

Темников В.А.¹, Шарий Т.В.², Темникова Е.Л.³, Конфорович И.В.³

¹Национальный авиационный университет, Киев

²Донецкий национальный университет

³Национальный технический университет Украины «Киевский политехнический институт»

ГОЛОСОВАЯ АУТЕНТИФИКАЦИЯ ОПЕРАТОРОВ, ИСПОЛЬЗУЮЩИХ В ПРОЦЕССЕ РАБОТЫ НОРМАТИВНО УСТАНОВЛЕННУЮ ФРАЗЕОЛОГИЮ

В данной статье рассматриваются вопросы проведения голосовой аутентификации операторов, особенностью функциональной деятельности которых является использование нормативно установленной фразеологии. Приведен порядок действий по расчету кепстральных коэффициентов, выбранных в качестве информативных параметров для проведения аутентификации операторов на базе смоделированной искусственной нейронной сети.

Ключевые слова: аутентификация операторов по голосу, параметризация речевого сигнала, кепстральные коэффициенты.

Введение

Существенно повысить безопасность на транспорте, в энергетике и других отраслях можно путем снижения влияния человеческого фактора на работу систем управления за счет повышения надежности работы операторов – субъектов эргатических систем [1].

Проблему повышения надежности операторов авторы предлагают решать путем осуществления автоматического непрерывного дистанционного негласного контроля за работой операторов в течение рабочей смены в режиме реального времени. Полученные авторами результаты исследований применимы к операторам, особенностью функциональной деятельности которых является использование нормативно установленной фразеологии. Примером таких операторов являются диспетчеры управления воздушным движением (авиадиспетчеры).

Разработанная авторами система контроля за работой операторов позволяет в режиме реального времени производить голосовую аутентификацию и идентификацию (если этого требуют результаты аутентификации) операторов, а также мониторинг их психофизиологического состояния (определение эмоционального состояния и состояния усталости) на основе результатов анализа непрерывной (слитной) речи (речевых фрагментов, произносимых операторами в процессе выполнения функциональных обязанностей).

В настоящей статье приведены предложения авторов по повышению точности аутентификации и идентификации операторов.

Основными этапами разработанного алгоритма построения системы аутентификации/идентификации операторов по непрерывной (слитной) речи являются:

1. Поиск в непрерывной (слитной) речи слов (речевых отрезков), модели которых составляют словарь, и выделение их для последующей аутентификации/идентификации операторов.

2. Проведение аутентификации операторов по выделенным словам на основе расчета параметров, описывающих найденные слова (речевые отрезки), и сравнения векторов параметров с моделями, составляющими базу данных, или использования обученных искусственных нейронных сетей.

Настоящая статья, в первую очередь, посвящена рассмотрению второго из указанных этапов.

Аутентификация операторов производится по голосу с применением теории распознавания образов на основе типовой структурной схемы, включающей в себя, среди прочих,

подсистемы предварительной обработки речевого сигнала, параметризации и принятия решения (классификации).

Подсистема предварительной обработки

В процессе предварительной обработки производятся: «усиление» высоких частот, сегментация, наложение оконной функции и нормализация по времени (DTW) [2]:

- «Усиление» высоких частот

Сначала необходимо усилить высокие частоты сигнала для выравнивания спектра (так как вокализованные участки речи характеризуются резко спадающим спектром). Для этого целесообразно использовать фильтр с конечной импульсной характеристикой вида $H(z) = 1 - az^{-1}$, где коэффициент a выбирается из промежутка $[0.4, 1.0]$.

Эксперименты показали, что значение коэффициента a целесообразно принять равным 0,9375.

- Сегментация

Сегментация состоит в разбиении сигнала на фреймы.

При малой длительности фрейма (20...40 мс) это дает возможность рассматривать речевой сигнал - на каждом фрейме - как однородную функцию, что необходимо для получения хорошей спектральной картины. Длина фрейма в значительной мере определяется частотой дискретизации. На основе результатов проведенного анализа длины фреймов были выбраны равными 240 отсчетов при шаге, равном 160 отсчетам, при частоте дискретизации 12 кГц.

- Наложение оконной функции

Простейшая прямоугольная оконная функция является малоэффективной, т.к. при последующем применении дискретного преобразования Фурье, вследствие ограниченной длительности сигнала, для которого оно будет применено (малой длительности фреймов, на которые разбивается исходный речевой сигнал), спектр, полученный после преобразования, будет «размыт». Для устранения этого эффекта используют более сложные оконные функции. В зависимости от целей и вида исходного сигнала могут быть применены оконные функции высокого и низкого разрешения. Примерами первых являются окна Хемминга, Ханна, Блекмана, вторых – окна Наталла, Блекмана-Наталла [3].

Авторами оконная функция выбиралась исходя из динамического диапазона сигнала так, чтобы уровень боковых лепестков спектра оконной функции был меньше динамического диапазона сигнала.

Уровни боковых лепестков окон Ханна и Хемминга равны -31,5 дБ и -42 дБ соответственно. Оба варианта являются удовлетворительными и дают приблизительно одинаковые результаты. В итоге предпочтение было отдано окну Ханна, т.к. оно имеет несколько большее разрешение:

$$w(n) = 0.5 \left(1 - \cos \left(2\pi \frac{n}{N} \right) \right), \quad 0 \leq n \leq N,$$

где n – номер отсчета, N – общее количество отсчетов.

- Нормализация по времени

Нестабильность темпа речи оператора, вызванная влиянием таких факторов, как интонация, акцент и др., порождает различия по длительности реализаций речевых сигналов, а также их отдельных составляющих. Для корректного сопоставления речевых образов необходимо производить их выравнивание по времени произнесения (длине). Выравнивание путём линейного сжатия или растяжения одной реализации слова до величины другой решает задачу лишь частично.

Авторы для нормализации речевых сигналов по времени предлагают применить метод DTW.

Подсистема параметризации

Сложность задачи распознавания личности по голосу состоит в определении информативных дикторозависимых параметров, которые бы гарантированно её характеризовали, и в разработке методов эффективного выделения этих параметров для обработки в оптимальном математическом представлении. «Удачное» выделение информативных параметров является основой для принятия системой аутентификации достоверного решения [4].

Речевой сигнал содержит такие характеристики, как тембр голоса, ритм, глубина и др., но далеко не все из них можно эффективно представить в формализованном виде. В качестве параметров речевых сигналов в литературе предлагается использовать, например, следующие параметры: частота основного тона и формантные частоты. Эти параметры весьма информативны, однако, являются не слишком устойчивыми; к тому же их количество часто является недостаточным для надёжного определения личности [5].

Более удобным для построения систем принятия решения (в частности, их построения на основе искусственных нейронных сетей) является использование в качестве параметров речевых сигналов спектральных коэффициентов, коэффициентов линейного предсказания (КЛП) и кепстральных коэффициентов.

Анализ показал, что использование спектральных составляющих «в чистом виде», как правило, приводит к неудовлетворительным результатам из-за большого числа параметров, их сильной зависимости от шумов и незначительных изменений исходного речевого сигнала. Использование же в качестве параметров речевых сигналов КЛП (особенно, если предполагается использовать достаточно большое число коэффициентов) требует проведения весьма большого количества вычислений; к тому же, они уступают в помехоустойчивости кепстральным коэффициентам.

В процессе выбора параметров было проведено сравнение «разброса» значений КЛП и кепстральных коэффициентов при произнесении контрольной фразы одним и тем же диктором несколько раз. На рис.1 представлены результаты такого сравнения для КЛП 10-го порядка.

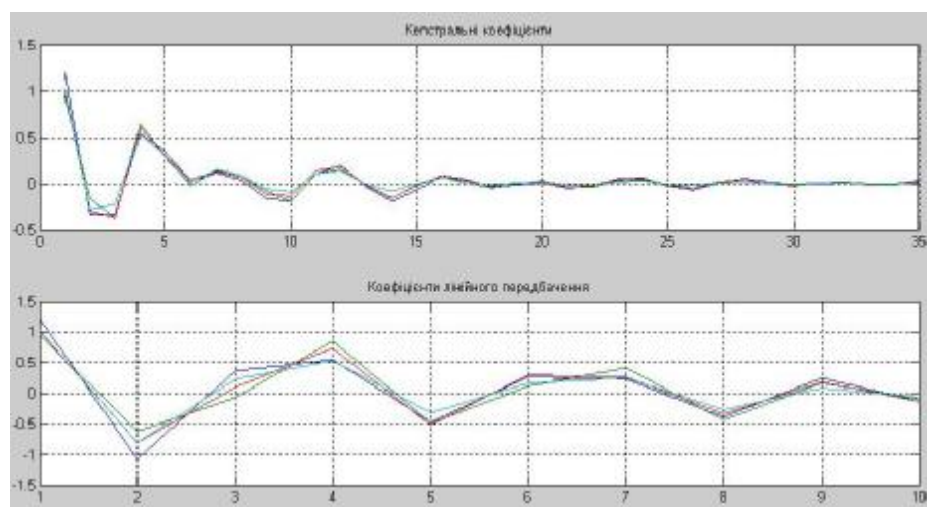


Рис.1. Сравнение «разброса» КЛП и кепстральных коэффициентов при неоднократном произнесении контрольной фразы одним и тем же диктором

Из рис.1 видно, что указанный «разброс» существенно меньше в случае использования кепстральных коэффициентов.

Приведенные выше обстоятельства обусловили выбор авторами в качестве параметров в разработанной системе параметризации кепстральных коэффициентов, которые определяем по-фреймово (по-кадрово).

Расчет кепстральных коэффициентов может производиться с применением различных методов, в частности, основанных на предварительном определении мел-кепстральных коэффициентов, КЛП, коэффициентов перцептивного линейного предсказания и др.

Авторы предлагают для расчета кепстральных коэффициентов использовать метод, основанный на предварительном вычислении КЛП.

КЛП рассчитывались автокорреляционным способом по формуле [6]: $a = R^{-1}r$, где R – автокорреляционная матрица в форме матрицы Тёплица, r – автокорреляционный вектор, a – вектор КЛП. Коэффициенты автокорреляции рассчитываются с использованием алгоритма Левинсона-Дюрбина. КЛП рассчитываются таким образом, чтобы ошибка в смысле наименьших квадратов была минимальна для заданного порядка предсказания p .

Порядок линейного предсказания, как правило, берут в пределах от 8 до 20 в зависимости от целей и способов реализации.

Определение кепстральных коэффициентов по КЛП производим по формуле:

$$c(n) = \begin{cases} 0, n < 0 \\ \log_e(A), n = 0 \\ a_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) c(k) a_{n-k}, 0 < n < p \\ \sum_{k=n-p}^{n-1} \left(\frac{k}{n}\right) c(k) a_{n-k}, n > p \end{cases}$$

где a_i – коэффициенты линейного предсказания, $c(i)$ – кепстральные коэффициенты, p – порядок линейного предсказания, n – номер кепстрального коэффициента, A – ошибка линейного предсказания.

Приведенная формула рекурсивна, что позволяет сгенерировать желаемое число коэффициентов для параметризации. Их качество напрямую зависит от количества КЛП.

Эксперименты, в процессе которых несколько дикторов много раз произносили различные фразы, показали хорошую повторяемость значений кепстральных коэффициентов. В результате было выявлено, что информативными для проведения аутентификации являются первые 12-20 кепстральных коэффициентов [7].

Количественными показателями качества параметризации могут служить коэффициенты корреляции кепстральных коэффициентов соответствующих фреймов слов, сказанных одним диктором.

Среднее значение корреляции для данных параметров равно 0.93. Пример корреляций кепстральных коэффициентов фреймов сигналов приведен на рис.2.



Рис.2. Коэффициенты корреляции кепстральных коэффициентов при неоднократном произнесении контрольной фразы одним и тем же диктором

И хотя этот результат является весьма высоким, использовать кепстральные коэффициенты, полученные пофреймово, при построении системы сравнения на основе искусственных нейронных сетей (ИНС), к сожалению, не всегда удобно из-за их большого количества. Поэтому авторами было предложено несколько изменить подход к получению кепстральных коэффициентов, что позволит уменьшить их количество до приемлемого для ИНС.

Коэффициенты линейного предсказания предлагается рассчитывать для всего слова (в этом случае длина фрейма равна длине слова). Так как в этом случае никакого четкого спектра получено быть не может, имеет место усредненный временной спектр. При этом, отпадает необходимость в использовании оконной функции. Расчет кепстральных коэффициентов производится, как было описано выше.

Существенное уменьшение количества вычислений дает возможность применять коэффициенты линейного предсказания больших порядков.

Эксперимент, проведенный для тех же параметров, что и в предыдущем случае, показал следующие результаты: время обработки сигнала получилось равным 0,05 с, среднее значение корреляции кепстральных коэффициентов – 0,973 (для некоторых параметров оно достигает значения 0,99).

Одновременно были рассчитаны коэффициенты корреляции для КЛП - их среднее значение получилось равным 0,947.

Полученные результаты, как и зависимости, приведенные на рис.1, свидетельствуют о том, что кепстральные коэффициенты характеризуют речевой сигнал лучше, чем КЛП.

Кепстральные коэффициенты, рассчитанные на основе предложенного метода, были применены в системе принятия решения (классификации), построенной на основе ИНС.

Подсистема принятия решения (классификации)

Как и в случае параметризации, начальные параметры для ИНС выбирались исходя из эмпирических соображений, кроме числа нейронов в первом слое (выбиралось равным числу параметров – 35) и числа нейронов в последнем слое (выбиралось равным числу распознаваемых личностей – 5).

Исходная выборка в 1161 сигнал была разбита на обучающую выборку и тестовую выборку в пропорции 3 к 1.

После многократных тестирований окончательные значения параметров были выбраны следующими: количество скрытых слоёв – 1, количество нейронов в скрытом слое – 20, количество эпох – 60.

В качестве метода обучения был выбран метод взвешенных сопряженных градиентов. В качестве количественной характеристики результатов использовалась матрица попаданий (confusion matrix), указывающая, на количество ошибок распознавания первого и второго рода.

Результаты работы ИНС: среднее значения «правильных» ответов – 99,69 %, время обучения и тестирования – 2,30 с, время обработки одного сигнала – 0,0064 с.

Такие высокие результаты работы ИНС объясняются высокой степенью корреляции кепстральных коэффициентов, выбранных в качестве параметров речевых сигналов при проведении их параметризации.

Выводы

В статье описана процедура проведения аутентификации/идентификации операторов как субъектов эргатических систем по непрерывной (слитной) речи с учетом особенностей их работы, заключающихся, в частности, в использовании операторами в процессе выполнения функциональных обязанностей нормативно установленной фразеологии.

Разработанная система аутентификации/идентификации построена с применением теории распознавания образов на основе типовой структурной схемы, включающей в себя подсистемы предварительной обработки, параметризации и принятия решения (классификации). При этом в процессе предварительной обработки производятся «усиление» высоких частот, сегментация, наложение оконной функции и нормализация по времени.

В качестве информативных параметров для проведения аутентификации операторов предложено применять определяемые по фрейм-кепстральным коэффициентам. Приведен порядок действий по расчету кепстральных коэффициентов.

С учетом исходных данных и результатов параметризации была создана искусственная нейронная сеть. Малое результирующее время обработки сигнала позволяет реализовать аутентификацию операторов в реальном времени.

Литература

1. Буров О.Ю. Ергономічні основи розробки систем прогнозування працездатності людини - оператора на основі психофізіологічних моделей діяльності: дис. д-ра техн. наук: 05.01.04/НДІ проблем військової медицини Збройних Сил України.— К., 2006. — 329 с.
2. Шарий Т. В. О проблеме параметризации речевого сигнала в современных системах распознавания речи // Вісник Донецького національного університету. Серія А. Природничі науки, 2008, вип.2 (2), Донецьк, 2008, С. 536—541
3. Weisstien, Eric W. Hanning Function // From MathWorld - A Wolfram Web Resource. - <http://mathworld.wolfram.com/HanningFunction.html>
4. Темников В.А., Темникова Е.Л. Параметризация автоматического контроля доступа операторов к ресурсам информационных систем по голосу // Вестник Восточноевропейского национального университета им.В.Даля. - №9 (151). – Ч.1. – 2010. – С.143-148
- Рамишвили Г. С. Автоматическое опознавание говорящего по голосу. М.: Радио и связь, 1981. – 224 с.
5. Methods for Speech Recognition - A Project for the Spectral Analysis course ELEC 532. Rice University, May 1998. - <http://www.clear.rice.edu/elec532/PROJECTS98/speech/cepstrum/cepstrum.html>
6. Темнікова О.Л., Темніков В.О., Конфорович І.В. Алгоритм побудови системи параметризації для автентифікації операторів за голосом // Третя наукова конференція «Прикладна математика та комп'ютеринг» (ПМК 2011). Збірник тез доповідей. – К.: Просвіта. – С. 333-336

Надійшла до редколегії 13.06.2011

Рецензент: д.т.н., проф. Молчанов А.А., Национальный технический университет Украины «Киевский политехнический институт», Киев

Темніков В.О., Шарій Т.В., Темнікова О.Л., Конфорович І.В.

ГОЛОСОВА АУТЕНТИФІКАЦІЯ ОПЕРАТОРІВ, ЩО ЗАСТОСОВУЮТЬ В ПРОЦЕСІ РОБОТИ НОРМАТИВНО ВСТАНОВЛЕНУ ФРАЗЕОЛОГІЮ

У даній статті розглядаються питання проведення голосової автентифікації операторів, особливості функціональної діяльності яких є використання ними нормативно встановленої фразеології. Наведено порядок дій з розрахунку кепстральних коефіцієнтів, обраних як інформативні параметри для проведення автентифікації операторів на базі змодельованої штучної нейронної мережі.

Temnikov V.A., Shariy T.V., Temnikova E.L., Konforovich I.V.

VOICE AUTHENTICATION OF OPERATORS THAT USE REGULATED VOCABULARY DURING WORK

In this article issues of performing voice authentication of operators which use regulated vocabulary are discussed. Course of action on the calculation of cepstral coefficients which were chosen as the informative parameters for the authentication of operators based on a modeled neural network, is presented.