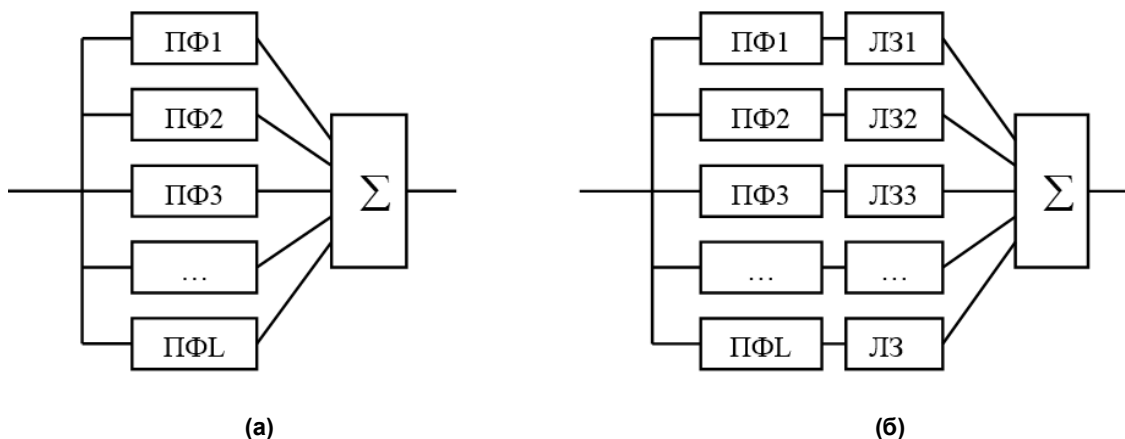


Рис. 1. Изображение гребенок с сумматором: без ЛЗ (а) и с ЛЗ (б)

Работ, посвященных исследованию данного вопроса, немного. В [4] отмечено, что удобной мерой нелинейности фазы $\theta(f)$ является отклонение групповой задержки $\tau(f) = -\frac{1}{2\pi} \frac{d\theta(f)}{df}$ от константы. Чувствительность слуховой системы

фазовых искажений сигнала в гребенке цифровых нерекурсивных фильтров с сумматором (рис. 1). Данная модель возникновения фазовых искажений сигнала представляет значительный практический интерес, поскольку гребенки фильтров широко используются в системах записи и воспроизведения, кодирования и декодирования акустических сигналов, в линиях связи, в системах коррекции слуха [2, 3].

При значительной нелинейности фазовой частотной характеристики (ФЧХ) такой гребенки искажения выходного сигнала становятся ощутимыми на слух. Введя линии задержки (ЛЗ) в частотные каналы (рис. 1,б), этот недостаток можно устранить. Поскольку использование ЛЗ усложняет систему в целом, возникает естественный вопрос о допустимой степени нелинейности ФЧХ линейной системы при передаче музыкальных и речевых сигналов.

человека к неравномерности $\tau(f)$ исследована в [5], где экспериментально показано, что 1-3 мс является пороговым значением, а для тренированного слуха этот порог снижается даже до 400 мкс. Вместе с тем, в [5] указано, что при использовании речевых сигналов фазовые искажения менее заметны. Для речевых сигналов

пороговые значения неравномерности групповой задержки так и не были определены, хотя для практических целей именно эти значения представили бы наибольший интерес.

В работе [6] этот недостаток устранен: показано, что для слуховой системы человека приемлемыми являются фазовые искажения речевых сигналов, если максимальная разница групповых времен задержки тракта в области высоких и низких частот не превышает 50 мс. К сожалению, для музыкальных сигналов аналогичное пороговое значение оставалось не определенным до последнего времени. Цель данной работы состоит в восполнении указанного пробела.

1. Модель возникновения фазовых искажений сигнала

Модель возникновения фазовых искажений сигнала, основанная на использовании гребенки октавных фильтров, предложена в [1] и модифицирована в [6]. При исследовании речевых сигналов охватывалась типичная для речевых сигналов полоса частот 90-11000 Гц [6]. Основные параметры фильтров, составляющих гребенку и рассчитанных методом Ремеза, приведены в табл. 1, где f_0 - центральная частота; Δf - полоса пропускания; n - порядок фильтра.

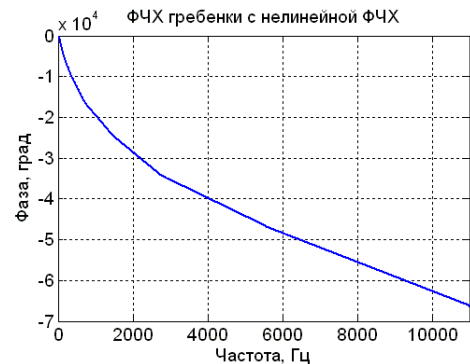
Таблица 1. Параметры гребенки октавных фильтров

f_0 , Гц	125	250	500	1000	2000	4000	8000
Δf , Гц	90	180	355	710	1400	2800	5600
n	4353	2903	2177	1320	927	545	437

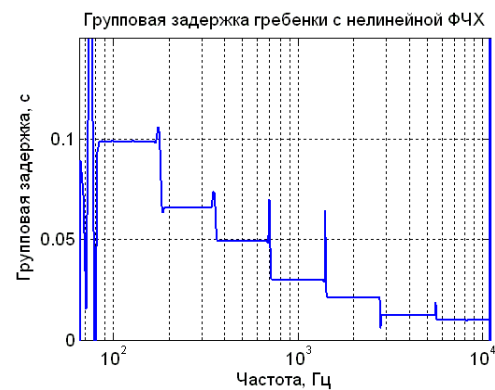
Амплитудная частотная характеристика (АЧХ) каждого фильтра гребенки практически прямоугольная, а ФЧХ – линейная в силу симметричности импульсных характеристик (ИХ) каждого из фильтров.

Хотя суммарные АЧХ обеих гребенок одинаковы, суммарная ИХ первой гребенки (рис. 1,а) несимметричная, а ФЧХ нелинейная (рис. 2,а), поскольку протяженности парциальных ИХ каждого из полосовых фильтров различны. Наличие ЛЗ во второй гребенке (рис. 1,б) позволяет добиться симметричности суммарной ИХ и, как следствие, линейности ФЧХ. Достигают этого, управляя временами задержки таким образом, чтобы максимумы парциальных ИХ каждого из каналов совпали.

Групповое время задержки $\tau(f)$ для гребенки рис. 1,а с нелинейной ФЧХ различно для различных частотных компонентов (рис. 2,б). Конфигурацию ФЧХ, где ВЧ компоненты сигнала отстают от НЧ компонентов (рис. 2,б, «убывающее время задержки»), обозначим ФЧХ1. Конфигурацию с «возрастающим временем задержки» обозначим ФЧХ2.



(а)



(б)

Рис. 2. Графики ФЧХ (а) и $\tau(f)$ (б) для гребенки рис. 1 (а)

При оценке «порогового» значения величины $\Delta\tau_{\max}$, когда слух человека начинает воспринимать искажения музыкального сигнала, рассмотрены обе конфигурации для значений $\Delta\tau_{\max} \approx 30, 50, 70$ и 90 мс.

2. Оценивание качества акустического сигнала

В данной работе, как и в [6], помимо субъективного оценивания, использованы четыре объективных показателя качества сигнала: сегментное отношение сигнал-шум (Segmental Signal to Noise Ratio - SSNR), логарифмически-спектральные искажения (Logarithmic Spectral Distortion - LSD), барк-спектральные искажения (Bark Spectral Distortion - BSD) и перцептуаль-

ное качество речи (Perceptual Evaluation of Speech Quality - PESQ) [7]. Первые два показателя – SSNR и LSD – весьма привлекательны в силу простоты вычислений, тогда как другие два показателя – BSD и PESQ – позволяют учесть, с различной степенью точности, особенности слуховой системы человека.

Аналитическое описание упомянутых выше показателей SSNR, LSD и BSD:

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

$$LSD = \frac{2}{KL} \sum_l \sum_{k=0}^{\frac{K}{2}-1} |G\{X(l,k)\} - G\{Y(l,k)\}|, \quad (2)$$

$$G\{X(l,k)\} = \max\{20 \lg(|X(l,k)|), \delta\},$$

$$\delta = \max_{l,k} \{20 \lg(|X(l,k)|)\} - 50$$

$$BSD = \frac{\sum_{l=1}^L \sum_{k=0}^{\frac{K}{2}-1} [B\{X(l,k)\} - B\{Y(l,k)\}]^2}{\sum_{l=1}^L \sum_{k=0}^{\frac{K}{2}-1} [B\{X(l,k)\}]^2}, \quad (3)$$

где $x(l,n)$ и $y(l,n)$ - n -я выборка l -го фрейма

входного и выходного сигналов фильтра $x(n)$ и $y(n)$, соответственно; $X(l,k)$ и $Y(l,k)$ - амплитудные спектры l -го фрейма сигналов $x(n)$ и $y(n)$, соответственно; $B\{X(l,k)\}$ и $B\{Y(l,k)\}$ - барк-спектры l -го фрейма сигналов $x(n)$ и $y(n)$, соответственно.

Аналитическое описание алгоритма вычисления показателя PESQ весьма громоздко и приведено в [7]. В данной работе использована вторая версия этого показателя (ITU-T Rec. P.862.2), позволяющая анализировать сигналы в широкой полосе частот (до 7 кГц).

3. Результаты оценивания качества музыкальных сигналов

При экспериментальном оценивании зависимости качества музыкального сигнала от характера и степени нелинейности ФЧХ фильтра использованы фрагменты восьми музыкальных произведений протяженностью 30-45 секунд каждый. При этом половина произведений принадлежала жанру «классическая музыка» («Ave Maria» Дж. Каччини, «Этюд №4, соч. 10, Ф. Шопена», 5-я симфония П. Чайковского, увертюра «Фауст» Р. Вагнера), а половина – жанру «популярная музыка» («Mama-mia» ABBA, «Shes_Leaving_Home» Beatles, «Я піду в далекі гори» К. Цісик, «Mademoiselle Hyde» L. Fabian). Все сигналы записаны с частотой дискретизации 22050 Гц и битовой глубиной 16 бит.

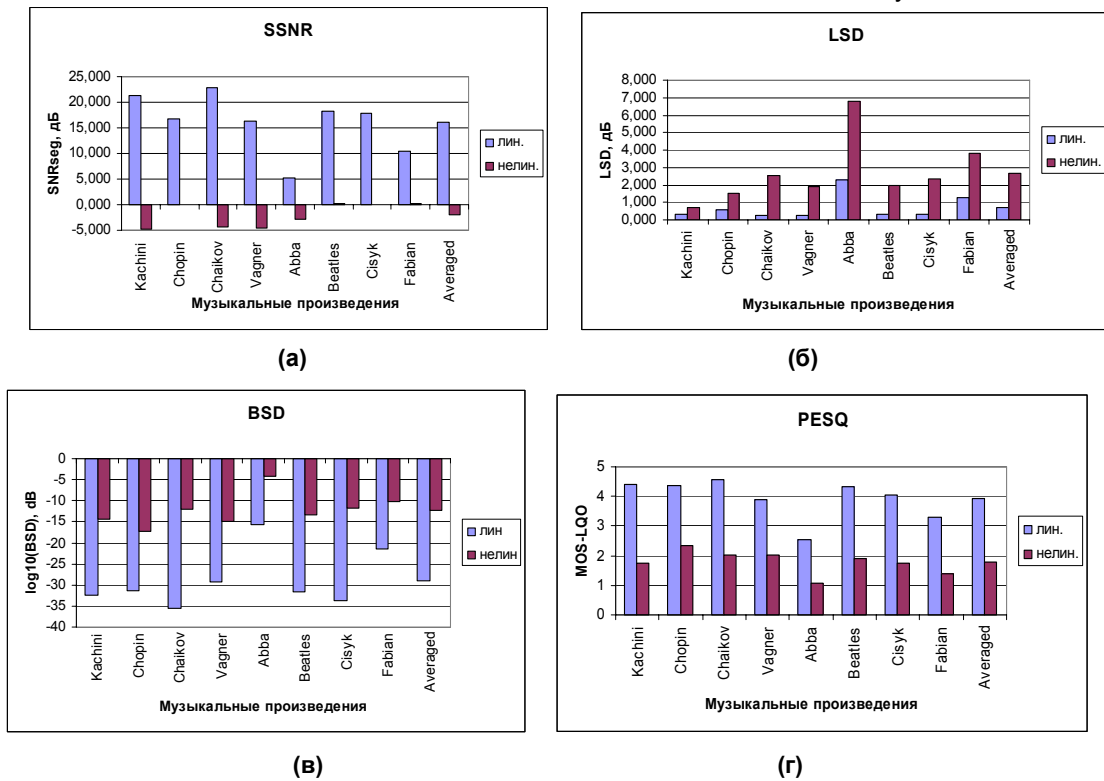


Рис. 3. Диаграммы значения объективных показателей

Оценивание качества звучания производилось как на слух, так и с использованием перечисленных выше объективных показателей. Оценки показателей SSNR, LSD BSD и PESQ для музыкальных сигналов, пропущенных через гребенки с линейной и нелинейной ФЧХ, приведены на рис. 3.

Нетрудно видеть, что все рассмотренные показатели качества адекватно отреагировали на нелинейность ФЧХ гребенки фильтров, засвидетельствовав существенное ухудшение качества музыкального сигнала.

Вместе с тем, приведенные графики свидетельствуют, что степень такого ухудшения неодинакова как для различных жанров, так и для отдельных произведений фиксированного жанра. В этой связи целесообразно произвести углубленный анализ причин такого различия.

Результаты фильтрации фрагмента записи фортепианного произведения (Шопен, этюд №4, соч. 10) фильтрами с линейной ФЧХ показаны на рис. 4,а, а с нелинейными ФЧХ показаны для $\Delta\tau_{\max} = 0,09$ с на рис. 4,б,в.

Данный пример интересен тем, что отдельные звуки чередовались примерно в таком же темпе, что и звуки быстрой речи (длительность отдельного звука была близкой 0,08 с). Малое время атаки, обусловленное ударным принципом звукоизвлечения фортепиано, обеспечило широкополосность спектра звука в его начальной части. На интервале установления затухания спектры звуков сузились. Поэтому

можно утверждать, что для быстрой фортепианной музыки наблюдается чередование широкополосных и узкополосных импульсов, аналогично тому, как это происходит в речевом сигнале.

Несмотря на отмеченное сходство с речевым сигналом, тем не менее, в данном примере с трудом удалось на слух определить ситуацию, не совсем «благополучную» с точки зрения звучания. Например, на рис. 4,а прямоугольником выделена область, в пределах которой одновременно звучат три звука, С#3, С#5 и С#6 (до-диез малой, второй и третьей октав), которым соответствуют частоты 138,6 Гц, 554,4 Гц и 1108,7 Гц (четвертый звук С#2 с частотой 69,3 Гц оказался вне полосы пропускания гребенки фильтров и был подавлен). Поскольку интенсивности трех звуков сложились, уровень результирующего сигнала заметно выше уровней соседних одиночных звуков – один из таких одиночных звуков (С6, 1046,5 Гц) выделен горизонтальным овалом. На начальном участке области, выделенной прямоугольником, просматривается начало звука С#6 (обведено вертикальным овалом), несколько опережающего остальные звуки. Однако на Рис. 4,б звук С#6 уже заметно опережает остальные два звука, что объясняется нелинейностью ФЧХ1. На слух данный участок сигнала воспринимается как несколько неприятное «кваканье». Этот эффект отсутствует для гребенки с ФЧХ2 (рис. 4,в), где высокочастотный звук С#6 запаздывает по отношению к низкочастотным звукам.

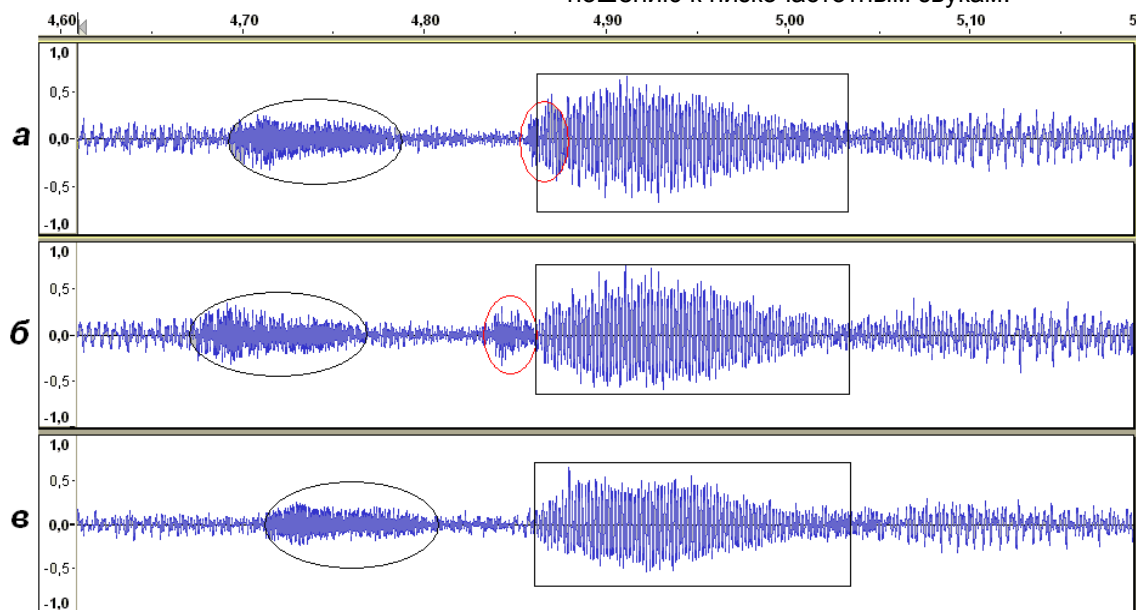


Рис. 4. Изображение сигнала: линейная ФЧХ (а), нелинейная ФЧХ1 (б), нелинейная ФЧХ2 (в)

Для органной, камерной, вокальной и инструментальной музыки классического жанра чередование широкополосных и узкополосных звуков является нетипичным, и можно предположить, что вероятность восприятия фазовых искажений для такой музыки будет малой. Действительно, экспериментальная проверка влияния нелинейности ФЧХ на восприятие звучания остальных трех произведений показала, что для $\Delta\tau_{\max} \leq 90$ субъективно искажения сигнала не обнаруживаются. Кроме того, прослушивание дополнительного произведения - «Болеро» Равеля, для которого характерен четкий ритм и где в оркестровке использованы медные духовые и ударные, - показало, что при насыщенном звучании музыкального полотна фазовые искажения обнаружить крайне сложно даже при $\Delta\tau_{\max} \approx 90$.

Казалось бы, для популярной музыки вероятность обнаружить фазовые искажения (при той же степени нелинейности ФЧХ) должна быть меньше, чем для классической музыки, поскольку исполнители популярной музыки обладают значительно большей свободой в плане вариаций темпа и тембра, не говоря уже о принципиальной возможности импровизации.

Тем не менее, факты не подтверждают это предположение. Как свидетельствуют графики рис. 3, а также результаты прослушивания, для произведений «Mama-mia» ABBA и «Mademoiselle Hyde» L. Fabian замечено существенное ухудшение качества звука при $\Delta\tau_{\max} \approx 70 - 90$ мс. Для первого произведения характерным является быстрое «упругопульсирующее» звучание, а также важная роль

в оркестре ударных и духовых инструментов. Из-за фазовых искажений нарушается «упругость» звучания, появляется ощущение нарушения ритма. Во втором случае ощутимые искажения наблюдаются в момент коротких барабанных ударов, воспринимаемых из-за фазовых искажений не как «сухие щелчки», а как «щелканье хлыста».

Следует отметить, что оцениваемое на слух пороговое значение $\Delta\tau_{\max}$, при котором становятся заметными фазовые искажения в отмеченных выше случаях, несколько выше, чем для речевых сигналов, и близко 70 мс. Данный вывод согласуется с результатами объективного оценивания (рис. 5), где максимальная крутизна графиков SSNR, BSD и PESQ наблюдается для $\Delta\tau_{\max} \leq 70$ мс (исключением является показатель LSD, скорость роста которого увеличивается и для $\Delta\tau_{\max} > 70$ мс).

Наиболее заметны фазовые искажения музыки в тех произведениях, где высока роль ударных инструментов. На рис. 6 приведены форма сигналов солирующих ударных для ситуации, когда максимальная разница групповых времен задержки ФЧХ в области высоких и низких частот составляет 90 мс. На рис. 6,а показан фрагмент формы волны сигнала на выходе фильтра с линейной ФЧХ. Кружками обведены импульсы бас-барабана, квадратом обведен импульс барабана-тома, квадратом с закругленными углами обведен импульс тарелки-крэша. В данном примере, как и в предыдущих, на слух фазовые искажения для ФЧХ1 воспринимаются как более сильные, по сравнению с ФЧХ2.

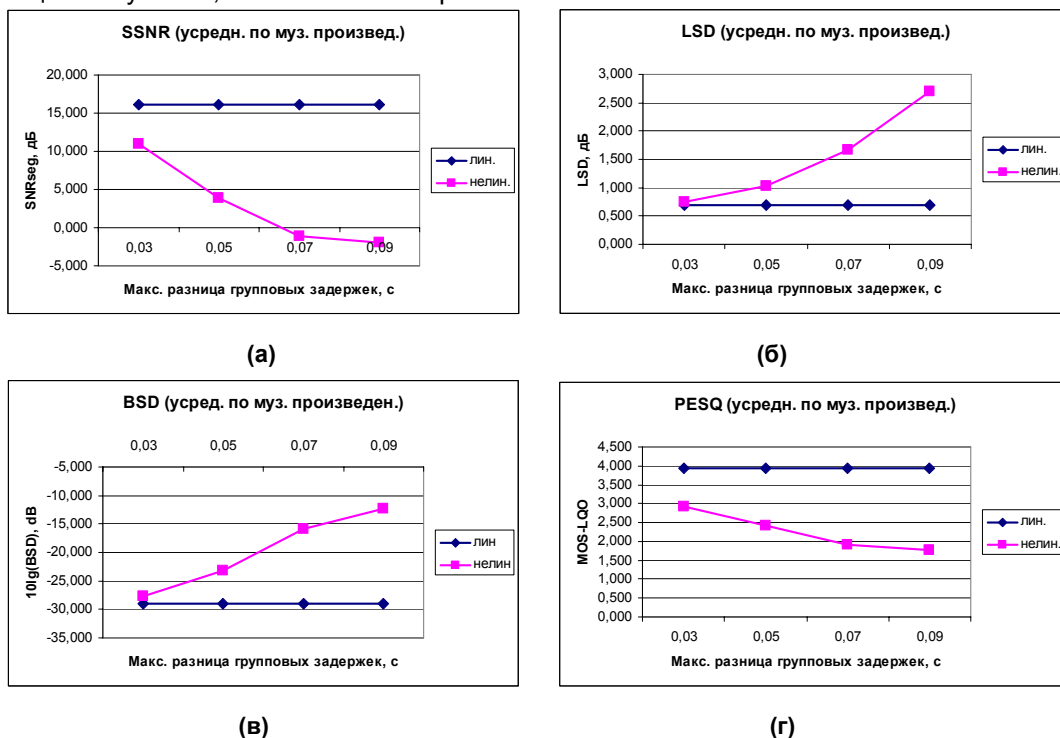


Рис. 5. Графики зависимости объективных показателей от $\Delta\tau_{\max}$

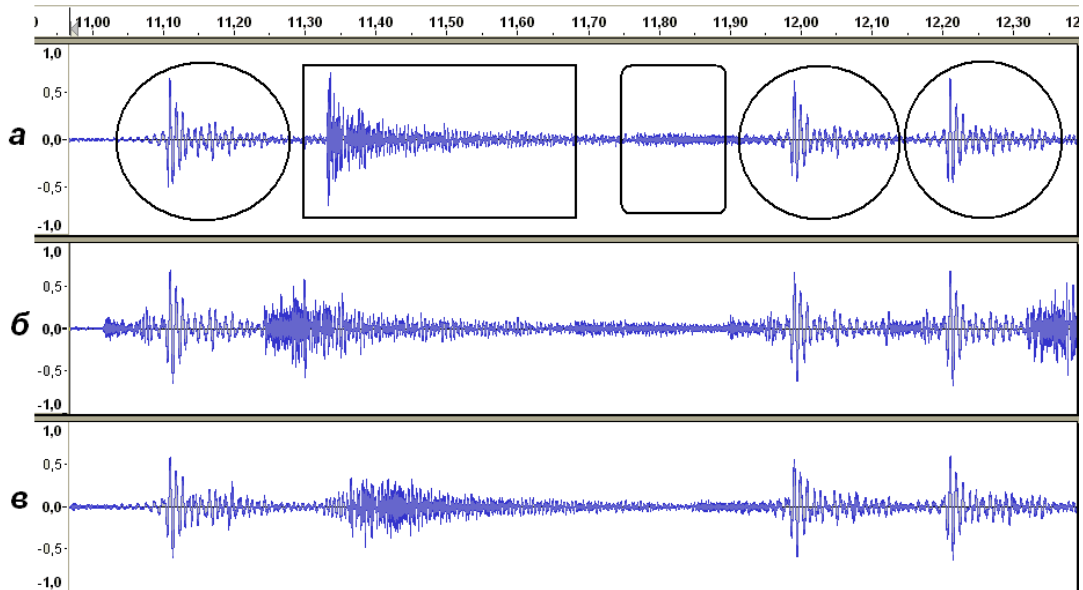


Рис. 6. Графики сигнала: линейная ФЧХ (а), нелинейная ФЧХ1 (б), нелинейная ФЧХ2 (в)

Обобщая полученные результаты, отметим, что музыкальные сигналы в меньшей степени, по сравнению с речевыми сигналами, подвержены искажениям, обусловленным нелинейностью ФЧХ тракта передачи. Пояснить это можно тем, что для музыки, в отличие от речи, быстрое чередование узкополосных и широкополосных звуковых импульсов является нетипичным. В наибольшей степени подвержены фазовым искажениям те музыкальные произведения или их фрагменты, где значительную роль играют ударные инструменты, при этом искажения для ФЧХ вида ФЧХ1 субъективно заметнее таковых для вида ФЧХ2.

Выводы

Экспериментальные исследования показали, что для слуховой системы человека приемлемыми являются фазовые искажения музыкальных сигналов, если максимальная разница групповых времен задержки в области высоких и низких частот не превышает 70 мс.

Объективные показатели SSNR, LSD, BSD и PESQ в целом адекватно реагируют на степень фазовых искажений музыкального сигнала. При этом пороговому значению $\Delta\tau_{\max} \approx 70$ мс соответствуют следующие пороговые значения объективных показателей: минус 1 дБ для SSNR; 1,7 дБ для LSD; минус 15 дБ для $10\lg\text{BSD}$ и 1,9 MOS для PESQ.

Фазовые искажения музыкального сигнала, обусловленные меньшей временной задержкой высокочастотных компонентов, по сравнению с низкочастотными компонентами, субъективно

воспринимаются как более сильные, по сравнению с ситуацией, когда большей задержке подвергаются НЧ компоненты речевого сигнала.

Список использованных источников

1. Дидковский В.С., Дидковская М.В., Продеус А.Н. Акустическая экспертиза каналов речевой коммуникации. Монография. – К.: Имэкс-ЛТД, 2008. – 420 с.
2. *Advances in Digital Speech Transmission* / Edited by Martin R., Heute U. and Antweiler C. - John Wiley & Sons Ltd, England, 2008. – 572 p.
3. *Blauert J. Communication acoustics* / Edited by Blauert J. - Springer Berlin Heidelberg New-York, 2005. – 385 p.
4. *Оппенгейм А., Шафер Р.* Цифровая обработка сигналов. – М.: Техносфера, 2006. – 858 с.
5. *Blauert J. Group delay distortions in electroacoustical systems* / Blauert J. // *J. Acoust. Soc. Am.* – Vol.63, No.5. – 1978. – P. 1478-1483.
6. *Продеус А.* Оценка влияния нелинейности фазовой частотной характеристики системы на качество речевых сигналов / Продеус А.Н., Пилипенко К.П., Калужный А.Я., Бартнев С.Г. // *Электроника и связь*, 2015, Т. 20, №2(85), С. 33-40.
7. *Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 - October 2005.*

Поступила в редакцию 15 апреля 2015 г.

УДК 004.934

А.М. Продеус, д.-р. техн. наук, **Н.В. Богданова**, канд. техн. наукНаціональний технічний університет України «Київський політехнічний інститут»,
ул. Політехнічна, 16, м. Київ-56, 03056, Україна.

Оцінка впливу нелінійності фазової частотної характеристики тракту на якість музичних сигналів

Показано, що для слухової системи людини прийнятними є фазові спотворення музичних сигналів, якщо максимальна різниця групових часів затримки в області високих і низьких частот не перевищує 70 мс. Аналіз об'єктивних показників якості сигналів у вигляді сегментного відношення сигнал-шум (SSNR), логарифмічно-спектральних спотворень (LSD), барк-спектральних спотворень (BSD) і перцептуальної оцінки якості мовлення (PESQ) свідчить, що всі розглянуті показники адекватно реагують на ступінь фазових спотворень музичного сигналу. Бібл. 7, рис. 6, табл. 1.

Ключові слова: гребінка фільтрів; фазові спотворення; якість музичного сигналу; показники якості.

UDC 004.934

A. Prodeus, Dr.Sc., **N. Bogdanova**, Ph.D.National Technical University of Ukraine "Kiev Polytechnic Institute",
st. Polytechnique, 16, Kiev-56, 03056, Ukraine.

Assessment of the impact of system phase response non-linearity on the music signals quality

It is shown that phase distortion of music signals are acceptable for human auditory system when the maximum difference of group delay times in the high and low frequencies is below 50. Analysis of objective indicators of quality of signal in the form of a segment signal-to-noise ratio (SSNR), log-spectral distortion (LSD), the bark spectral distortion (BSD) and perceptual evaluation of speech quality (PESQ) indicates that all these measures adequately respond to the degree of phase distortion of the music signal. Reference 7, figures 6, table 1.

Keywords: filter bank; phase distortion; music quality; quality indicators.

References

1. Didovskiy, V. S., Didovskaia, M. V., Prodeus, A. N. (2008). Acoustic assessment of speech communication channels. Monograph. K.: Imex-Ltd, 420 p. (Rus)
2. Martin, R., Heute, U. and Antweiler, C. (Ed.). (2008). Advances in Digital Speech Transmission. John Wiley & Sons Ltd, England, P. 572.
3. Blauert, J. (Ed.). (2005). Communication acoustics. Springer Berlin Heidelberg New-York, P. 385.
4. Oppenheim, A., Schaffer, R. (2006). Digital signal processing. M.: Techospera, P. 858. (Rus)
5. Blauert, J. (1978). Group delay distortions in electroacoustical systems. J. Acoust. Soc. Am., Vol. 63, No.5, Pp. 1478-1483.
6. Prodeus, A., Pylypenko, K., Kalyuzhny, A., Bartenev, S. (2005). Assessment of the impact of system phase response non-linearity on the speech signals quality. Electronics and Communication, Vol. 20, No2(85), Pp. 33-40. (Rus).
7. (2005). Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 - October 2005.