(19) Евразийское патентное ведомство

(11) 043719 (13) B1

# (12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
**2023.06.16**

(21) Номер заявки
**202293449**

(22) Дата подачи заявки
**2020.11.12**

(51) Int. Cl. ***G10L 25/30*** (2013.01)
***G10L 25/60*** (2013.01)
*G10L 21/0216* (2013.01)

## (54) СПОСОБ АВТОМАТИЧЕСКОЙ ОЦЕНКИ КАЧЕСТВА РЕЧЕВЫХ СИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ВЫБОРА КАНАЛА В МНОГОМИКРОФОННЫХ СИСТЕМАХ

(43) **2023.01.20**
(86) **PCT/RU2020/000600**
(87) **WO 2022/103290 2022.05.19**

(71)(73) Заявитель и патентовладелец:
**ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "ЦРТ-ИННОВАЦИИ" (RU)**

(72) Изобретатель:
**Волкова Марина Викторовна, Новосёлов Сергей Александрович, Лаврентьева Галина Михайловна, Анджукаев Церен Владимирович, Гусев Алексей Евгеньевич (RU)**

(74) Представитель:
**Нилова М.И. (RU)**

(56) ANDERSON R AVILA ET AL.: "Non-intrusive speech quality assessment using neural networks", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 16 March 2019 (2019-03-16), XP081154240, abstract, sections 2, 3 (last paragraph)
US-A1-2020082843
GB-A-2456296
J.URGEN TCHORZ ET AL.: "SNR Estimation Based on Amplitude Modulation Analysis With Applications to Noise Suppression", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 11, no. 3, 1 May 2003 (2003-05-01), XP011079710, ISSN: 1063-6676, DOI: 10.1109/TSA.2003.811542, Retrieved from the Internet: URL:https://ieeexplore.ieee.org/document/1208288, figure 1
US-B1-9972339

(57) Изобретение относится к области автоматической оценки качества речевого сигнала, в частности к способу обучения нейронных сетей оценивать отношение сигнал/шум, время реверберации, класс шума, присутствующего в записи, и давать общую оценку качества как функцию от указанных оценок на целом речевом сигнале или его фрагменте. Предложенный способ обучения нейронной сети оценивать характеристики качества входного речевого сигнала содержит этапы, на которых выполняют подготовку набора обучающих речевых сигналов, для каждого из которых известно отношение сигнал/шум, время реверберации и класс шума, выбираемый из заданного множества классов шумов, применяют к каждому обучающему речевому сигналу детектор речевой активности с выделением из указанного обучающего речевого сигнала обучающих признаков, выполняют обучение нейронной сети одновременно предсказывать на основе полученных обучающих признаков отношение сигнал/шум, время реверберации, класс шума и обобщенную оценку качества для входного речевого сигнала.

043719 B1

043719 B1

## Область техники

Настоящее изобретение относится к области автоматической оценки качества речевого сигнала, в частности к способу обучения нейронных сетей оценивать отношение сигнал/шум, время реверберации, класс шума, присутствующего в записи, и давать общую оценку качества как функцию от указанных оценок на целом речевом сигнале или его фрагменте.

Оценка качества речевого сигнала может использоваться в различных приложениях речевой обработки, например, для автоматического выбора наилучшего микрофона в многомикрофонной системе записи звуковых сигналов. В случае голосовой биометрии она может использоваться для определения наиболее качественных сегментов речи в записях, выполненных в различных акустических условиях, для построения голосовой модели диктора по выбранным фрагментам.

## Уровень техники

В области обработки сигналов для оценки того или иного искажения обычно используются количественные характеристики, такие как отношение сигнал/шум и время реверберации.

Отношение сигнал/шум (SNR, signal to noise ratio) выражается через отношение мощностей сигнала и шума и может быть представлено с использованием следующего математического выражения:

$$SNR = \frac{P_{\text{сигнал}}}{P_{\text{шум}}} = \left(\frac{A_{\text{сигнал}}}{A_{\text{шум}}}\right)^2$$

где SNR - отношение сигнал/шум,

$P_{\text{сигнал}}$ - средняя мощность сигнала,

$P_{\text{шум}}$ - средняя мощность шума,

$A_{\text{сигнал}}$ - среднеквадратическое значение амплитуды сигнала,

$A_{\text{шум}}$ - среднеквадратическое значение амплитуды шума.

Время реверберации (RT, reverberation time) является одним из основных параметров для описания акустической среды области пространства, в которой проводилась запись речи. Чаще всего, оно определяется как время, за которое уровень звукового давления уменьшается на 60 дБ (в 1 млн раз по мощности или в 1000 раз по звуковому давлению). В литературе время реверберации, определенное таким образом, обычно обозначается как RT60 или T60. Существуют устоявшиеся методы для определения RT60 по известной импульсной характеристике помещения, однако в реальных сценариях работы со звукозаписями, полученными из произвольных источников, импульсная характеристика не доступна. Поэтому становится актуальной задача приблизительной оценки времени реверберации по данной звукозаписи без дополнительных сведений об акустических условиях.

Наиболее известным способом оценки качества речевых сигналов является MOS-оценка (Mean Opinion Score), представляющая собой субъективно-статистические испытания с помощью группы слушателей-экспертов. Несмотря на достаточно высокую надежность такой оценки при большом количестве слушателей, она является самой ресурсозатратной. Среди объективных методов, не зависящих от экспертов, часто используются оценки, основанные на сравнении оригинального (эталонного) и кодированного (искаженного) сигналов. Например, перцептивная оценка качества речи (PESQ) и ее усовершенствованная версия - перцептивная объективная оценка качества восприятия речи (POLQA). Несмотря на широкую распространенность в телекоммуникационных системах, такие оценки ограничены в применении к речевой обработке в реальном времени, так как эталонный сигнал при этом недоступен.

Из уровня техники известны методы автоматической оценки качества сигнала, не требующие эталонного образца для сравнения. Такие решения используют методы машинного обучения для получения систем, как правило, нейронных сетей, предсказывающих MOS-оценку (CN 104581758 A, EP 3494575 A1) или оценки PESQ, POLQA (CN 108346434 A, CN 108322346 A).

Однако обобщенные оценки, полученные таким образом, тяжело интерпретировать в терминах количественных характеристик акустической среды, таких как отношение сигнал/шум и время реверберации. Знание этих параметров необходимо для предсказания условий работоспособности систем обработки речи: например, для построения голосовой модели диктора хорошими условиями считаются SNR ≥ 15 Дб, RT60 < 600 мс. Поэтому необходимо помимо обобщенной оценки качества сигнала иметь представление о количественных характеристиках среды, в которых проводилась запись.

В US 9396738 B2 рассматривается оценка количественных характеристик качества сигнала с помощью нейронных сетей, в частности, определяется отношение сигнал/шум (signal to noise), спектральная ясность (spectral clarity), коэффициент асимметрии (skew), коэффициент эксцесса (kurtosis) и средняя частота основного тона (pitch average). Однако основная цель данных измерений - оценка искаженного сигнала, переданного по сети, для телекоммуникационных приложений. Поэтому такие факторы акустической среды, как время реверберации, в данной оценке качества не рассматриваются.

Сам процесс вычисления времени реверберации сигнала чаще всего основан на использовании импульсной характеристики помещения, в котором была сделана запись (EP 2238590 A1, WO 2015010983). Очевидно, что такие методы можно использовать, только если импульсная характеристика помещения известна, что практически невозможно в реальных применениях автоматической речевой обработки.

При неизвестной импульсной характеристике помещения (также используется термин "слепая

оценка", blind estimation) оценка времени реверберации как правило сводится к классическим методам обработки сигналов. Например, в US 9558757 B1 предлагается определение скорости затухания звука через построение автокорреляционной функции интенсивности сигнала по времени. При этом обнаруживаются только значения времени реверберации, превосходящие определенный порог. К недостаткам такого подхода можно отнести низкую чувствительность к сигналам с малым временем реверберации, невозможность применения на коротких фрагментах речи и неустойчивость к зашумленным данным.

Способ оценки времени реверберации с использованием мультиканального микрофона, использующий глубокие нейронные сети, описан в US 20200082843 A1. Согласно данному способу, анализируют сигналы, полученные с помощью мультиканального микрофона. Однако применение данного способа для обработки речевых сигналов от одного микрофонного входа достаточно затруднительно.

Способ определения характеристик, отбора и адаптации обучающих акустических сигналов для автоматической системы распознавания речи известен из US 9922664 B2. Согласно данному способу осуществляют подготовку обучающих данных, имитирующих целевые условия окружающей среды, в том числе уровня шума и реверберации, которые впоследствии могут использоваться для обучения глубокой нейронной сети. Обученную нейронную сеть далее могут использовать для классификации образцов речевых данных для симулирования кодеков, соответствующих данным образцам речевых данных. Однако, данный способ не позволяет, в частности, осуществить обучение нейронной сети одновременно предсказывать или оценивать на основе выделенных обучающих данных отношение сигнал/шум, время реверберации, класс шума и обобщенную оценку качества для входного речевого сигнала.

Таким образом, существует потребность в способах обучения нейронных сетей осуществлять одновременную оценку характеристик входного речевого сигнала без использования дополнительных сведений об акустических условиях, при которых получен данный речевой сигнал.

**Раскрытие сущности изобретения**

Согласно одному варианту реализации настоящего изобретения предложен способ обучения нейронной сети оценивать характеристики качества входного речевого сигнала, согласно которому выполняют подготовку набора обучающих речевых сигналов, для каждого из которых известно отношение сигнал/шум, время реверберации и класс шума, выбираемый из заданного множества классов шумов, применяют к каждому обучающему речевому сигналу детектор речевой активности с выделением из указанного обучающего речевого сигнала обучающих признаков, выполняют обучение нейронной сети одновременно оценивать на основе выделенных обучающих признаков отношение сигнал/шум, время реверберации, класс шума и обобщенную оценку качества для входного речевого сигнала.

При использовании данного способа оценка качества сигнала может производиться только по данному входному сигналу, без необходимости сравнения с эталонным или неискаженным сигналом, а для оценки времени реверберации не требуется знать импульсную характеристику помещения, в котором была сделана запись речи.

Согласно одному примеру реализации подготовка набора обучающих речевых сигналов может включать в себя обеспечение множества чистых речевых сигналов, имеющих минимальное значение отношения сигнал/шум и времени реверберации, множества стационарных шумовых сигналов различных классов и множества импульсных характеристик, соответствующих различным помещениям, для которых известно время реверберации, применение к каждому чистому речевому сигналу операции свертки с импульсной характеристикой из указанного множества импульсных характеристик с получением множества реверберированных сигналов, сложение полученных реверберированных сигналов со стационарными шумовыми сигналами различных классов с получением множества искаженных зашумлением сигналов, имеющих различное отношение сигнал/шум, формирование итогового набора обучающих речевых сигналов из искаженных зашумлением сигналов, сбалансированный по отношению сигнал/шум, времени реверберации и классу шума, а также подсчет интегральной оценки качества для каждого искаженного зашумлением сигнала как функции от отношения сигнал/шум и времени реверберации.

Согласно еще одному примеру реализации каждый шумовой сигнал может быть подвергнут реверберации с использованием импульсной характеристики того же помещения, которая была выбрана для реверберации соответствующего чистого речевого сигнала. Кроме того, каждый шумовой сигнал может быть подвергнут реверберации с использованием импульсной характеристики помещения, отличной от импульсной характеристики, которая была выбрана для реверберации соответствующего чистого речевого сигнала.

Согласно еще одному примеру реализации предложен способ, содержащий этап использования регрессионной модели предиктора, обучаемой с использованием стоимостной функции на основе среднеквадратической ошибки для оценки отношения сигнал/шум, времени реверберации и обобщенной оценки качества.

Оценка класса шума может быть выполнена с использованием классификатора, обученного с помощью бинарной кросс-энтропии.

Согласно еще одному варианту реализации предложен способ автоматического выбора канала в многомикрофонной системе с использованием нейронной сети, обученной на основе обучающих признаков, согласно которому получают входные речевые сигналы из множества каналов многомикрофонной

системы, применяют детектор речевой активности к каждому входному речевому сигналу с выделением из него признаков, характеризующих входной речевой сигнал и соответствующих обучающим признакам, подают выделенные признаки, характеризующие входной речевой сигнал, на вход нейронной сети и осуществляют их одновременную оценку, получают с выхода нейронной сети оцененные значения отношения сигнал/шум, времени реверберации, обобщенной оценки качества и предсказанного класса шума для каждого входного речевого сигнала, и выбирают канал из множества каналов многомикрофонной системы, из которого был получен входной речевой сигнал, имеющий оцененные значения, удовлетворяющие заданному условию.

При этом заданное условие может представлять собой максимальное значение обобщенной оценки качества.

**Краткое описание чертежей**

Предложенное изобретение далее описано более подробно со ссылкой на прилагаемые фигуры чертежей, среди которых:

фиг. 1 - последовательность действий, обеспечивающая получение обучающего множества;

фиг. 2 - схема получения оценок качества речевого сигнала из исходной аудиозаписи с помощью обученной модели.

**Осуществление изобретения**

Согласно одному варианту предложен способ обучения нейронных сетей оценивать характеристики качества входного речевого сигнала. Согласно способу выполняют подготовку набора обучающих речевых сигналов, для каждого из которых известно отношение сигнал/шум, время реверберации и класс шума, выбираемый из заданного множества классов шумов. Далее, согласно способу к каждому обучающему речевому сигналу применяют детектор речевой активности с выделением из указанного обучающего речевого сигнала обучающих признаков, и выполняют обучение нейронной сети одновременно оценивать на основе выделенных обучающих признаков отношение сигнал/шум, время реверберации, класс шума и обобщенную оценку качества для входного речевого сигнала.

Как показано на фиг. 1, на первом этапе подготовки обучающего набора данных берут множество чистых речевых сигналов 101, имеющих минимальное значение сигнал/шум и время реверберации, множество стационарных шумовых сигналов 102 различных классов, множество импульсных характеристик 103, соответствующих различным помещениям, время реверберации (T60) для которых известно. При этом в качестве источника чистых речевых сигналов и стационарных шумовых сигналов могут быть использованы существующие базы речи и шумов. Требуемые импульсные характеристики могут быть сгенерированы с использованием специальной утилиты.

Согласно одному варианту реализации, в качестве стационарных шумовых сигналов использовалась база из 79 классов шумов, таких как шум клавиатуры, дождя, гул толпы людей, производственных станков и т.д. В качестве множества импульсных характеристик использовалась специально сгенерированная база импульсных характеристик, которая имитировала 40000 комнат различных размеров с временем реверберации от 0 до 2 с, причем для каждой комнаты были сгенерированы по 4 импульсные характеристики, имитирующие различные положения источника звука внутри этой комнаты.

На втором этапе подготовки обучающего набора данных к каждому чистому речевому сигналу применяют операцию свертки с импульсной характеристикой произвольно выбранной комнаты с получением множества реверберированных речевых сигналов 104.

Каждой комнате может соответствовать несколько импульсных характеристик в зависимости от положения источника звука в этом помещении.

В дополнительном примере реализации каждый шумовой сигнал также может быть подвергнут реверберации 105. При этом реверберация может быть осуществлена с использованием импульсной характеристики того же помещения, которая была выбрана для реверберации соответствующего чистого речевого сигнала. Если импульсных характеристик в помещении несколько, может быть использована импульсная характеристика, отличная от используемой для реверберации чистого речевого сигнала. Это позволяет сымитировать различные положения источников речи и шума в пространстве и создать более реалистичную базу.

На третьем этапе подготовки обучающего набора данных выполняют сложение каждого реверберированного сигнала из множества реверберированных сигналов, полученных на предыдущем этапе, с стационарными шумовыми сигналами различных классов с получением в результате множества зашумленных сигналов 106 с различными значениями отношения сигнал/шум.

В одном из вариантов реализации, для получения корректного отношения сигнал/шум мощность речевого сигнала рассчитывается только на речевых сегментах без учета пауз, для этого к речевому сигналу применяется детектор речевой активности (voice activity detector, VAD).

На четвертом этапе подготовки обучающего набора данных формируют итоговое сбалансированное обучающее и тестовое множество 107 из подготовленных речевых сигналов, искаженных зашумлением и реверберацией. При этом для каждого из таких сигналов известны параметры SNR, RT60 и класс шума.

На пятом этапе осуществляют подсчет обобщенной или интегральной оценки качества (QE) для каждого искаженного реверберацией и зашумлением речевого сигнала как некоторой функции от парамет-

ров искажения SNR и RT60. Согласно частному случаю реализации, вычисление QE осуществляют с помощью следующих математических выражений:

$$S_{SNR} = \frac{1}{1+e^{-0.25\cdot(SNR_{\text{дБ}}-15)}},$$

$$S_{RT60} = \frac{1}{1+e^{0.0125\cdot(RT60_{\text{мс}}-600)}},$$

$$OQ = \sqrt{S_{SNR}\cdot S_{RT60}},$$

где $S_{SNR}$ - оценка уровня SNR речевого сегмента,

$S_{RT60}$ - оценка уровня реверберации речевого сегмента,

OQ - интегральная оценка качества речевого сегмента,

$SNR_{\text{дБ}}$ - значение SNR речевого сегмента в децибелах,

$RT60_{\text{мс}}$ - значение времени реверберации в миллисекундах для речевого сегмента.

При подготовке обучающего набора данных также может быть использован существующий набор данных, который удовлетворяет условию сбалансированности по диапазонам SNR и T60, а также по классам шумов.

Далее выполняют выделение обучающих признаков из подготовленных речевых сигналов. В частности, применяют к подготовленным речевым сигналам детектор речевой активности (VAD) с целью выделения и сохранения только речевых участков. Далее, осуществляют выделение из полученных сигналов обучающих признаков, например, мел-частотных кепстральных коэффициентов (MFCC) или банка (набора) полосовых фильтров (FBANK) или других признаков, характеризующих аудиосигнал, известных из уровня техники. Полученные обучающие признаки подвергают дальнейшей обработке, например, путем применения нормализации кепстрального среднего (cepstral mean normalization, CMN).

Выделенные обучающие признаки далее используют для обучения сверточной нейронной сети в режиме многозадачности. Сверточную нейронную сеть обучают одновременно оценивать отношение сигнал/шум (SNR), время реверберации (RT60), класс шума и обобщенную оценку качества (OQ) на входных признаках, полученных на предыдущем этапе. Это осуществляется благодаря использованию четырех выходов в архитектуре нейронной сети и комбинированной функции потерь (loss function), основанной на суммировании четырех стоимостных функций с различными весовыми коэффициентами.

В одном из вариантов реализации, для автоматической оценки SNR, RT60 и OQ используют регрессионную модель предиктора, обучаемую с использованием стоимостной функции на основе среднеквадратической ошибки (MSE). Автоматическая оценка класса шума может быть основана на использовании классификатора, обучаемого с помощью бинарной кросс-энтропии (BCE).

В качестве неограничивающего примера, ниже приведена формула для расчета комбинированной функции потерь:

$$L = 10\cdot MSE(OQ) + 0{,}001\cdot MSE(RT60_{\text{мс}}) + MSE\left(SNR_{\text{дБ}}\right) + 10\cdot BCE(\text{классшума})$$

где L - комбинированная функция потерь,

MSE (OQ) - функция потерь на основе среднеквадратической ошибки для интегральной оценки качества,

MSE ($RT60_{\text{мс}}$) - функция потерь на основе среднеквадратической ошибки для оценки RT60,

MSE ($SNR_{\text{дБ}}$) - функция потерь на основе среднеквадратической ошибки для оценки SNR,

BCE (классшума) - бинарная кросс-энтропийная функция потерь для классификации шума.

Поскольку предполагается, что разрабатываемая нелинейная модель предсказания качества речевого сигнала должна оценивать качество на коротких фрагментах речи (от 1 до 2 с), обучение модели в одном из вариантов реализации также происходит на коротких речевых фрагментах. В реальности человеческая речь и естественные шумы не являются строго стационарными. Это значит, что глобальное значение отношения сигнал/шум, полученное на этапе подготовки данных и единое для целого файла, должно быть скорректировано для каждого короткого сегмента этого файла. В качестве неограничивающего примера ниже приведена формула для подсчета локального отношения сигнал/шум:

$$SNR_{local} = 10log(\frac{\alpha^2 E_{stc}^{rev}}{\beta^2 E_{noise}^{rev}}),$$

где $E_{src}^{rev}$ - энергия реверберированного речевого сигнала до зашумления и $E_{noise}^{rev}$ - энергия реверберированного шума. Коэффициенты $\alpha$ и $\beta$ для каждого сигнала находятся с помощью решения системы линейных уравнений по четырем фрагментам сигнала:

$$X_{aug}(i) = \alpha X_{src}^{rev}(i) + \beta X_{noise}^{rev}(i)\ for\ \ i \in \{1,\dots,n\},$$

где $X_{aug}(i)$ - i-й фрагмент аугментированного сигнала, $X_{src}^{rev}(i)$ и $X_{noise}^{rev}(i)$ - его реверберированные речевая и шумовая части.

В качестве неограничивающего примера ниже приведена архитектура нейронной сети, которая может быть использована для оценки характеристики качества речевого сигнала.

Остаточная сеть ResNet18 состоит из 8 ResNet блоков, каждый из которых образован двумя сверточными слоями с 64 фильтрами размера 3×3 и связью с пропуском соединения (skip connection) через два слоя. Данная связь осуществляется при помощи простого поэлементного сложения входа блока и

выхода последнего слоя блока, если размерности совпадают, либо с применением операции свертки для согласования размерностей.

Верхний уровень образован слоем глобального усреднения (global average pooling layer), 512-мерный выход которого может быть назван "вектором качества" (quality embedding). Этот вектор подается затем на три линейных слоя: для предсказания отношения сигнал/шум (SNR), времени реверберации (RT60) и оценки качества (OQ). Для оценки качества используется функция активации сигмоида.

Для классификации шума используют дополнительный двухслойный классификатор с функцией активации softmax или ее модификациями по количеству классов шумов (в рассматриваемом варианте - 79).

Таким образом, предложенный способ обучения нейронной сети позволяет получить как обобщенную оценку качества речевого сигнала, так и его конкретные акустические характеристики (отношение сигнал/шум и время реверберации), что может быть использовано как для приложений голосовой биометрии, так и для выбора лучшего канала в многомикрофонных системах согласно заданному критерию.

Согласно еще одному варианту осуществления предложен способ автоматического выбора канала в многомикрофонной системе, осуществляемый с использованием описанной выше обученной нейронной сети.

Согласно данному способу из множества каналов многомикрофонной системы получают входные речевые сигналы. Далее, как показано на фиг. 2, к каждому отдельному речевому сигналу 201 применяют детектор речевой активности 202 с выделением признаков, характеризующих данный входной речевой сигнал и соответствующих обучающим признакам, которые были использованы для обучения нейронной сети, например, мел-частотных кепстральных коэффициентов (MFCC) или банка (набора) полосовых фильтров (FBANK) или других признаков, характеризующих аудиосигнал, известных из уровня техники.

Далее подают полученные признаки 203, характеризующие входной речевой сигнал, на вход нейронной сети 204 и осуществляют их одновременную оценку с получением на выходе нейронной сети оцененных значения отношения сигнал/шум 205, времени реверберации 206, обобщенной оценки качества 207 и предсказанного класса шума 208 для каждого входного речевого сигнала. На основе указанных оцененных значений выбирают канал из множества каналов многомикрофонной системы, из которого был получен входной речевой сигнал, имеющий оцененные значения, удовлетворяющие заданному условию. При этом в качестве заданного условия может быть использовано максимальное значение обобщенной оценки качества. Однако специалисту будет понятно, что в качестве заданного условия могут быть использованы другие показатели, известные в уровне техники.

Настоящее изобретение не ограничено конкретными вариантами реализации, раскрытыми в описании в иллюстративных целях, и охватывает все возможные модификации и альтернативы на каждом этапе осуществления.

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ обучения нейронной сети оценивать характеристики качества входного речевого сигнала на локальных сегментах входного речевого сигнала длиной до двух секунд, согласно которому

выполняют подготовку набора обучающих речевых сигналов, для каждого из которых известно отношение сигнал/шум на каждом локальном сегменте входного речевого сигнала, время реверберации и класс шума, выбираемый из заданного множества классов шумов,

получают речевые сегменты посредством применения к каждому обучающему речевому сигналу детектора речевой активности для удаления из речевого сигнала сегментов, не содержащих речь, затем из полученных речевых сегментов выделяют обучающие признаки,

выполняют обучение нейронной сети одновременно оценивать на основе выделенных обучающих признаков отношение сигнал/шум, время реверберации, класс шума и обобщенную оценку качества, представляющую собой функцию от отношения сигнал/шум и времени реверберации, для входного речевого сигнала.

2. Способ по п.1, согласно которому подготовка набора обучающих речевых сигналов включает в себя

обеспечение множества чистых речевых сигналов, имеющих минимальные значения отношения сигнал/шум и времени реверберации, множества стационарных шумовых сигналов различных классов и множества импульсных характеристик, соответствующих различным помещениям, для которых известно время реверберации,

применение к каждому чистому речевому сигналу операции свертки с импульсной характеристикой из указанного множества импульсных характеристик с получением множества реверберированных сигналов,

сложение полученных реверберированных сигналов с стационарными шумовыми сигналами различных классов с получением множества искаженных зашумлением сигналов, имеющих различное отношение сигнал/шум,

формирование итогового набора обучающих речевых сигналов из искаженных зашумлением сигна-

лов, сбалансированный по отношению сигнал/шум, времени реверберации и классу шума,

подсчет значения обобщенной оценки качества для каждого искаженного зашумлением сигнала как функции от отношения сигнал/шум и времени реверберации.

3. Способ по п.2, согласно которому каждый шумовой сигнал подвергают реверберации с использованием импульсной характеристики того же помещения, которая была выбрана для реверберации соответствующего чистого речевого сигнала.

4. Способ по п.2, согласно которому каждый шумовой сигнал подвергают реверберации с использованием импульсной характеристики помещения, отличной от импульсной характеристики, которая была выбрана для реверберации соответствующего чистого речевого сигнала.

5. Способ по п.1, согласно которому для оценки отношения сигнал/шум, времени реверберации и обобщенной оценки качества используют регрессионную модель предиктора, обучаемую с использованием стоимостной функции на основе среднеквадратической ошибки.

6. Способ по п.1, согласно которому оценку класса шума выполняют с использованием классификатора, обученного с помощью бинарной кросс-энтропии.

7. Способ автоматического выбора канала в многомикрофонной системе с использованием нейронной сети, обученной с использованием способа обучения нейронной сети по п.1, согласно которому

получают входные речевые сигналы из множества каналов многомикрофонной системы,

получают речевые сегменты посредством применения детектора речевой активности к каждому входному речевому сигналу для удаления сегментов, не содержащих речь, затем из полученных речевых сегментов выделяют обучающие признаки,

подают выделенные обучающие признаки на вход нейронной сети и осуществляют их одновременную оценку,

получают с выхода нейронной сети оцененные значения отношения сигнал/шум, времени реверберации, предсказанное значение обобщенной оценки качества и предсказанный класс шума для каждого сегмента входного речевого сигнала, и

выбирают канал из множества каналов многомикрофонной системы, из которого был получен входной речевой сигнал, имеющий оцененные значения, удовлетворяющие заданному условию.

8. Способ по п.7, в котором заданное условие представляет собой максимальное значение обобщенной оценки качества.

9. Способ обучения нейронной сети оценивать характеристики качества входного речевого сигнала, в том числе на локальных сегментах входного речевого сигнала длиной до двух секунд, согласно которому

выполняют подготовку набора обучающих речевых сигналов, для каждого из которых известно отношение сигнал/шум на каждом локальном сегменте входного речевого сигнала, время реверберации и класс шума, выбираемый из заданного множества классов шумов,

получают речевые сегменты посредством применения к каждому обучающему речевому сигналу детектора речевой активности для удаления сегментов, не содержащих речь, затем из полученных речевых сегментов выделяют обучающие признаки,

выполняют обучение нейронной сети одновременно предсказывать на основе полученных обучающих признаков отношение сигнал/шум, время реверберации, класс шума и обобщенную оценку качества, представляющую собой функцию от отношения сигнал/шум и времени реверберации, для входного речевого сигнала,

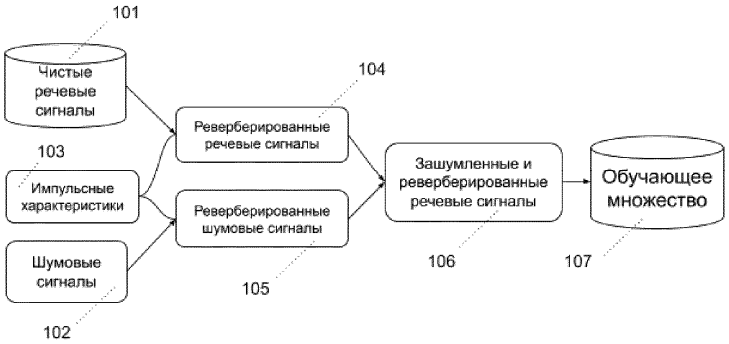
причем при обучении нейронной сети на локальных сегментах входного речевого сигнала длиной до двух секунд корректируют локальное значение отношения сигнал-шум на каждом локальном сегменте входного речевого сигнала, при этом подсчет локального отношения сигнал/шум осуществляется по формуле:

$$SNR_{local} = 10log(\frac{\alpha^2 E_{stc}^{rev}}{\beta^2 E_{noise}^{rev}}),$$

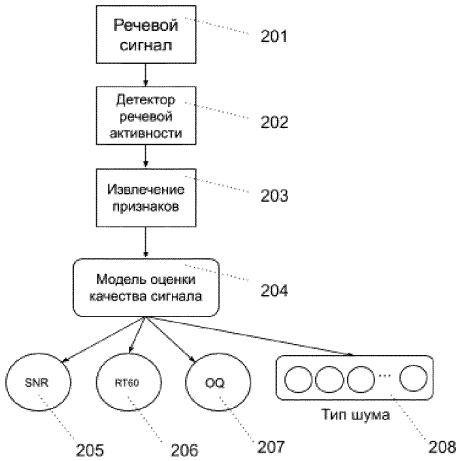
где $E_{src}^{rev}$ - энергия реверберированного речевого сигнала до зашумления и $E_{noise}^{rev}$ - энергия реверберированного шума, коэффициенты α и β для каждого сигнала находятся с помощью решения системы линейных уравнений по четырем фрагментам сигнала:

$$X_{aug}(i) = \alpha X_{src}^{rev}(i) + \beta X_{noise}^{rev}(i)\ for\ \ i \in \{1, \ldots, n\},$$

где $X_{aug}(i)$ - i-й фрагмент аугментированного сигнала, $X_{src}^{rev}(i)$ и $X_{noise}^{rev}(i)$ - его реверберированные речевая и шумовая части.


101
Чистые речевые сигналы
104
Реверберированные речевые сигналы
103
Импульсные характеристики
Зашумленные и реверберированные речевые сигналы
Обучающее множество
Реверберированные шумовые сигналы
Шумовые сигналы
106
107
105
102

Фиг. 1


Речевой сигнал
201
Детектор речевой активности
202
Извлечение признаков
203
Модель оценки качества сигнала
204
SNR
RT60
OQ
Тип шума
205
206
207
208

Фиг. 2