OSTIS-2014
(Open Semantic Technologies for Intelligent Systems)

УДК 004.934

ПРИМЕНЕНИЕ МЕТОДА СЕМАНТИЧЕСКОГО ДИФФЕРЕНЦИАЛА ДЛЯ ОЦЕНКИ ПОКАЗАТЕЛЕЙ КАЧЕСТВА КОНВЕРСИИ ГОЛОСА

Захарьев В.А., Петровский А.А.

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь

zahariev@bsuir.by
palex@bsuir.by

В докладе рассмотрены возможности применения метода семантического дифференциала для оценки показателей качества конверсии голоса. Стандартные подходы, основанные на субъективных методах оценок MOS и ABX, в первую очередь предназначены для оценки степени разборчивости речи, и не могут быть напрямую применены для определения степени близости голосов дикторов. Представлена методика проведения оценки качества конверсии на основе построения семантического пространства слушателей.
Ключевые слова: семантический дифференциал, психолингвистика, конверсия голоса.

Введение

Речевой интерфейс является одной из форм организации взаимодействия с интеллектуальной системой. По эффективности и удобству данный способ является оптимальным для пользователя, поскольку сама по себе речь, как канал коммуникации между людьми, является наиболее естественной формой еѐ организации. Поэтому в данный момент вопросам реализации речевых интерфейсов в рамках исследования и разработки интеллектуальных систем уделяется немалое внимание.

Важное место в структуре речевого интерфейса интеллектуальной системы занимает подсистема синтеза речи по тексту, обеспечивающая реализацию обратной связи от системы к человеку. Развитие такого типа систем на настоящем этапе достигло значительного прогресса. Перед ними ставятся задачи не просто обеспечения заданных показателей разборчивости синтезируемого речевого сигнала, но и предъявляются требования по натуральности речи, наличию широкого спектра просодических шаблонов, поддержки множества языков синтеза и различных голосов дикторов. Последний аспект является наиболее важным и интересным для исследования, особенно в контексте перехода от систем синтеза к системам клонирования голоса, конкретного диктора [Лобанов, 2008]. Для решения задачи по реализации свойств мультимодальности для синтезатора речи применяются различные технологии обработки исходного речевого сигнала, в том числе конверсия голоса.

Конверсия голоса является технологией обработки речевого сигнала, позволяющей реализовать процесс трансформации параметров голоса, характеризующих речь исходного диктора, в параметры целевого. Объектами конверсии голоса, как технологии обработки сигналов, являются стабильные во времени свойства говорящего, проявляющиеся в речевом сигнале через изменение его акустических параметров [Stylianou, 2009].

Необходимо отметить, что очень важным этапом при разработке и реализации такой технологии является оценка результатов качества конверсии. Аудиторная степень “близости” или соответствия сконвертированного голоса голосу целевого диктора, положенная в основу многих методов оценки качества конверсии является очень обобщенной и нечѐткой метрикой, к тому же сильно субъективированной. В работе предлагается рассмотреть возможность использования метода семантических дифференциалов для построения более объективной методики оценки качества конверсии голоса.

1. Стандартные методы оценки показателей качества конверсии голоса

Для оценки качества конверсии голоса в системах конверсии используются два подхода на

основе объективных и субъективных методов оценок. Суть объективных методов заключаются в определении степени несоответствия акустических параметров сконвертированного голоса голосу целевого диктора. Другими словами, на определении расстояния между двумя векторами в пространствах акустических признаков дикторов. Например, на евклидовой метрике расстояния между параметрами результирующей и целевой спектральной огибающей закодированной с помощью коэффициентов линейного предсказания. Недостатком данного вида методов является невысокая корреляция полученных оценок с субъективными представлениями о качестве результирующего речевого сигнала для конкретного слушателя. Второй подход опирается на использование субъективных методов оценки, учитывающих мнения реальных субъектов выступающих в роли экспертов.

1.1. Метод на основе средней оценки мнений

Наиболее широко используемая методика субъективной оценки качества речевого сигнала описана в Рекомендации ITU-T P .800 и известна как методика средней оценки мнений (Mean Opinion Score – MOS), изначально применяющаяся для оценки качества передачи в телефонных сетях [ITU-T, 1996]. В соответствии с MOS качество речи, получаемое при прохождении сигнала от источника через систему связи к приемнику, оценивается как арифметическое среднее от всех оценок, выставляемых экспертами после прослушивания речевого материала. Количественные результаты этих тестов отображают усредненное качество, уровень усилий слушателя, разборчивость, естественность звучания. Экспертные оценки определяются по пяти балльной шкале: 5 – отлично, 4 – хорошо, 3 – приемлемо, 2 – плохо, 1 – неприемлемо. В случае оценки качества конверсии голоса экспертам предоставляется записи сконвертированной и целевой фонограмм, и предлагается кроме основных оценок, определить меру соответствие голосов дикторов в речевом материале по дополнительной шкале “узнаваемости”.

1.2. Метод на основе слепого тестирования

Вторым распространѐнным субъективным методом определения оценок качества конверсии голоса, является подход на основе слепого тестирования или так называемый ABX-тест [Clark, 1982]. В области аудио так обычно называют метод определения слышимой разницы между двумя звуковыми фрагментами. Метод наиболее эффективен для определения потенциальных различий находящихся вблизи порога слышимости. Главным преимуществом теста является то, что эксперт не знает какой именно фрагмент (сконвертированной речи или речи целевого диктора) воспроизводится в данный момент. Другой важный момент – возможность многократно повторения теста, что значительно уменьшает влияние случайности. Производится тест следующим образом. Эксперт присваивает кнопке A один фрагмент, кнопке B - второй, а программа проведения тестирования случайным образом присваивает кнопке X один из них (эксперту неизвестно, какой именно). Далее, посредством нажатий на кнопки, эксперт может в любой последовательности неограниченно слушать A, B и X, после чего должен определить чему соответствует X: A или B. Затем тест повторяется. По мере выполнения каждого прохода, программа подсчитывает вероятность "заблуждения" эксперта, т.е. вероятность того что он не услышит разницу между фрагментами речи содержащими голос сконвертированный системой и голос целевого диктора. Например, если эксперт правильно ответил при первом проходе, вероятность будет 50 %, если и второй проход даст такой же результат – 25%, и т.д. Положительным (эксперт действительно слышит различия) результат считается после как минимум 13-ти правильных ответов в 16-ти проходах.

Оба из перечисленных методов успешно и широко используются для оценки качества систем конверсии голоса. Однако, они также не лишены определенных недостатков. Легко видеть, что данные методы не были разработаны специально для проведения тестирования систем конверсии голоса. Так, например, тест MOS использовался для оценки качества речи передающейся по различным каналам связи, а тест ABX является общей методикой слепого тестирования применяемого во многих областях статистических исследований. Следствием этого является не самая высокая степень адекватности оценок получаемых по данным методам, метрики которых не отражают в полной степени характеристики голоса говорящего с точки зрения аудиторов. Они опираются на такие общие плохо определѐнные, размытые и сильно интегрированные категории оценок отраженные в таких шкалах как “разборчивость”, “узнаваемость”, “соответствие”, слабо связанных с особенностями восприятия экспертами личностных черт голоса диктора. Еще одной проблемой является неустойчивость данных оценок, которая выражается в их вариабельности во времени в зависимости от различных условий и речевого материала. Например, для одной пары дикторов для различных фраз данные оценки могут быть различными. Для построения более обоснованного пространства оценок предлагается рассмотреть возможность применения метода семантического дифференциала.

2. Оценка на базе метода семантического дифференциала

2.1. Метод семантического дифференциала

Метод семантического дифференциала (СД) принадлежит к методам экспериментальной семантики и является одним из методов построения субъективных семантических пространств. Техника

СД предназначалась для измерения различий в интерпретации понятий испытуемыми. Этот метод был разработан в 1952 г. группой американских психологов во главе с Ч. Осгудом в ходе исследования механизмов синестезии [Osgood, 1952]. Как полагает Ч. Осгуд, метод СД позволяет измерять так называемое коннотативное значение – те состояния, которые следуют за восприятием символа-раздражителя и необходимо предшествуют осмысленным операциям с символами, тем самым позволяя выявить общую меру, на основе которой выносятся человеческие оценки. В методе СД измеряемые объекты (понятия, изображения) оцениваются по ряду биполярных градуированных (трех-, пяти-, семибалльных) шкал, полюса которых заданы с помощью вербальных антонимов. Оценки понятий по разным шкалам, вообще говоря, коррелируют друг с другом, и с помощью факторного анализа удается выделить пучки таких высококоррелирующих шкал, сгруппировать их в факторы. Психологическим механизмом, объясняющим взаимосвязь и группировку шкал в факторы, Осгуд считал явление синестезии – психологический феномен восприятия, заключающийся в том, что раздражение одного органа чувств, вызывает специфические ощущения в другом органе чувств, незадействованном в данном конкретном случае, например, переживание цветового образа в ответ на музыкальную фразу в цветомузыке. Механизмы синестезии признаются основой метафорических переносов в высказываниях типа “бархатный голос”, “кислая физиономия” и т.п.

Переход от признаков, заданных шкалами, к факторам фактически является построением семантического пространства. Фактор можно рассматривать как смысловой инвариант содержания входящих в него шкал, и в этом смысле факторы являются формой обобщения прилагательных-антонимов, на базе которых строится СД, а его факторная структура отражает структуру антонимии в лексике. Группировка шкал в факторы позволяет перейти от описания объектов с помощью признаков, заданных шкалами, к более емкому описанию с помощью меньшего набора категорий-факторов, представив содержание объекта в виде совокупности факторов (многочлен) с различными коэффициентами веса. Нагрузки объекта по каждому из выделенных факторов определяются как среднее арифметическое оценок объекта по шкалам, входящим в этот фактор.

Семантическое пространство является своеобразным метаязыком описания значений, позволяющим путем разложения их содержания в фиксированном алфавите категорий-факторов проводить семантический анализ этих значений, выносить суждения об их сходстве и различии путем вычисления расстояний между соответствующими значениям координатными точками в пространстве. Традиция использования метода семантического дифференциала, как перспективного метода изучения восприятия слуховых сигналов, возникла в отечественной науке еще в начале 70-х годов по инициативе В.И.Галунова, и продолжена в работах В.Х.Манерова, В.Ф. Петренко и др [Галунов, 1978., Петренко, 1988].

2.2. Методика оценки на базе метода семантического дифференциала

Для оценки показателей качества конверсии голоса методика использования семантического дифференциала представляется совокупностью следующих основных шагов:

Шаг 1. Для группы экспертов путем поочерёдного прослушивания фонограмм исходного и целевого диктора производится формирование и тестирование списка прилагательных, положенных в основу шкал сравнения. От выбранных признаков зависит тот уровень осознанности, на котором аудитор будет оценивать измеряемый объект. Обязательное условие – они не должны содержать прямых характеристик объекта (голос – мужской, тембр – низкий), а обязаны иметь ассоциативные характеристики (голос – мягкий, тембр – звонкий). Возможно применение уже апробированного набора шкал. За основу факторного пространства предлагается взять предложенной в работе [Манеров, 2012] полученных при изучении восприятия мужских и женских голосов, протестированную на большом наборе из представителей разных социальных слоев, обладателей различной речевой культуры, обеспечивавшей большую репрезентативность выборки. Более подробно шкалы, сгруппированные по факторам представлены на шаге 2.

Шаг 2. Далее экспертам предлагается прослушать два разбитых, на несколько блоков, набора фраз, озвученных целевым диктором и сконвертированных системой. Система установок аудитора по отношению к значимым для него характеристикам голоса обнаруживается в его оценочных суждениях, которые классифицируются сознанием по схеме логических дихотомий. Пары противоположных эмоционально-оценочных прилагательных, распределенных по факторам представлены ниже

- Фактор силы (мужественный – женственный, басовитый – тонкий, сильный – слабый);
- Фактор активности (бодрый – вялый, активный – пасивный, быстрый – медленный);
- Оценка (теплый – холодный, хороший – плохой, нежный – грубый, мягкий – жесткий);
- Ненормативность (аспект оценки, связанный с качеством тембра и артикуляции: звучный – сиплый, звонкий – глухой, нехриплый – хриплый, четкий – расплывчатый, красивый – некрасивый);

Таким образом, даже при наличии в методике большого числа денотативных шкал была получена четкая структура, близкая к Осгудовской. Природа четвертого фактора получается смешанной, в нем присутствует и эстетическая оценка, и

характеристики денотативной природы, связанные речевой культурой, качеством звучания и артикуляции. Оценки усредняются по количеству блоков в тестовой выборке и количеству экспертов.

Шаг 3. Производится математическая обработка полученной матрицы данных. Для этого используется процедура корреляционного и факторного анализа, позволяющая выявить латентные (скрытые) критерии оценивания, в которые складываются первоначальные шкалы. Поскольку нами были взяты за основу уже разработанные шкалы, факторы по которым они группируются, для удобства сразу были отражены на предыдущем шаге. В результате, производится выделение факторов, формирующих семантическое пространство голосов слушателей-экспертов.

Шаг 4. С помощью методов линейной алгебры определяется расстояния между прослушанными голосами в семантическом пространстве и на основе его величины делается вывод о качестве конверсии голоса. Степень качества конверсии голоса, производимая той или иной системой определяется на основе метрики расстояния в данном семантическом пространстве, и является величиной обратно пропорциональной данному расстоянию.

Заключение

В докладе рассмотрены вопросы создания методики оценки на основе применения метода семантического дифференциала для определеиия показателей качества конверсии голоса. Стандартные подходы, основанные на субъективных методах оценок MOS и ABX, не всегда полностью удовлетворяют требованиям по адекватности оценок сравнения различного рода систем конверсии. Поскольку методики, которые они предлагают, а также категории оценок, не учитывающие всего феноменологического богатства смысловых, эмоциональных, мотивационных, осознаваемых и неосознаваемых составляющих субъективного образа – голоса диктора, будут, по крайней мере, неполными. При решении данной проблемы необходимо глубокое проникновение в смысловое пространство слушателей, причем не только как на уровне отдельных экспертов, но и на уровне их группы. Предложенная методика позволяет прийти к формализации представления о свойствах голосов и факторов, которые могли бы наиболее точно отразить степень восприятия близости одного диктора к другому, тем самым позволяя определить более объективную оценку соответствия целевого и сконвертированных голоса. Применение метода семантического дифференциала даѐт возможность перейти от непосредственного сравнения голосов дикторов, к сравнению, условно выражаясь, инвариантного представления экспертов о самих дикторах в их семантическом пространстве.

Предложенная методика является еще одним шагом на пути устранения границ в областях междисциплинарных исследований на основе подходов искусственного интеллекта. Подтверждая возможности использования психосемантики для решения вопросов связанных с оценкой качества обработки сигналов.

Библиографический список

[**Лобанов, 2008**] Компьютерный синтез и распознавание речи / Б. М. Лобанов, Л. И. Цирульник — Минск : Белорусская наука, 2008.— 344 с.

[**Stylianou, 2009**] Voice conversion: a survey / Y. Stylianou // Proc. of International Conference on Acoustics, Speech and Signal Processing. — Taipei, 2009. — P. 3585 - 3588.

[**ITU-T, 1996**] Methods for subjective determination of transmission quality // ITU-T Recommendation P. 800: Telephone transmission quality. — 1996. — 37 p.

[**Clark, 1982**] High-Resolution Subjective Testing Using a Double-Blind Comparator / D. L. Clark //Journal of the Audio Engineering Society. — Vol. 30 No. 5, May 1982. — P. 330-338.

[**Osgood, 1952**] The nature and measurement of meaning, / C.E. Osgood // Psychological Bulletin. — vol. 49, 1952. — P.197-237

[**Галунов, 1978**] Влияние индивидуальных и эмоциональных изменений параметров артикуляторного тракта на характеристики речевого сигнала / В.И. Галунов, С.Л. Коваль, И.Б. Тампель // Речь, эмоции и личность. — Л.: Наука, 1978. — С. 83-90.

[**Петренко, 1988**] Психосемантика сознания / В. Ф. Петренко. — М.: Изд-во МГУ, 1988. — 207 с.

[**Манеров, 2012**] Опыт использования метода семантического дифференциала при изучении слухового восприятия речевых сообщений / В.Х.Манеров // Международная научно-практическая конференция «Преемственность психологической науки в России: традиции и инновации» — Спб., 2012. — С. 452–458.

SEMANTIC DIFFERENTIAL METHOD FOR THE VOICE CONVERSION QUALITY TESTING

Zahariev V.A. , Petrovsky A.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

zahariev@bsuir.by

palex@bsuir.by

The report examines the possibility of using the semantic differential method for assessing the quality of voice conversion. Standard approaches based on subjective methods of estimating the MOS and ABX, primarily designed to assess the degree of intelligibility, and can not be directly used to determine the degree of closeness of speakers voices. Presents a methodology for assessing the quality of the conversion on the basis of building a listeners semantic space.

Keywords: semantic differential, psycholinguistics, voice conversion.