

УДК 004.934.2

Д.И. Новохрестова, Е.Ю. Костюченко, И.А. Ходашинский

Алгоритм и методика количественной оценки схожести речевых сигналов

Предложено решение задачи автоматизированной количественной оценки произношения слогов в рамках оценки качества речи в речевой реабилитации. Представлен алгоритм количественной оценки схожести двух звуковых сигналов различной длины с использованием гибридной меры совпадения. Гибридная мера основана на расчете трех метрик (DTW-расстояние, коэффициент корреляции и метрика Минковского) и нечетком классификаторе в качестве механизма комбинации рассчитанных значений по метрикам. Среднее количество совпадений оценок, получаемых с использованием предложенного алгоритма, и оценок по ранее применяемой методике составляет 83%. Предложенная методика оценки схожести речевых сигналов с использованием нескольких опорных сигналов позволяет учитывать вариативность речи и индивидуальные особенности произношения фонем. Учет особенностей возможен при использовании в качестве опорных сигналов предоперационной речи пациента, проходящего речевую реабилитацию.

Ключевые слова: алгоритм оценки схожести, оценка качества речи, речевая реабилитация.

DOI: 10.21293/1818-0442-2022-25-3-45-51

Увеличение количества впервые диагностированных случаев онкологических заболеваний органов речевого тракта [1, 2] привело к необходимости автоматизации процесса оценки качества речи. Оценка речи является неотъемлемой частью речевой реабилитации, которая требуется после хирургического этапа комбинированного лечения рака полости рта и ротоглотки. При этом сам процесс реабилитации и оценки должен иметь персонализированный подход согласно концепции развития здравоохранения России и национальному проекту «Здравоохранение».

Методы оценки качества речи

Глобально методы оценки качества речи можно разделить на две группы: объективные и субъективные. Под объективными в большинстве случаев понимается расчет количественного показателя через оценку значений параметров речевого сигнала, под субъективными – оценка с использованием мнений экспертов (аудиторов).

Самыми известными и часто используемыми объективными методами оценки являются подходы, основанные на сравнении исходного сигнала и этого же сигнала после проведения манипуляций над ним (передача по каналам связи, обработка и др.), или оценки параметров речевой составляющей/шума внутри сигнала. К этим методам можно отнести PESQ [3, 4], PSQM [3], E-model [5] и метод, основанный на оценке соотношения сигнал/шум (SNR и segSNR) [6]. Относительно решаемой задачи эти методы обладают существенными недостатками. В случае сравнения с исходным сигналом подразумевается, что и исходный, и искаженный сигнал является одной и той же реализацией некоторого набора фонетических единиц. В случае расчета параметров сигнала не учитывается понятность и корректность произносимой речи.

Если говорить про субъективные оценки, то эталонным методом при передаче по каналам связи является метод MOS [7]. Этот метод подразумевает

экспертное оценивание, при этом предъявляются строгие требования к группе аудиторов (экспертов).

Ранее применяемая в речевой реабилитации методика основана на экспертной оценке произношения слогов по ГОСТ Р 50840–95 «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости» [8]. Стандарт описывает методику оценки слоговой разборчивости группой аудиторов. Реальное применение стандарта в речевой реабилитации имеет следующие недостатки: невозможность сбора полной группы аудиторов (по стандарту – от пяти человек, в практике – один эксперт-логопед), список произносимых слогов не адаптирован под особенности изменения речи постоперационных пациентов, список слогов быстро запоминается экспертом, оценка каждого слога по отдельности является бинарной (1 – слог произнесен полностью корректно, 0 – иначе). Поэтому была сформулирована задача разработки алгоритма и методики количественной оценки схожести речевых сигналов. Алгоритм позволяет оценить схожесть двух речевых сигналов различных длин на основе гибридной меры расчета, а учитывающая особенности речи диктора методика – оценить произношение слога с использованием нескольких опорных сигналов.

Цель работы – повысить эффективность оценки речи в речевой реабилитации за счет автоматизации процесса получения количественной оценки качества речи и учета индивидуальных особенностей пациента.

Гибридная мера расчета

В алгоритме количественной оценки схожести двух речевых сигналов предложено использовать гибридную меру расчета. Гибридная мера представляет собой комбинацию трех метрик: DTW-расстояния, коэффициента корреляции и метрики Минковского. В качестве механизма комбинирования используется нечеткий классификатор.

Использованные метрики и анализ результатов их расчета по отдельности описаны в [9]. Однако

при таком подходе к их использованию возникает задача ручного подбора порогового значения для бинаризации оценок, что приводит к невозможности полной автоматизации процесса получения оценки. Поэтому предложено использовать комбинацию метрик, где итоговое значение оценки является выходом нечеткого классификатора [10]. На вход нечеткому классификатору подаются три значения, рассчитанные по вышеуказанным метрикам.

Для обучения классификатора использовались результаты оценки аудиозаписей произношения слогов реальными пациентами, прошедшими речевую реабилитацию в НИИ онкологии г. Томска. Эксперимент и используемый нечеткий классификатор описаны в [11]. Для каждой группы фонем был обучен отдельный классификатор, точности классификации для тестовых и обучающих выборок приведены в табл. 1. Точности были получены на 10-кратной кроссвалидации. Средняя точность для тестовых выборок составила 0,83. Полученные значения точностей на тестовых выборках являются точечными оценками относительной частоты случаев правильно выставленной метки класса относительно общего объема тестовой выборки. В табл. 1 представлены доверительные интервалы для полученных значений согласно интервальной оценке неизвестной вероятности по относительной частоте [12] (заданный уровень надежности 0,95).

Таблица 1
Точности классификации для групп фонем

Группа фонем	Тестовая выборка	Обучающая выборка
Группа [к] и [к']	0,87 ± 0,021	0,88
Группа [с] и [с']	0,84 ± 0,022	0,85
Группа [т] и [т']	0,79 ± 0,025	0,81

Алгоритм количественной оценки схожести

На вход алгоритму подаются два аудиосигнала с одинаковыми характеристиками самих сигналов (частота дискретизации, формат аудиофайла). Аудиосигналы содержат разные реализации одного и того же слога. Длины аудиосигналов могут быть (и в большинстве случаев являются) различными. Также на вход подается указание на проблемную фонему в записанном слоге для применения построенного классификатора или выбора метрики DTW в качестве итоговой.

С учетом подобранных параметров итоговый алгоритм представляет собой следующую последовательность шагов:

Шаг 1. Преобразование сигналов в последовательность значений. Оба сигнала преобразуются в массив значений амплитуд сигналов – числовые последовательности.

Шаг 2. Временная нормализация двух числовых последовательностей – приведение последовательностей к единой длине. Этот шаг выполняется с использованием алгоритма динамической трансформации временной шкалы – DTW-алгоритм [13, 14].

Шаг 3. Фиксируется одна из количественных оценок – DTW-расстояние. Это значение стоимости

пути между двумя последовательностями – последний элемент матрицы преобразований.

Шаг 4. На основе матрицы преобразований и построенного на шаге 2 оптимального пути составляются трансформированные числовые последовательности.

Шаг 5. Нахождение количественных оценок между трансформированными числовыми последовательностями с использованием двух метрик: коэффициента корреляции и метрики Минковского (параметр $p = 3$).

Шаг 6. Для выбранной проблемной фонемы выбирается нечеткий классификатор, обученный для данной группы проблемных фонем. Если фонема не входит в список выделенных проблемных фонем, то делается указание на это.

Шаг 7. Итоговая количественная оценка находится как выход нечеткого классификатора (на вход подаются значения DTW-расстояния, коэффициента корреляции и метрики Минковского) в случае оценки слога с одной из выделенных проблемных фонем или как DTW-расстояние иначе.

Шаг 8. Вывод итоговой оценки.

На рис. 1 приведено графическое представление алгоритма в виде NS-диаграммы [15].

Методика оценки схожести речевых сигналов на основе нескольких опорных

Предложенный выше алгоритм позволяет количественно оценить схожесть двух звуковых сигналов при их представлении в виде числовых последовательностей. Однако такую оценку, представляющую собой положительное действительное число, невозможно однозначно интерпретировать в контексте оценки схожести речи к эталонной (в рамках решаемой задачи – предоперационной) речи конкретного пациента. Предлагаемая методика оценки схожести звуковых сигналов за счет применения нескольких опорных сигналов позволяет получить интерпретируемую оценку схожести произношения в заранее известном интервале значений. Использование нескольких опорных сигналов позволяет учитывать вариативность речи до операции. По предлагаемой методике можно оценить подобие звуковых сигналов, представляющих одинаковую по содержанию фонетическую единицу речи – слог. То есть можно сравнивать между собой записи только одного и того же слога.

Примем следующие обозначения:

- эталонный сеанс – сеанс записи речи пациента, проводимый до проведения хирургического лечения, речь близка к нормальной;
- эталонный набор записей/аудиосигналов – записанные в процессе эталонного сеанса оценки речи аудиосигналы;
- эталонная запись – одна запись из эталонного набора;
- оцениваемый сеанс – сеанс записи речи пациента на одном из этапов речевой реабилитации, речь пациента может быть как искажена, так и близка к нормальной (зависит от времени проведения сеанса и состояния речевого аппарата пациента);

Получить числовую последовательность 1 из аудиофайла речевого сигнала 1			
Получить числовую последовательность 2 из аудиофайла речевого сигнала 2			
Применить DTW-алгоритм к числовым последовательностям 1 и 2			
Запомнить значение Y - DTW-расстояние (значение последнего элемента матрицы преобразований)			
Получить трансформированные числовые последовательности 1 и 2 согласно матрице преобразований и кратчайшему пути			
Вычислить значение Z – коэффициент корреляции Пирсона между трансформированными числовыми последовательностями 1 и 2			
Вычислить значение W – расстояние по метрике Минковского между трансформированными числовыми последовательностями 1 и 2			
Выбрать проблемную фонему			
[к] или [к']	[с] или [с']	[т] или [т']	Иначе
Получить C как выход нечеткого классификатора FC1 (вход – Y, Z, W)	Получить C как выход нечеткого классификатора FC2 (вход – Y, Z, W)	Получить C как выход нечеткого классификатора FC3 (вход – Y, Z, W)	$C = Y$
Вывести количественную оценку C			

Рис. 1. Алгоритм нахождения количественной оценки схожести двух речевых сигналов различной длины

– оцениваемый набор записей – записанные в процессе оцениваемого сеанса оценки речи аудио-сигналы, оценка которого должна быть получена по предложенной методике;

– оцениваемая запись – одна запись из оцениваемого набора.

Опорными сигналами служат эталонные записи из нескольких эталонных сеансов. С точки зрения теоретического решения поставленной задачи количество опорных сигналов ограничено только вычислительными возможностями, большее количество опорных сигналов позволит учитывать большую вариативность речи. С практической точки зрения в речевой реабилитации рекомендуемым количеством опорных сигналов, а следовательно, и эталонных сеансов, являются два сигнала и два сеанса соответственно. Это количество позволяет учитывать вариативность речи, в то же время запись такого количества сеансов возможна в процессе проведения предоперационной подготовки. Рекомендовано проводить запись эталонных сеансов в разные дни или с перерывом в несколько часов в течение одного дня. Это способствует уменьшению нагрузки на органы речевого аппарата.

Методика заключается в выполнении следующей последовательности действий:

1. Выбор нескольких эталонных записей из нескольких эталонных сеансов и оцениваемой записи из оцениваемого сеанса. Все записи представляют собой реализацию одного и того же слога.

2. Нахождение количественных оценок между парами эталонных записей по алгоритму, описанному ранее.

3. Нахождение среднего значения количественных оценок, полученных в пункте 2.

4. Нахождение количественных оценок между каждой из эталонных записей и оцениваемой записью по алгоритму.

5. Нахождение среднего значения количественных оценок, полученных в пункте 4.

6. Нахождение отношения среднего значения оценок между эталонными и оцениваемым сигналами к среднему значению оценок между эталонными сигналами – относительная оценка оцениваемого сигнала.

7. Вывод о схожести оцениваемого сигнала и эталонных по относительной оценке.

Относительная оценка заключена в интервал от 0 до 1 (включая обе границы), и чем качественнее (более похоже на эталон) произнесен слог, тем ближе результат к 1.

Само отношение средних оценок теоретически может быть больше 1, это означает, что оцениваемая речь максимально схожа с эталонной и речь пациента не нуждается в речевой реабилитации с точки зрения анализа количественной оценки схожести речевых сигналов. В таком случае окончательное решение о необходимости проведения реабилитации принимается лечащим врачом – логопедом. Если отношение в процессе вычисления получается больше 1, то, согласно алгоритму расчета, относительная оценка принимается равной 1.

Относительная оценка может принимать значение 1 при условии полного восстановления речи до предоперационного уровня. Если рассматривать с точки зрения наличия предпосылок, то все из них

говорят об ухудшении речи пациента после проведения оперативного вмешательства и невозможности достижения предоперационного уровня. Проведенное оперативное вмешательство (зачастую резекция одного из органов речевого аппарата) приводит к невозможности полного восстановления правильного произношения всех проблемных фонем. Анализ аудиозаписей пациентов НИИ онкологии г. Томска до и после хирургического лечения, в процессе и после речевой реабилитации подтверждает описанные предпосылки.

Методика оценки схожести речевых сигналов при наличии двух эталонных (опорных) сигналов может быть представлена в виде следующего алгоритма:

1. Выбрать два эталонных сеанса ε_1 и ε_2 (опорные сеансы).

2. Выбрать сеанс для оценки (оцениваемый сеанс) o .

Для первого слога в оцениваемом сеансе:

3. Найти оценку слогу по алгоритму для пар $o-\varepsilon_1$, $o-\varepsilon_2$, $\varepsilon_1-\varepsilon_2$: C_{o,ε_1} , C_{o,ε_2} , $C_{\varepsilon_1,\varepsilon_2}$.

4. Найти относительную оценку сигнала C_o :

4.1. $C_{o,\varepsilon} = (C_{o,\varepsilon_2} + C_{o,\varepsilon_1}) / 2$ – среднее значение оценок в парах (оцениваемая-эталонная).

4.2. Относительная оценка $C_o = C_{o,\varepsilon} / C_{\varepsilon_1,\varepsilon_2}$.

Оценка C_o в интервале $[0;1]$. Чем ближе к 1, тем лучше.

5. Повторить шаги 3–4 для остальных слогов в сеансе.

6. Найти оценку сеанса как среднее арифметическое оценок слогов, входящих в оцениваемый сеанс.

Описанная последовательность действий в виде NS-диаграммы представлена на рис. 2.

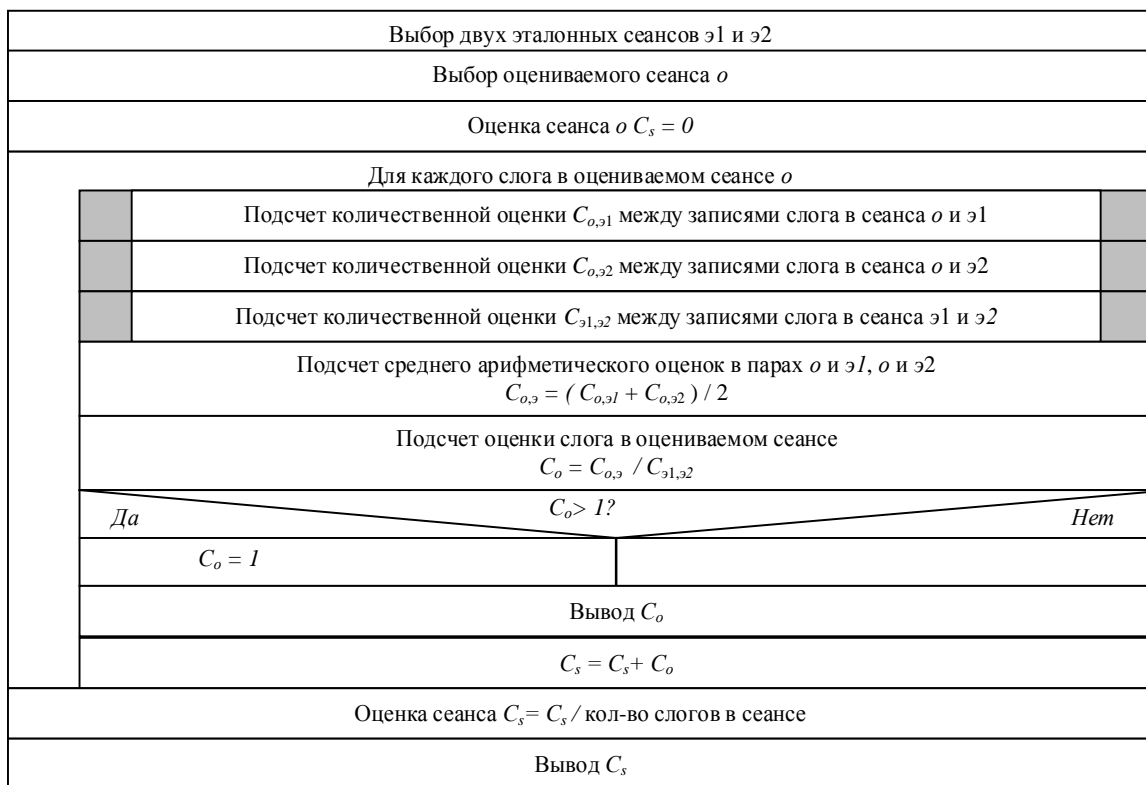


Рис. 2. Методика оценки при двух эталонных сеансах

Сравнение оценок по методике с использованием эталонных сеансов от разных дикторов

Использование в качестве эталонных сигналов записей речи того же пациента (диктора), чью речь необходимо оценить, позволяет учитывать особенности речи пациента, в то время как использование в качестве эталона записей другого человека даже с идеальным произношением приводит к существенному занижению оценки качества произношения слогов.

Было рассчитано, как изменяется оценка сеанса в зависимости от набора эталонных сеансов и их принадлежности тому или иному диктору. Были выбраны сеансы, в которых речь диктора близка к иде-

альной, т.е. оценка по старой методике большинства записей в сеансе была бы равна единице (допускалось не более пяти возможных «нулей» в оценках). В табл. 2 приведен пример сравнения оценок. В сеансах S1 и S2 диктора D1 и сеансах S4 и S5 диктора D2 все слоги были произнесены правильно (оценки произношения равны единицам), в сеансе S3 диктора D1 было допущено четыре ошибки (в слогах была неправильно произнесена проблемная фонема, остальные фонемы – правильно).

Из данных таблицы видно, что использование в качестве эталона сеансов другого диктора приводит к занижению оценок произношения слогов.

Для установления и проверки направленности изменений получаемых оценок были получены

оценки сеансов четырех пациентов. Для каждого из пациентов была проведена оценка на основе сравнения с сеансами этого же пациента и с сеансами других пациентов/дикторов. Проверка производилась с использованием критерия Вилкоксона [12].

Таблица 2
Сравнение оценок сеанса с использованием эталонных сеансов разных дикторов

Оцениваемый сеанс	Эталонные сеансы			
	D1			D2
	S1 + S2	S1 + S3	S2 + S3	S4 + S5
S1	–	–	0,716	0,252
S2	–	0,982	–	0,296
S3	0,992	–	–	0,383

Для подсчета значения критерия были получены разности значений оценок сеанса, а также проведена ранжировка полученных значений разности. Наблюдаемое значение критерия Вилкоксона представляет собой сумму рангов нетипичных сдвигов. Все сдвиги (разности) значений оценок имеют одинаковую направленность, т.е. при использовании в качестве эталонов сеансов этого же пациента значения оценки больше, чем в случае использования сеансов другого диктора. Поэтому наблюдаемое значение критерия Вилкоксона равно 0. Критическое значение для объема выборки $n = 36$ и уровня значимости 0,01 равно 185. На основе полученных значений была принята нулевая гипотеза о том, что оценки сеансов пациента, полученные с использованием эталонных сеансов этого же пациента, превосходят оценки, получаемые с использованием эталонных сеансов другого пациента. Поэтому использование в качестве эталона предоперационной речи пациента в предложенной методике позволяет учитывать индивидуальные особенности речи пациента.

Заключение

Предложенный в работе алгоритм с применением временной нормализации и гибридной меры совпадения предназначен для количественной оценки схожести двух речевых сигналов. Используемая в алгоритме гибридная мера оценки схожести основана на расчете DTW-расстояния, коэффициента корреляции и расстояния Минковского и применении нечеткого классификатора, на вход которого подаются рассчитанные значения. Для каждой группы проблемных фонем были обучены классификаторы на основе оценок аудиозаписей пациентов НИИ онкологии. Среднее количество совпадений оценок, получаемых по алгоритму, и оценок по ранее используемой методике составило 83%.

Предложенная методика оценки схожести речевых сигналов с применением нескольких опорных сигналов позволяет учитывать вариативность произношения слогов и особенности речи дикторов (пациентов). Получаемые по методике оценки записей слогов и средние оценки сеансов представляют собой интерпретируемые оценки в заранее известном интервале. В [16] описано практическое применение и внедрение предложенных методик в процесс

речевой реабилитации, проводимой после комбинированного лечения онкологических заболеваний органов полости рта и ротоглотки на базе действующего медицинского учреждения – НИИ онкологии г. Томска.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90082.

Литература

1. Злокачественные новообразования в России в 2020 году (заболеваемость и смертность) // под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. – М.: МНИОИ им. П.А. Герцена (филиал ФГБУ «НМИЦ радиологии» Минздрава России), 2021. – 252 с.
2. Состояние онкологической помощи населению России в 2020 году // под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. – М.: МНИОИ им. П.А. Герцена (филиал ФГБУ «НМИЦ радиологии» Минздрава России), 2021. – 239 с.
3. ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-P.862>, свободный (дата обращения: 17.06.2022).
4. Rix A.W. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs // 2001 Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing. – 2001. – No. 2. – P. 749–752.
5. Recommendation G.107 (06/15): The E-model: a computational model for use in transmission planning [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-G.107-201506-I/en>, свободный (дата обращения: 17.06.2022).
6. Способы оценки субъективного качества речи [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/post/177099/>, свободный (дата обращения: 17.06.2022).
7. Streijl R.C. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives / R.C. Streijl, S. Winkler, D.S. Hands // Multimedia Systems. – 2016. – Vol. 22, No. 2. – P. 213–227.
8. ГОСТ Р 50840–95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости [Электронный ресурс]. – Режим доступа: <http://docs.cntd.ru/document/1200027288>, свободный (дата обращения: 17.06.2022).
9. Experimental Analysis of Expert and Quantitative Estimates of Syllable Recordings in the Process of Speech Rehabilitation / D. Novokhrestova, E. Kostuchenko, I. Hodashinsky, L. Balatskaya // Lecture Notes in Computer Science. – 2021. – Vol. 12997 LNAI. – P. 483–491.
10. Корушев Н.П. Алгоритм формирования базы правил нечеткого классификатора на основе алгоритма кластеризации k-средних и метаэвристического алгоритма «китов» / Н.П. Корушев, И.А. Ходашинский // Доклады ТУСУР. – 2021. – Т. 24, № 1. – С. 42–47.
11. Новохрестова Д.И. Гибридная мера для сравнения записей слогов // Сб. избр. статей научной сессии ТУСУР: в 3 т. – Томск: В-Спектр, 2022. – Т. 2. – С. 75–79.
12. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов. – 10-е изд., стереотипное. – М.: Высшая школа, 2004. – 479 с.
13. Романенко А.А. Выравнивание временных рядов: прогнозирование с использованием DTW [Электронный ресурс]. – Режим доступа: <http://jmla.org/papers/doc/2011/>

no1/Romanenko2011Dynamic.pdf, свободный (дата обращения: 17.06.2022).

14. Salvador S. Fastdtw: Toward accurate dynamic time warping in linear time and space / S. Salvador, P. Chan // Workshop on Mining Temporal and Sequential Data, Seattle. – 2004. – P. 70–80.

15. Nassi I. Flowchart Techniques for Structured Programming / I. Nassi, B. Shneiderman // ACM SIGPLAN Notices. – 1973. – Vol. 8, No. 8. – P. 12–26.

16. Восстановление речевой функции у больных раком полости рта и ротоглотки с использованием инновационных технологий / Л.Н. Балацкая, Е.Л. Чойзонов, Е.А. Красавина, Е.Ю. Костюченко, Д.И. Новохрестова // Вопросы онкологии. – 2020. – Т. 66, № 3. – С. 247–251.

Новохрестова Дарья Игоревна

Аспирант каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР) Ленина пр-т, 40, г. Томск, Россия, 634050
ORCID ID: 0000-0002-4931-1681
Тел.: +7 (382-2) 70-15-29
Эл. почта: ndi@fb.tusur.ru

Костюченко Евгений Юрьевич

Канд. техн. наук, доцент, доцент каф. КИБЭВС ТУСУРа Ленина пр-т, 40, г. Томск, Россия, 634050
ORCID ID: 0000-0001-8000-2716
Тел.: +7 (382-2) 70-15-29
Эл. почта: key@keva.tusur.ru

Ходашинский Илья Александрович

Д-р техн. наук, профессор каф. КИБЭВС ТУСУРа Ленина пр-т, 40, г. Томск, Россия, 634050
ORCID ID: 0000-0002-9355-7638
Тел.: +7 (382-2) 70-15-29
Эл. почта: hia@fb.tusur.ru

Novokhrestova D.I., Kostyuchenko E.Y., Hodashinsky I.A.
Algorithm and method for quantitative assessment of the speech signals similarity

The paper proposes a method to solve the task of automated quantitative assessment of the syllable pronunciation. This quantitative assessment is used to evaluate the speech quality during speech rehabilitation. An algorithm for quantifying the similarity of two audio signals of different lengths is presented. The algorithm uses a hybrid match measure. The hybrid measure is based on calculation of three metrics (DTW-distance, correlation coefficient and Minkowski metric) and using a fuzzy classifier as a mechanism for combining the calculated values. The average number of coincidences of estimates by the proposed algorithm and estimates by the previously applied method is 83%. A method for quantifying the similarity of speech signals using several reference signals is proposed. The method allows to consider the variability of speech and the individual characteristics of the phoneme's pronunciation. This is achieved by using records of the patient's preoperative speech as reference signals.

Keywords: similarity assessment algorithm, speech quality assessment, speech rehabilitation.

doi: 10.21293/1818-0442-2022-25-3-45-51

References

1. *Zlokachestvennyye novoobrazovaniya v Rossii v 2020 godu (zabolevaemost' i smertnost')* [Malignant neoplasms in Russia in 2020 (morbidity and mortality)]. Moscow, National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, 2021. 252 p. (in Russ.)
2. *Sostoyanie onkologicheskoi pomoshchi naseleniyu Rossii v 2020 godu* [The state of oncological care for the population of Russia in 2020]. Moscow, National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, 2021, 239 p. (in Russ.)
3. ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Available at: <https://www.itu.int/rec/T-REC-P.862> (Accessed: June 17, 2022).
4. Rix A.W. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. *2001 Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, no. 2, pp. 749–752.
5. Recommendation G.107 (06/15): The E-model: a computational model for use in transmission planning. Available at: <https://www.itu.int/rec/T-REC-G.107-201506-I/en> (Accessed: June 17, 2022).
6. *Sposoby otsenki sub'ektivnogo kachestva rechi* [Methods for assessing the subjective quality of speech]. Available at: <https://habrahabr.ru/post/177099/> (Accessed: June 17, 2022).
7. Streijl R.C., Winkler S., Hands D.S. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems*, 2016, vol. 22, no. 2, pp. 213–227.
8. GOST R 50840–95. Speech transmission over varies communication channels. Techniques for measurements of speech quality, intelligibility and voice identification. Available at: <http://docs.cntd.ru/document/1200027288> (Accessed: June 17, 2022)
9. Novokhrestova D., Kostuchenko E., Hodashinsky I., Balatskaya L. Experimental Analysis of Expert and Quantitative Estimates of Syllable Recordings in the Process of Speech Rehabilitation. *Lecture Notes in Computer Science*, 2021, vol. 12997 LNAI, pp. 483–491.
10. Koryshev N.P., Hodashinsky I.A. Algorithm to forming a rule base for a fuzzy classifier designed on the basis of the K-means clustering algorithm and the whale optimization algorithm. *Proceedings of TUSUR University*, 2021, vol. 24, no. 1, pp. 42–47. DOI: 10.21293/1818-0442-2021-24-1-42-47 (in Russ.)
11. Novokhrestova D. I. *Gibridnaya mera dlya sravneniya zapisei slogov* [Hybrid measure for comparing syllable records]. *Sbornik izbrannykh statei nauchnoi sessii TUSUR* [Collection of selected articles of TUSUR Scientific Session]: in 3 part. Tomsk: V-Spektr, 2022 – Томск: В-Спектр, 2022, part. 2, pp. 75–79 (in Russ.).
12. Gmurman V.E. *Teoriya veroyatnostei i matematicheskaya statistika: Uchebnoe posobie dlya vuzov* [Probability Theory and Mathematical Statistics: Textbook for Universities]. Moscow, Vysshaya shkola Publ., 2004, 479 p.
13. Romanenko A.A. *Sposoby otsenki sub'ektivnogo kachestva rechi* [Time Series Flattening: Forecasting with DTW]. Available at: <http://jmla.org/papers/doc/2011/no1/Romanenko2011Dynamic.pdf> (Accessed: June 17, 2022).
14. Salvador S., Chan P. Fastdtw: Toward accurate dynamic time warping in linear time and space. *Workshop on Mining Temporal and Sequential Data*, Seattle, 2004, pp. 70–80.
15. Nassi I., Shneiderman B. Flowchart Techniques for Structured Programming. *ACM SIGPLAN Notices*, 1973, vol. 8, no. 8, pp. 12–26.

16. Balatskaya L.N., Choinzonov E.L., Krasavina E.A., Kostyuchenko E.Y., Novokhrestova D.I. [Voice restoration in patients with oral cavity and oropharyngeal cancers using innovative technologies]. *Problems in Oncology*, 2020, vol. 66, no. 3, pp. 247–251. DOI: 10.37469/0507-3758-2020-66-3-247-251 (in Russ.).

Evgeny Y. Kostyuchenko

Candidate of Science in Engineering, Assistant Professor, KIBEVS TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID ID: 0000-0001-8000-2716
Phone: +7 (382-2) 70-15-29
Email: key@keva.tusur.ru

Darya I. Novokhrestova

Postgraduate student, Department of Complex Information Security of Electronic Computer Systems (KIBEVS), Tomsk State University of Control Systems and Radioelectronics (TUSUR)
40, Lenin pr., Tomsk, Russia, 634050
ORCID ID: 0000-0002-4931-1681
Phone: +7 (382-2) 70-15-29
Email: ndi@fb.tusur.ru

Илья А. Ходашинский

Doctor of Science in Engineering, Professor, KIBEVS TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID ID: 0000-0002-9355-7638
Phone: +7 (382-2) 70-15-29
Email: hia@fb.tusur.ru