МОБИЛЬНАЯ СИСТЕМА ОЦЕНКИ КАЧЕСТВА ПЕРЕДАЧИ РЕЧИ

MOBILE VOICE QUALITY ASSESSMENT SYSTEM

M. Gusev
I. Guseva

Annotation

In the article the algorithm for estimating the quality of voice and audio signals are describes. It offers a software implementation of the algorithm, oriented to work on mobile devices. The experimental results confirm the effectiveness and the practical significance of the proposed solutions.

Keywords: sound quality evaluation, MOS, Android, psychoacoustics, AquA.

Гусев Михаил Николаевич
К.т.н., Санкт–Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч–Бруевича, Санкт–Петербург
Гусева Инна Юрьевна
Санкт–Петербургский политехнический университет Петра Великого, Санкт–Петербург

Аннотация

В статье рассматривается алгоритм оценки качества передачи речи и звуковых сигналов. Предлагается программная реализация алгоритма, ориентированная на работу в мобильных устройствах. Приводятся результаты экспериментов, подтверждающие эффективность и практическую значимость предложенных решений.

Ключевые слова:

Оценка качества передачи речи, MOS, Андроид, психоакустика, AquA.

Введение

Оказание коммерческих услуг сотовой связи в сетях GSM (Global System for Mobile telecommunications) началось в середине 1991 года. В 1993 году существовало уже 36 сетей стандарта GSM в 22 странах мира, а к 1994 году число абонентов GSM сетей в мире достигло 1,3 миллиона. Сегодня более трети населения земного шара использует услуги на базе стандарта GSM.

Качество услуг сотовой связи является ключевым фактором, влияющим на конкурентоспособность операторов связи. Соответственно, исследование вопросов обеспечения качества услуг мобильной связи является актуальной задачей.

Для повышения качества оказания услуг сотовой связи в условиях конкуренции на рынке телекоммуникационных услуг необходима разработка методов, средств и регламентов оценки качества связи. Результаты такой разработки будут востребованы как операторами связи так и федеральными органами исполнительной власти, осуществляющими контроль деятельности в области связи, а также гражданами и организациями, являющимися пользователями услуги связи.

В связи со сказанным представляется целесообразным разработка системы оценки качества передачи речи на базе смартфонов.

Выбор платформы и средств разработки

Согласно исследованиям компании IDC (International Data Corporation) рынок между мобильными операционными системами распределен следующим образом [1] (**табл. 1**):

Таблица 1.

Доли рынка мобильных платформ.

Платформа	Доля рынка, %
Android	68,3
iOS	18,8
Windows	2,6
BlackBarry, Linux	10,3

Целесообразным представляется разработка программного обеспечения для наиболее популярной мобильной платформы – Android.

Само мобильное приложение под Android разрабатывается на языке Java. Для реализации математических функций удобно использовать язык C/C++. Среда разра–

ботки принципиального значения не имеет – может использоваться как Android Studio так и Eclipse.

Ядро системы оценки качества

В качестве ядра системы оценки качества передачи речевого сигнала воспользуемся алгоримом AQuA [2, 3], дополненным психоакустической моделью, предложенной в [4].

Схема алгоритма оценки качества представлена на **рис. 1**.

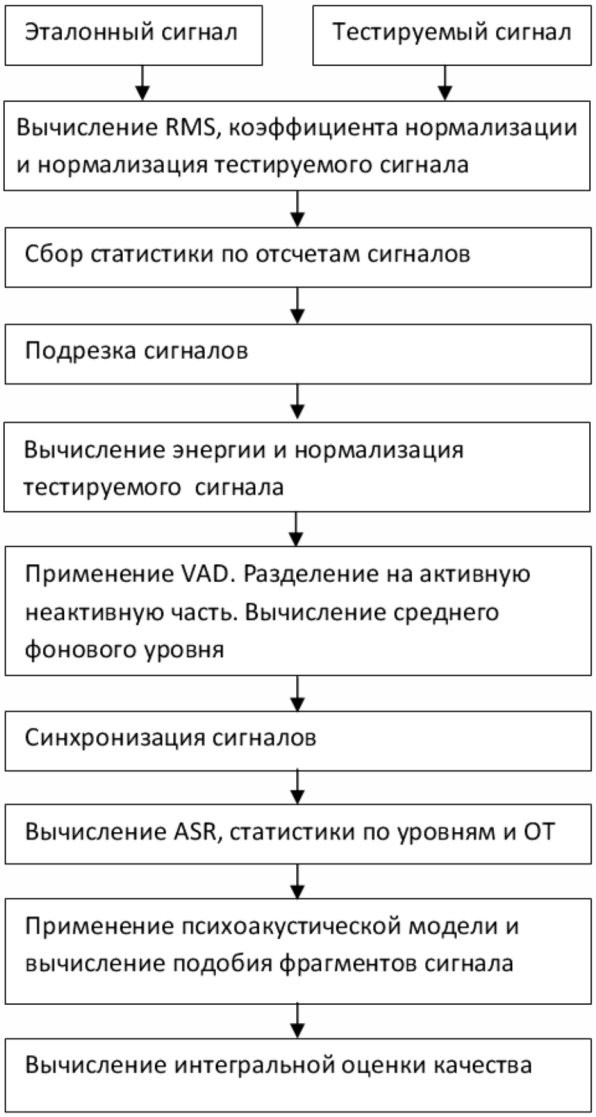

Рисунок 1. Схема алгоритма AQuA.

Рассмотрим подробнее основные этапы работы алгоритма.

На вход системы оценки качества подаются два сигнала: исходный и тестовый. Качество исходного (или эталонного) сигнала принимается на 100%, он считается идеальным, не содержащим искажений. Предполагается, что тестовый сигнал получен из эталонного путем передачи по каналам сотовой связи, и может содержать различные искажения.

Вычисление RMS и нормализация

На первом шаге обработки вычисляются RMS (Root Mean Square) сигналов и, если включена нормализация по RMS, выполняется нормализация – RMS тестового сигнала приводится к RMS эталонного сигнала.

Для каждого сигнала вычисляется пара значений RMS (1) и RMS_{bound} (2):

$$RMS = \frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{x_i}{32768}\right)^2 \quad (1)$$

$$RMS_{bound} = \frac{1}{N}\sum_{i=0}^{N-1}\left(\begin{cases}\frac{x_i}{32768}, & xl \le \frac{x_i}{32768} < xh \\ 0, & иначе\end{cases}\right)^2, \quad (2)$$

где

$$xl = \begin{cases}1.1\cdot\min(X), & 1.1\cdot\min(X) \ge 0{,}0078125 \\ 0{,}0078125, & 1.1\cdot\min(X) < 0{,}0078125\end{cases},$$

$$xh = 0{,}9\cdot\max(X)$$

Значение RMS_{bound} представляет собой попытку исключить из рассчета коэффициента нормализации низкоуровневую часть сигнала, а также пиковые значения отсчетов, образующиеся в результате искажений.

Статистика по отсчетам и "подрезка" сигналов

Статистика, собираемая по сигналу, включает в себя минимум, максимум и среднее значение энергии и отсчетов, а также относительное количество клиппированных отсчетов сигнала (3) и SNR (signal-to-noise ratio).

$$Clp = \frac{100}{N}\sum_{i=1}^{N-1}\begin{cases}1, & x_{i-1} = x_i = \min(X) \text{ или} \\ & x_{i-1} = x_i = \max(X) \\ 0, & иначе\end{cases} \quad (3)$$

Далее, для повышения точности синхронизации сигналов по VAD (Voice Activity Detection), выполняется удаление начальной и конечной тишины из обоих обрабатываемых сигналов. Удаление тишины производится по пороговому значению энергии отсчетов в дБ, причем может использоваться как абсолютное пороговое значение, так и значение, вычисляемое относительно минимальной

энергии сигнала. Уровень энергии отсчета вычисляется по формуле (4).

$$e_i = 10 \cdot \lg\left(\frac{x_i^2}{240} + const \right). \quad (4)$$

Нормализация по энергии

Если включена нормализация сигнала по энергии, вычисляются средние значения максимумов сигналов $\mathrm{Max_{Avg}}$, коэффициент нормализации сигнала, и выполняется нормализация тестируемого сигнала. При расчете значения $\mathrm{Max_{Avg}}$ используются только максимумы, превосходящее среднее значение амплитуды сигнала (5).

$$Max_{Avg} = \frac{1}{NC} \sum_{i=1}^{N-1} \begin{cases} |x_i|, & x_i > avg(X)\,\delta \\ & x_{i-1} < x_i < x_{i+1} \\ 0, & иначе \end{cases}, \quad (5)$$

где

$$NC = \sum_{i=1}^{N-1} \begin{cases} 1, & x_i > avg(X)\delta \\ & x_{i-1} < x_i < x_{i+1} \\ 0, & иначе \end{cases}.$$

Использование среднего значения максимумов и ограничения, вызвано желанием исключить из нормализации низкоуровневую часть сигнала.

Применяемый алгоритм VAD и синхронизация сигналов по результатам его работы детально описано в [2, 3].

ASR, статистики по уровням и ОТ

ASR (Active Speech Ratio) вычисляется на основе результатов работы алгоритма VAD как отношение количества активных фреймов в сигнале к общему количеству фреймов в сигнале, выраженное в процентах.

Статистики по уровням и ОТ (Основной Тон) строятся по гистограммам уровней, скорости изменения значений отсчетов и основного тона. Коэффициент искажений вычисляется как (6):

$$Dist = \frac{100}{ND} \sum_{i=1}^{N-1} \begin{cases} \dfrac{|h1_i - h2_i|}{h1_i + h2_i}, & h1_i - h2_i > 0 \\ 0, & иначе \end{cases}, \quad (6)$$

где

$$ND = \sum_{i=1}^{N-1} \begin{cases} 1, & h1_i - h2_i > 0 \\ 0, & иначе \end{cases}.$$

Психоакустическая модель включает в себя три уровня:

- пси–фильтрацию;
- нормализацию уровней;
- перевод в различимые градации.

Основу психоакустической модели составляют различные полученные экспериментально зависимости, оформленные в виде таблиц значений.

Пси–фильтрация [3] – наиболее сложный уровень обработки. На рис. 2 представлена укрупненная схема пси–фильтра. Поступающий фрейм данных сохраняется в блоке текущего фрейма, и передается на вход пси–маскера. На основе поступивших данных формируется пре–маска. Пре–маска накладывается на предыдущий фрейм данных, и результат маскирования принимается за выходное значение фильтра.

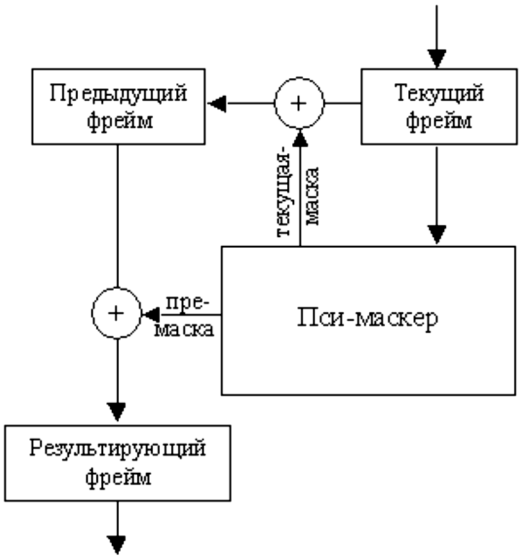

Рисунок 2. Укрупненная схема пси-фильтра.

Кроме того, маскером формируется текущая маска, являющаяся объединением внутренней – и пост – масок. Текущая маска накладывается на текущий фрейм данных, и результат маскирования сохраняется в блоке предыдущего фрейма.

Процесс маскирования описывается формулой (7):

$$s_i = \begin{cases} s_i, & s_i > m_i \\ 0, & s_i \le m_i \end{cases}, \quad (7)$$

где

s_i– компонента спектра с номером i;

m_i – компонента маски с номером i.

Процедура построения масок включает следующую последовательность действий:

- обработка порога слышимости;
- маскирование флюидных уровней;
- разделение спектра на тоны и шумы;
- построение масок от тональных компонент;
- построение масок от шумовых компонент;
- объединение масок от тональных и шумовых компонент.

Порог слышимости характеризует чувствительность уха к интенсивности звуковой энергии. Один из возможных вариантов определения порога слышимости зафиксирован в стандарте ISO/R-226 (International Organization for Standartization). Обработка порога слышимости заключается в построении соответствующей маски mt_i путем интерполяции (8):

$$mt_i = HThres_{k-1} + \frac{(F_i - THFreq_{k-1})(HThres_k - HThres_{k-1})}{THFreq_k - THFreq_{k-1}}. \quad (8)$$

Частота, соответствующая индексу определяется как (9):

$$F_i = \frac{(i-0.5)\cdot SampleRate}{2\cdot(SpecSize-1)}. \quad (9)$$

Маскирование флюидных уровней позволяет избежать ошибок вычислений, связанных с эффектом растекания спектра. Значения флюидных уровней рассчитываются относительно максимальной компоненты спектра. Пересчет компонент маски выполняется по формуле (10):

$$mt_i = \max\left(mt_i, 0.01\cdot \max_{j}^{SpecSize}(s_j)\right). \quad (10)$$

Разделение спектра на тональные и шумовые компоненты связано с различиями в процессе построения масок.

При разделении спектра используется простейший алгоритм, выделяющий пики:

- ищутся локальные максимумы, уровень которых превышает некоторое пороговое значение;
- слева и справа от локальных максимумов ищутся локальные минимумы;
- компоненты спектра, между найденными парами локальных минимумов считаются тональными;
- оставшиеся компоненты спектра – шумовыми.

Степень маскировки определяется как разность в децибелах между уровнем порога слышимости маскируемого тона в присутствии маскирующего тона и уровнем порога слышимости маскируемого тона в тишине. Общее описание всех возможных кривых маскировки представляется весьма затруднительным, поэтому в рамках решаемой задачи было решено использовать упрощенную модель маскировки, близкую к используемой в стандарте MPEG (Moving Picture Coding Experts Group).

Для каждой выделенной тональной компоненты строятся маски путем интерполяции промежуточных кривых маскировки в зависимости от ее уровня энергии. Результирующая маска (msi) определяется как набор максимумов из значений с совпадающими индексами.

При построении маски от шумовых компонент (mni) для каждой тональной компоненты определяется ее собственная критическая полоса и в пределах этой критической полосы определяется уровень маскирующего шума. Далее степень маскирования определяется согласно кривым маскировки, представленным в [3].

На вход процедуры объединения поступают следующие маски: маска порога слышимости и флюидных уровней (mti), маски тональных (msi) и шумовых (mni) компонент. Результирующая маска вычисляется по формуле (11):

$$m_i = \begin{cases} \max(mt_i, ms_i), & l_i = 1 \\ \max(mt_i, mn_i), & l_i = 0 \end{cases}, \quad (11)$$

где l_i – признак является ли i–тая компонента спектра тональной или шумовой.

Второй уровень психоакустической модели осуществляет перевод интенсивностей компонент спектра в соответствующие значения уровня воспринимаемой громкости. Для пересчета используется семейство кривых равной громкости [3]. По значению частоты и интенсивности компоненты спектра определяется пара кривых равной громкости, между которыми находится нормализуемое значение. Затем с помощью линейной интерполяции определяется соответствующее значение громкости в фонах.

Под различимой градацией понимается минимально заметное на слух изменение амплитуды сигнала. Частотная разрешающая способность слуха – не учитывается. Известно, что в зависимости от уровня громкости и частоты сигнала разрешающая способность слуха варьируется от 2 до 40%.

Общая громкость сигнала определяется как сумма максимальной громкости по всем компонентам спектра и 0.3 средней громкости по всем остальным компонентам спектра. Т.к. при расчете градаций используются воспринимаемые уровни громкости, вызов третьего уровня психоакустической модели возможен только после применения второго уровня.

Для рассчитанного уровня громкости интерполируется кривая амплитудной разрешающей способности. Для

каждой компоненты спектра определяется минимально–различимое изменение громкости и текущий уровень громкости компоненты спектра делится на найденное значение.

Вычисление оценки качества

Вычисление подобия выполняется раздельно для активной и неактивной фаз сигнала, причем сравниваются пары синхронизированных фрагментов. Сравнение (12) производится по энергиям в критических полосах, вычисленных по интегральным спектрам сигналам:

$$Q = 100 - \sum_{n=1}^{Np} \begin{cases} 100 \cdot Vc_n \min(1, \left| \frac{E_n^{src} - E_n^{tst}}{E_n^{src}} \right|), & E_n^{src} > E_{\min} \\ 0, & 8=0G5 \end{cases}. \quad (12)$$

Оценка для каждой фазы определяется как среднее по всем парам фрагментов.

Оценка для всего сигнала вычисляется как сумма взвешенных оценок активной и неактивной фаз.

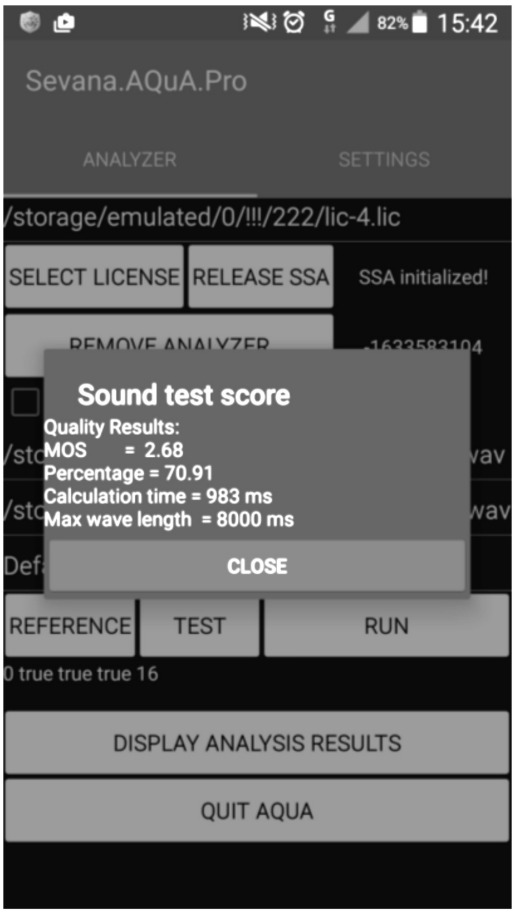

Рисунок 3. Основное окно приложения.

Программная реализация

Программа для мобильного устройства выполнена ввиде двух модулей: динамически загружаемой библиотеки оценки качества (нативный C++) и пользовательского интерфейса (JAVA).

В библиотеке предусмотрена работа как со звуковыми файлами, так и со звуковым потоком. Допускается создание нескольких экземпляров контекстов оценки качества, в которых оценка качества выполняется параллельно.

В GUI (Graphical User Interface) поддерживает только работу со звуковыми файлами. Реализован функционал выбора файлов, вызов библиотеки оценки качества, визуализация результатов оценки и настройка параметров алгоритма AQuA. На **рис. 3** представлено основное окно приложения.

Видно, что библиотека инициализирована, создан контекст оценки, выбраны эталонный и тестируемый файлы, и заданы настройки по умолчанию. После нажатия на кнопку "RUN" вызывается функционал библиотеки, вычисляющий оценку качества по двум звуковым файлам. Результаты оценки отображаются на экране как показано на **рис. 4**.

Программа предоставляет возможность просмотреть подробный отчет по результатам анализа звуковых файлов (**рис. 5**). Отчет отображается по нажатию кнопки "DISPLAY ANALYSIS RESULTS".

По значениям спектральных пар и интегральным спектрам сигнала могут быть построены графики, представленные на **рис. 6 и 7**.

Окно настроек параметров алгоритма AQuA представлено на **рис. 8**. При выборе параметра отображается окно ввода значения, в котором приводится его имя, краткое описание и диапазон допустимых значений (**рис. 9**).

Эксперимент

Тестирование мобильного приложения показало совпадение оценок качества получаемых в программах, разработанных для мобильных телефонов и персонального компьютера.

Для сравнения метода AQuA с рекомендацией ITU-T P.562 [5] была использована речевая база данных ITU-T для тестов кодеков [6].

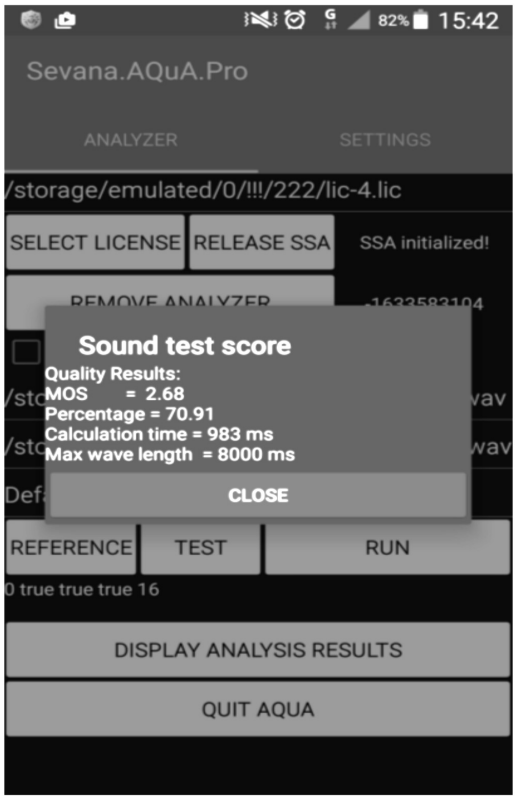

Рисунок 4. Отображение результатов оценки.

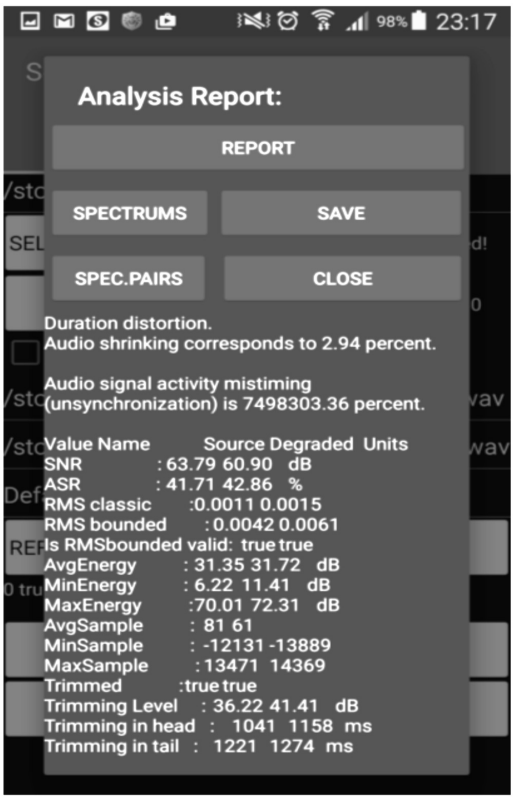

Рисунок 5. Отчет по результатам анализа.

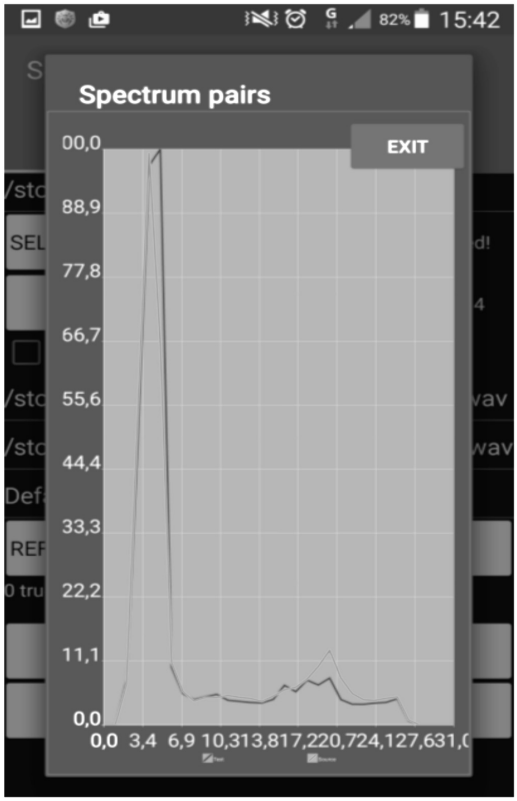

Рисунок 6. Отображение спектральных пар.

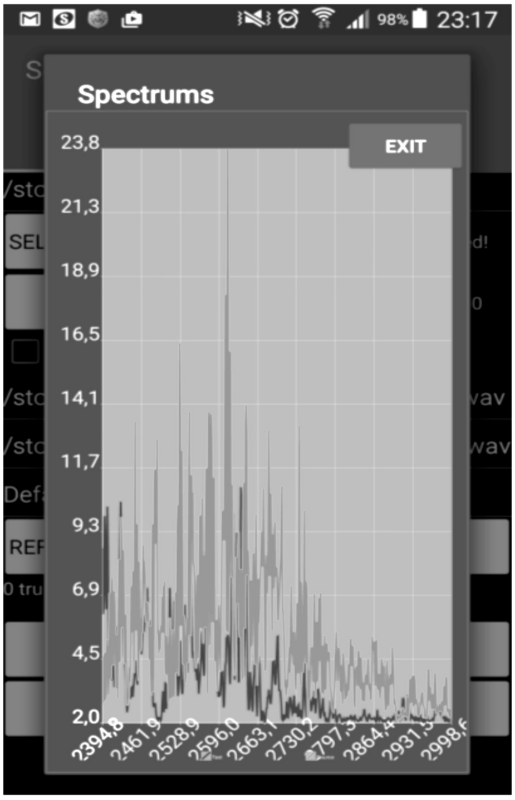

Рисунок 7. Отображение интегрального спектра.

В **табл. 2** приводятся суммы ошибок (модуль разности экспертного и вычисленного значения MOS) полученные в результате работы стандартного ПО и предложенного метода.

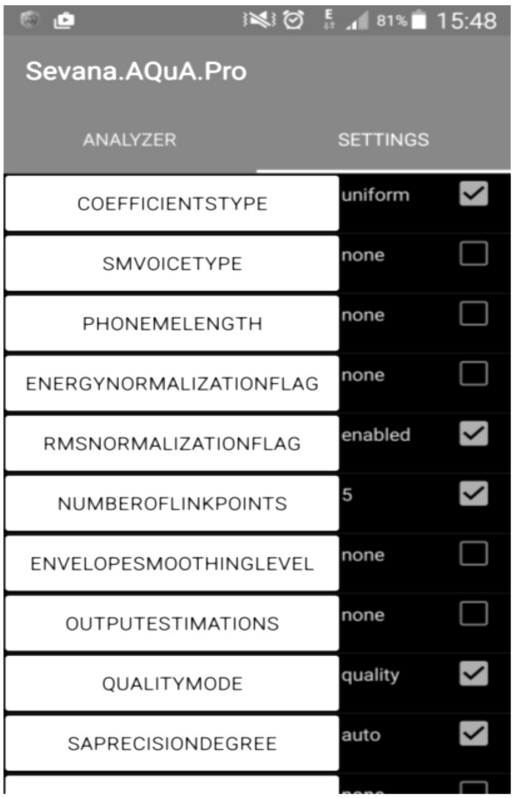

Рисунок 8. Окно настроек алгоритма AquA.

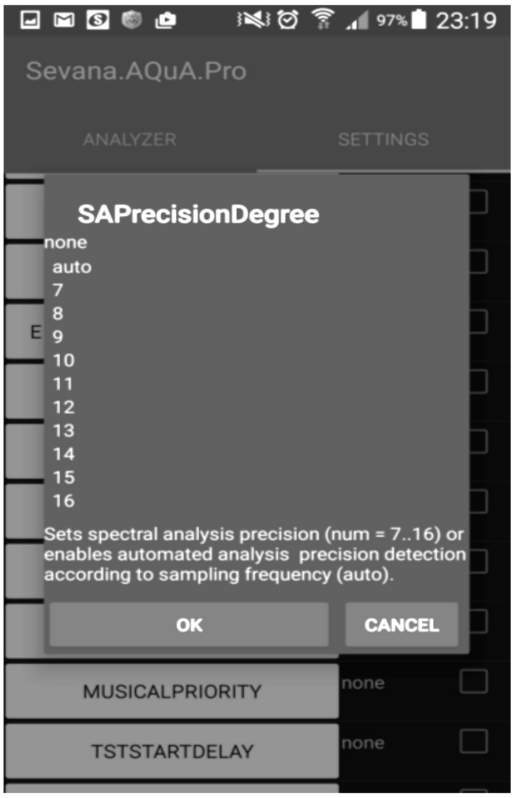

Рисунок 9. Окно ввода параметра "SAPrecisionDegree".

Видно, что в ряде случаев предложенный метод дает лучшую точность оценок.В данный момент ведутся работы по совершенствованию метода.

Известны результаты исследования [7], показывающего, что точность оценок алгоритмов PESQ и AQuA для GSM сетей связи совпадает, а в случае CDMA сетей точность оценок AQuA оказывается выше.

Таблица 1.

Сравнение алгоритмов ITU-T P.562 и AquA.

Язык	Сумма ошибок			
	PESQ-OS	MOS-LQO	MOS-WB-LQO	AQUA
Японский	105,75	92,40	59,31	90,43
Французский	66,32	59,20	80,03	64,27
Английский	51,02	50,74	135,92	59,65

ЛИТЕРАТУРА

1. Винницкий А. Рынок мобильных систем к 2016 году не изменится // электронный источник http://appleinsider.ru/analysis/rynok–mobilnyx–os–k–2016–godu–ne–izmenitsya.html / доступ 25.02.2016

2. Пат. 2312405 Российская Федерация, МПК G 10 L 19 / 02 (2006.01), G10L15/00 (2006.01). Способ осуществления машинной оценки качества звуковых сигналов, Гусев М. Н., Дегтярёв В.М., Жарков И.В.; заявитель и патен–тообладатель Гусев М.Н. – № 2005128572/09; заявл. 13.09.2005; опубл. 10.12.2007, Бюл. №34(ч.2) – 2с: ил.

3. Гусев М.Н. Расчет и измерение качества речевых сигналов/ Гусев М.Н., Дегтярев В.М. // Геликон Плюс, СПб., 2008, 275с

4. Пат. 2435232 Российская Федерация, МПКG 10 L 15 / 14 (2006.01). Способ машинной оценки качества передачи речи, Гусев М. Н.; заявитель и па–тентообладатель Гусев М. Н. – №2010133428/08; заявл. 09.08.2010; опубл. 27.11.2011, Бюл. №33 – 2с: ил.

5. Perceptual evaluation of speech quality (PESQ): An objective method for end–to–end speech quality assessment of narrow–band telephone networks and speech codecs / ITU–T Recommendation P.862 // Режим доступа: http://www.itu.int/rec/T–REC–P.862/en

6. ITU–T coded–speech database [Электронный ресурс] / Supplement 23 to ITU–T P–series Recommendations // Режим доступа: http://www.itu.int/rec/T–REC–P.Sup23–199802–I/en

7. Bruno Daniel M.L. Characterisation of noisy speech channels in 2G and 3G mobile networks // Master Thesis to obtain the degree of master at the Instituto Superior de Engenharia do Porto, 2013 / электронный источник http://www.sevana.fi/MSc_Thesis_–_Bruno_Daniel_Moreira_Leite_–_2013.rar