



OSTIS-2011

(Open Semantic Technologies for Intelligent Systems)

УДК 004.934.1

МЕТОДЫ СЕМАНТИЧЕСКОГО АНАЛИЗА ДЛЯ ПОСТРОЕНИЯ ГОЛОСОВЫХ ИНТЕРФЕЙСОВ: ПОСТРОЕНИЕ ВЕКТОРА ПРИЗНАКОВ РЕЧЕВОГО СИГНАЛА НА ОСНОВЕ АНТРОПОМОРФИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ

А.А.Петровский (palex@bsuir.by)

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

В данной статье рассматриваются вопросы создания адаптивных алгоритмов построения вектора признаков речевых сигналов с использованием антропоморфической обработки информации и методов семантического анализа.

В настоящее время методы оценки векторов признаков на основе спектрального и кепстрального анализа достаточно хорошо разработаны и широко используются в системах распознавания речи. Однако такого рода алгоритмы обладают целым рядом серьезных ограничений, не позволяющих осуществлять качественный первичный анализ речевых сигналов в условиях противодействия (шумов). Обладая сравнимым с человеческим уровнем эффективности в отсутствии противодействия, искусственные распознаватели существенно проигрывают человеку в способности распознавать речь в присутствии различных типов помех, характерных для реальной акустической обстановки. Поэтому для широкого применения голосовых интерфейсов в технике необходимо повышение эффективности построения векторов признаков путем использования адаптивных систем анализа и обработки речевых сигналов с использованием методов искусственного интеллекта.

Модификация алгоритмов выделения признаков с целью увеличить степень инвариантности получаемых с их помощью векторов признаков является одним из возможных путей решения поставленной проблемы. В рамках этого подхода, доступный инструментарий повышения эффективности в основном ограничивается набором линейных методов и алгоритмов обработки сигналов. Такая ситуация объясняется принципиальной трудностью синтеза нелинейных алгоритмов обработки сигналов с желаемыми свойствами. Принципиальная возможность достижения большей эффективности распознавания при помощи использования нелинейных методов и алгоритмов обработки сигналов является следствием сравнения уровней производительности существующих искусственных систем и человека в реальных акустических условиях. Развитие аудиторного аппарата в ходе биологической эволюции сделало его универсально-эффективным средством оценки окружающей акустической обстановки и выделения полезного сигнала из входного потока акустических колебаний. Сведения, полученные при исследованиях аудиторного аппарата человека доказывают нелинейность его свойств.

Сравнение производительности человека и машин также приводит к предположению, что более точное моделирование механизма обработки информации человеческим аудиторным аппаратом приведёт к созданию искусственных систем, в большей степени соответствующих уровню человека, т.е. обладающих свойством универсальной эффективности распознавания в

широком спектре возможных противодействий среды. Современные методики выделения признаков широко используют знания о свойствах аудиторного аппарата человека. Примерами перцептуально обоснованных методик могут служить алгоритмы RASTA и PLP. Но, в основном, эти знания почерпнуты из психоакустического субъективного описания целостной аудиторной системы в экспериментах с простыми звуками.

Психоакустика доказывает, что аудиторный аппарат человека – нелинейная система, параметры которой зависят от времени. Следовательно, не существует возможности точно предсказать её отклик на такой сложный и нестационарный сигнал как речь при использовании только той информации, которая заключена в психоакустическом описании. Путь решения данной проблемы лежит в использовании объективно зафиксированных физиологических свойств элементов аудиторной системы.

Антропоморфический принцип моделирования аудиторной системы человека открывает принципиальную возможность расширить круг методов и алгоритмов обработки сигналов подходящих для повышения эффективности распознавания. Он добавляет нелинейные методы и алгоритмы, построенные по “образу и подобию” важнейших процессов, происходящих внутри аудиторной системы человека и обеспечивающих эффективную обработку поступающей информации биологической системой. Успешное применение антропоморфических алгоритмов выделения признаков в составе искусственных систем распознавания речи способно серьёзно продвинуть эти системы по пути их широкого применения в качестве удобного и производительного естественного интерфейса между человеком и машиной.

При этом необходимо решить следующие задачи:

1. Осуществить выбор наиболее важных процессов, происходящих внутри аудиторного аппарата человека, с точки зрения обработки сигналов биологической системой для их моделирования в составе антропоморфического алгоритма обработки сигналов;
2. Исследовать свойства моделей, описывающих выбранные физиологические процессы, выделить их предположительный алгоритмический смысл и сделать возможной интеграцию сконструированных на этой основе антропоморфических методов и алгоритмов обработки сигналов в составе антропоморфического алгоритма выделения признаков

На основе обзора исследований нарушений слуха человека показано, что недостаточность функции слуховой улитки, выражающаяся в нарушении механизмов взаимного влияния компонент сложного входного сигнала, служит причиной пониженной способности пациентов к распознаванию речи в условиях противодействия. Соответственно, область поиска физиологических механизмов ограничена механизмами внутреннего уха, вносящими вклад в психоакустически наблюдаемое маскирование. Среди них выделены явления суппрессии и адаптивности откликов внутренних волосковых клеток.

Аудиторная суппрессия заключается в объективном снижении отклика участка базилярной мембраны на стимуляцию тоном с частотой, совпадающей с центральной для данного участка, в присутствии других компонент сложного сигнала. Это явление совместно с распространением возбуждения служат основными механизмами частотного маскирования. Аудиторная суппрессия моделируется в рамках представления об активной улитке.

Адаптивность откликов внутренних волосковых клеток является одним из механизмов временного маскирования. Это явление заключается в неравномерности отклика с течением времени: мгновенный отклик на только что представленный стимул значительно превышает установившееся значение отклика на стационарный сигнал. Адаптивность отклика объясняется свойствами химического синапса между волосковой клеткой и волокном аудиторного нерва и

количественно описывается при помощи т.н. резервуарной модели химического синапса, которая описывает циркуляцию нейротрансмиттера в химическом синапсе.

Предлагается конструирование антропоморфических алгоритмов, основанных на модели активной улитки и резервуарной модели химического синапса. Каждый из эквивалентных кохлеарных фильтров, характеризующих отклик отдельных участков базилярной мембраны, предполагается моделировать каскадом цифровых линейных постоянных и перестраиваемых полюсных гамма-тон фильтров (ПГТФ). Параметры перестраиваемых фильтров контролируются через цепь обратной связи, которая является феноменологической моделью активного процесса в улитке и включает в себя описание электромотильных откликов внешних волосковых клеток. Далее, на основе результатов обработки речевого сигнала моделью активной улитки делается алгоритмическая интерпретация эффекта суппрессии. В результате суппрессии ожидается повышение контраста перцептуальной сонограммы, что приведет к усилению тоновых компонент в низкочастотной области и шумовых – в высокочастотной, что согласуется со спектральной структурой речи. Алгоритмы оптимизируются с точки зрения потребления вычислительных ресурсов реализации антропоморфического алгоритма выделения признаков в составе распознавателя речи на базе гибрида искусственная нейронная сеть – скрытая Марковская модель (ИНС/СММ).

Аудио сигнал на входе системы распознавания можно представить в виде последовательности некоторых акустических “событий”, отличающихся разной степенью линейности, стационарности, зашумленности, и т.д. В связи с этим для обработки каждого такого “события” необходимо использовать алгоритм, наилучшим образом ему соответствующий и позволяющий повысить точность оценки вектора признаков. Использование

В связи с этим можно выделить следующие важнейшие задачи формирования пространства признаков на основе антропоморфической обработки информации в распознавателях речи в условиях противодействия:

1. Синтез нелинейных алгоритмов обработки сигналов на основе антропоморфического принципа, пригодных для реализации в составе технических устройств с целью улучшения их характеристик при работе в шумах, характерных для реальной акустической обстановки, который, в отличие от известных, позволяет создавать нелинейные алгоритмы обработки сигналов с малым количеством оцениваемых в ходе синтеза алгоритма параметров. Оптимальность соответствия этих алгоритмов поставленной задаче основывается на достаточном уровне точности моделирования биологической системы обработки сигналов и предположении, что моделируемая биологическая система оптимально соответствует этой задаче в результате процесса биологической эволюции.
2. Метод синтеза цифровой активной кохлеарной модели на базе банка перестраиваемых цифровых фильтров, который, в отличие от известных, минимизирует необходимое количество вычислительных ресурсов и является пригодным для применения в составе алгоритмов выделения признаков. В рамках этого метода предусмотрено рассмотрение критериев устойчивости перестраиваемых цифровых фильтров в составе модели и оценка уровня их аппроксимации в условиях минимизации частоты дискретизации представления сигналов в устройстве.
3. Метод синтеза цифровой реализации резервуарной модели синапса между волосковой клеткой и волокном аудиторного нерва в виде механизма сигнало-зависимой регуляции усиления, пригодной для использования в технических устройствах с ограниченным вычислительным ресурсом, который, в отличие от известных методов,

основывается на эквивалентном представлении модели в виде комбинации фильтров с постоянными параметрами и механизма регулировки усиления, что позволяет использовать получаемые цифровые реализации модели синапса внутри технических устройств с минимизацией частоты дискретизации представления сигналов.

4. Антропоморфический алгоритм выделения признаков-векторов, состоящий из алгоритмов повышения контраста перцептуальной сонограммы и усиления нестационарных компонентов речевого сигнала, оптимизированный с точки зрения потребления вычислительных ресурсов, для систем искусственного распознавания речи повышенной эффективности распознавания в условиях шумов, который, в отличие от известных методов, является нелинейным способом обработки сигналов.

Достигнутые успехи в решении данных задач нашли отражение в следующих публикациях

1. Ivanov A.V., Petrovsky A.A. A composite physiological model of the inner ear for audio coding. - Berlin: 116th AES Convention, 2004. – 20 p. - (Preprint / 6082).
2. Petrovsky A., Ivanov A., Baszun J. An Attempt To Adequately Estimate Intelligibility Of The Speech Perceived Through Cochlear Implant In Noisy Environment Based On The Neural Network Approach // Journal of the University of Applied Sciences Mittweida. - 1999. - N 3. - P.321-330.
3. Ivanov A., Petrovsky A. MLPs and Mixture Models for the Estimation of the Posterior Probabilities of Class Membership // Lecture notes in Artificial Intelligence. - Berlin: Springer-Verlag. - 1999.- P.215-218.
4. Ivanov A., Petrovsky A. Auditory Models for Robust Feature Extraction: Suppression // Signal Processing: Proc. of IEEE Workshop, Poznan, Poland, 10 October 2003. - P.23-28.
5. Ivanov A.V., Petrovsky A.A. Anthropomorphic feature extraction algorithm for speech recognition in adverse environments // Speech and Computer: Proc. 9th International Conference on / SPECOM'2004, St. Petersburg, Russia, 20-22 September 2004. – P.17-21.
6. A.Shadevsky, A.Petrovsky. Applying of bio-inspired speech enhancement method in strong noised acoustic environment. – Structures-waves-human health: acoustical engineering. Editor R.Panuszka. – vol. XIV, No1, Krakow 2005. – pp.151-156
7. A.Shadevsky, A.Petrovsky. Bio-inspired voice activity detector based on the human speech properties in the modulation domain // in book “Information processing and security systems”, Springer, 2005, NY, USA. – pp.43-54
8. A.Ivanov, A.Petrovsky. Analysis of the IHC adaptation for the anthropomorphic speech processing systems // EURASIP Journal on Applied Signal Processing, June 2005, No 9. – pp.1323-1333