Институт информационных и вычислительных технологий
МОН РК

Казахский Национальный Университет имени аль-Фараби

Университет Туран

Люблинский технический университет, Польша


МАТЕРИАЛЫ

III Международной научной конференции

«Информатика и прикладная математика»,

посвященная 80-летнему юбилею

профессора Бияшева Р.Г.

и 70-летию профессора Айдарханова М.Б.

26-29 сентября 2018 года, Алматы, Казахстан

Часть 2

Алматы 2018

УДК 004(063)
ББК 32.973
И74

И 74 **Информатика и прикладная математика:** Мат. III Межд. науч. конф. (26-29 сентября 2018 г). Часть 2. – Алматы, 2018. – 449 с.

ISBN 978-601-332-165-3

В сборнике опубликованы доклады, представленные по 5 секциям от Республики Казахстан, Российской Федерации, США, Латвии, Польши, Республики Беларусь, Украины, Азербайджана, Узбекистана, Японии, Кореи, Ирана, Португалии, Испании, Великобритании, Греции, Кыргызской Республики и других.

Рассмотрены актуальные вопросы в области математики, информатики и управления: математического моделирования сложных систем и бизнес-процессов, исследования и разработки защищенных и интеллектуальных информационных и телекоммуникационных технологий, математической теории управления, технологий искусственного интеллекта.

Материалы сборника предназначены для научных работников, докторантов и магистрантов, а также студентов старших курсов.

УДК 004(063)
ББК 32.973

ISBN 978-601-332-165-3

современных средств и систем ведения разведки, дающих ясное представление о противнике в реальном времени, является ключевой проблемой на пути радикального повышения боевой эффективности воинских формирований любого уровня. Основными требованиями к современным средствам дистанционного ведения наземной разведки, обеспечивающим действия против современного наземного противника в конфликте низкой интенсивности, являются своевременность, достоверность, полнота и точность получаемой информации, а также скрытность, низкая уязвимость и возможность функционирования на территории, занятой противником [4].

Выполнение данных задач возможно решить при тесном взаимодействии целого ряда специалистов различного профиля военного и гражданского сектора, которые смогли бы выполнять работы не только по модификации существующих образцов БЛА, но и созданию новых отечественных экспериментальных моделей, отвечающих современным требованиям, систем и комплексов.

Список использованной литературы

1. Ногуманов Д.У., Юсупов С.В., Мутив Н.Х. Направление и перспективы развития Вооруженных Сил Республики Казахстан до 2050 года по опыту зарубежных стран. Учебное пособие. АО «ЦВСИ». Астана 2013 г., стр.26-28.
2. В.Щербаков. Пентагон составил тридцатилетний план развития авиации. Минобороны США отдает приоритет многофункциональным беспилотным летательным аппаратам. Независимое военное обозрение №8 (5-11 марта) 2010г.
3. В.Романов. Программа использования БЛА в США. Ежемесячный информационно-аналитический иллюстрированный журнал МО РФ «Зарубежное военное обозрение», № 12-2013г., стр.86.
4. Современное состояние и перспективы развития беспилотных авиационных систем XXI века: аналитический обзор по материалам зарубежных информационных источников/ Под общ.ред. академика РАН Е.А.Федосова. М.: ГосНИИАС, 2012г.-177 с.

АЛГОРИТМЫ И АРХИТЕКТУРЫ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ

Мамырбаев О.Ж., Мекебаев Н.О., Турдалыулы М.

Институт информационных и вычислительных технологий
КН МОН РК, Казахстан
NURBAPA@MAIL.RU

Аннотация

Цифровая обработка речевого сигнала и алгоритма распознавания голоса очень важна для быстрого и точного автоматического озвучивания распознавание технология. Голос - сигнал бесконечной информации. Прямой анализ и синтез сложного речевого сигнала обусловлен тем, что информация содержится в сигнале.

Речь - это самый естественный способ общения людей. Задача распознавания речи - преобразовать речь в последовательность слов с помощью компьютерной программы.

В этой статье представлен алгоритм извлечения MFCC для распознавания речи. MFCC алгоритм снижает вычислительную мощность на 53% по сравнению с обычным алгоритмом. Автоматическое распознавание речи с использованием Matlab.

Ключевые слова: *распознавание речи, звуковой сигнал, нейронный сети, MFCC, ASR.*

Введение

Автоматическое распознавание речи является динамично развивающимся направлением в области искусственного интеллекта.

Распознавание речи - это процесс распознавания слов, произнесенных человеком на основе автоматического речевого сигнала. Продолжительность слова может быть разной на разных частотах, и одна и та же длина зависит от одних и тех же слов, разные части слов отличаются от кажущегося темпа разницы в окружающей среде. Вам нужно выровнять время, чтобы получить расстояние между речами (представленное как последовательности векторов). Для решения проблем использовалось сравнение понятия динамического выравнивания по спектральной последовательности слов.

Задача распознавания речи на сегодняшний день является актуальной проблемой.

Большинство современных методов, используемых для ее решения, требуют больших вычислительных ресурсов, объем которых часто бывает ограничен. Невозможность широкого применения многих алгоритмов сегодня, например, в мобильных устройствах заставляет исследователей искать более эффективные методы.

В данной статье является описание алгоритма и анализ методов распознавания речи, выявление недостатков каждого из них. Разработка программы по распознаванию речи и проведение эксперимента.

Модуль распознавания речи включает в себя два основных процесса цифрового сигнала: извлечение функций и соответствие функций. Первый обрабатывает слово, произнесенное пользователем, и генерирует его функции. Произнесенная речь сначала преобразуется в цифровой домен, и цифровая дискретизированная речь обрабатывается для извлечения функций с использованием подхода MFCC, который дает оценку фильтра голосового тракта. В

разделе I. расмотривается предварительная обработка речевого сигнала, в разделе II. расмотривается выделение признаков речи с помощью алгоритма MFCC, в разделе III. расмотривается архитектуры система автоматического распознаваия речи.

I. Предварительная обработка речевого сигнала

При записи речи на амплитуду звукового сигнала влияет целый ряд факторов: громкость голоса диктора, его удаленность от микрофона и т.д. Перечисленные факторы приводят к большой вариативности громкости речевого сигнала [1]. Особенно сильно это явление заметно при использовании разнородной звукозаписывающей аппаратуры. Для устранения разброса громкости применяется процедура нормализации по амплитуде. С помощью этого приема амплитуда сигнала заключается в границы$[\frac{\Delta}{2},-\frac{\Delta}{2}]$ (рис.1) $\tilde{p}$[n] выполняется по следующей формуле:

$$\tilde{p}\,[n] = \frac{\Delta}{\max\limits_{m} |s[m]|} \cdot p[n] \qquad (1)$$

где Δ – ширина полосы нормализации, симметричной относительно оси абсцисс (например, на рис. 1. Δ= 1).

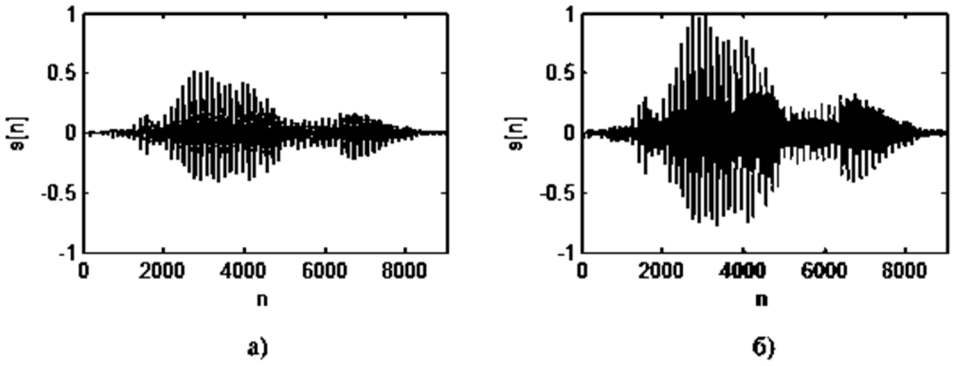

Рис. 1. Оцифрованный речевой сигнал до (а) и после (б) нормализации

Для оценки вариативности громкости рассмотрим набор из Q примеров произнесения одного или нескольких слов. Найдем среднее значение громкости M_q [2] для -го примера длиной в N отсчетов и среднее M_Q для Q примеров:

$$\mathrm{M(q)} = \sum_{n=1}^{N} |p[n]|, q = 1, \dots, Q; \qquad (2)$$

$$\mathrm{M_Q} = \frac{1}{Q}\sum_{q=1}^{Q} M(q) \qquad (3)$$

После этого рассчитаем относительное отклонение $D(q)$ громкости каждого примера от среднего:

$$\mathrm{D}_q = |1 - \frac{M(q)}{M_Q}| \quad (4)$$

Из формул (2), (3) видно, что на результат влияет как абсолютное значение отсчета, так и количество этих отсчетов в примере, поэтому необходимо оценить вариативность громкости набора примеров одного класса, длина которых примерно одинакова, и варьирование громкости базы в целом.

На рисунке 2.а)-б) показаны $D(q)$ для Q = 100 записанных примеров слова «три» до и после нормализации соответственно. Разброс громкости составил 28.5% для исходных примеров и 14.3% для нормализованных. Графики 2.в)-г) отображают $D(q)$ для речевой базы из Q = 2000 примеров различных произнесений. Разброс громкости составил 24.8 % для исходной базы и 23.11 % нормализованной.

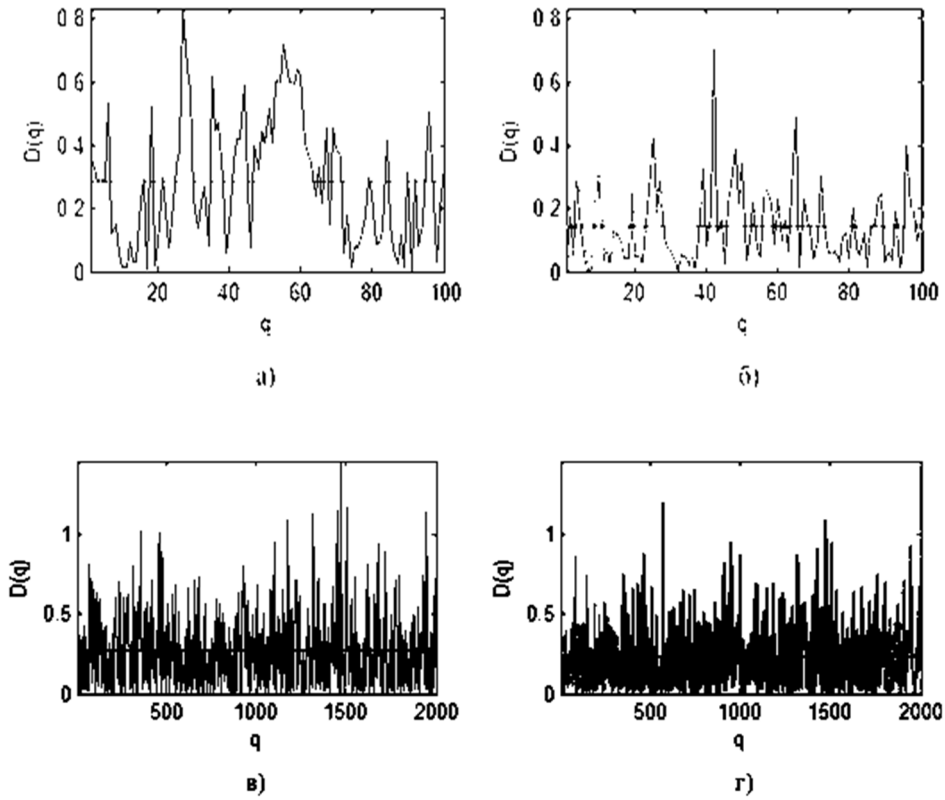

Рис. 2. Отклонение энергии примера от среднего значения энергии для (а, в) исходных примеров, (б, г) нормализованных

Как видно из проведенного исследования, применение нормализации всегда позволяет уменьшить разброс громкости для различных произнесений. Данные результаты были получены для речевой базы, собранной в одинаковых условиях на единственной доступной аппаратуре, поэтому нормализация сыграла, в общем, незначительную роль. Однако, нормализация необходима в работе реальной системы, когда прием речевого сигнала ведется в различных условиях.

II. Выделение признаков речи с помощью алгоритма MFCC

Таким образом, на вход нашей системы подается звуковой сигнал. Звук делится на фреймы – участки по 25 мс с перекрытием фреймов равным 10 мс. Для обработки звукового сигнала его следует преобразовать либо в виде спектра сигнала, либо в виде прологарифмированного спектра, с последующим масштабированием, поскольку это соответствует особенностям человеческого восприятия звука (Mel-шкала). Затем сигнал представляется в виде MFCC (Мел кепстральные коэффициенты) путем применения дискретного косинусоидального преобразования. MFCC обычно является вектором из тринадцати вещественных чисел, он представляют собой энергию спектра сигнала. Данный метод учитывает волновую природу сигнала, mel-шкала выделяет наиболее существенные частоты, воспринимаемые человеком, а количество MFCC коэффициентов можно задать любым числом, что позволяет сжать фрейм и уменьшить количество обрабатываемой информации [3].

Рассмотрим алгоритм MFCC-преобразования получаемого звукового сигнала.

Получаемый звуковой сигнал дискретизируется:

$$x[n],\ 0 \le n < N. \quad (5)$$

Представляем его в качестве Фурье преобразования:

$$X_Q[k]=\sum_{n=0}^{N-1} x[n] e^{\frac{-2\pi i}{N}kn},\ 0 \le k < N. \quad (6)$$

Рассчитываем гребенку фильтров, используя окно:

$$H_m = \begin{cases} 0 & k < f[m-1]; \\ \frac{(k-f[m-1])}{(f[m]-f[m-1])} & f[m-1] \le k < f[m]; \\ \frac{(f[m+1]-k)}{(f[m+1]-f[m])} & f[m] \le k < f[m+1]; \\ 0 & k > f[m+1], \end{cases} \quad (7)$$

где f[m] будет равно

$$f[m] = \left(\frac{N}{F_3}\right) B^{-1}\left(B((f_1) + m\frac{B(f_h)-B(f_1)}{M+1}\right) \quad (8)$$

B(b) – представляем наши частоты в виде Мел-шкалы:

$$B^{-1}(b) = 700\left(exp\left(\frac{b}{1125}\right)^{-1}\right) \quad (9)$$

где энергия окон будет равна

$$S[m] = ln\left(\sum_{k=0}^{N-1} |X_Q[k]|^2 H_m[k]\right),\ 0 \le \mathrm{m} < \mathrm{M} \quad (10)$$

Получаем коэффициенты MFCC:

$$c[n] = \sum_{N=0}^{M-1} S[m] cos\left(\pi n\left(m + \frac{1}{2}\right)/M\right),\ 0 \le \mathrm{n} < \mathrm{M} \quad (11)$$

Пусть наш фрейм представляется в виде дискретного вектора значения согласно формуле (9).

Вычислим спектр сигнала:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-2\pi kn}{N}}, \qquad 0 \le \mathrm{k} < \mathrm{N} \quad (12)$$

Обработаем сигнал окном Хэмминга, чтобы сгладить пульсации сигнала на краях [6].

$$H[k] = 0.54 - 0.46 \cdot cos\left(\frac{2\pi k}{N-1}\right) \quad (13)$$

$$X[k]=X[k]\cdot H[k],\ 0 \le \mathrm{k} < \mathrm{N} \quad (14)$$

По оси OX откладывается частота в Герцах, по оси OY – магнитуда, чтобы не связываться с комплексными величинами (рис. 3):

Mel представление показывает значимость отдельных частот звука для человека, зависит и от конкретных частот звука, и от громкости, и от тембра человека. Mel-шкала вычисляется следующим образом (прямое и обратное преобразование):

$$M = 1127 \cdot log\left(1 + \frac{F}{700}\right) \quad (15)$$

$$F = 700 \cdot (e^{M \cdot 1127} - 1) \quad (16)$$

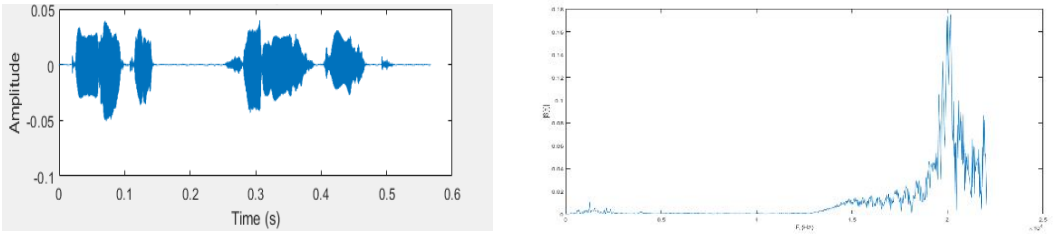

Рис. 3. Представление исходного сигнала в качестве Фурье преобразования

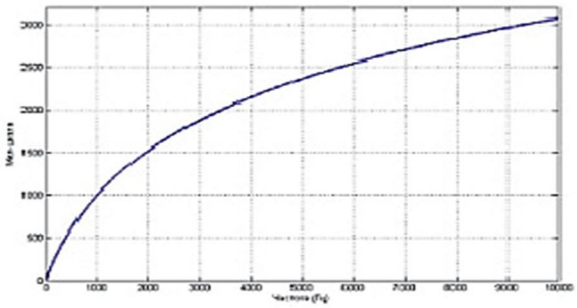

Рис. 4. График зависимости Мел-шкалы от частоты

График зависимости Мел-шкалы от частоты представлен на рис. 4.

Наибольшее распространение в системах распознавания речи получили именно эти единицы измерения, поскольку они соответствуют особенностям восприятия звука человеком.

Рассмотрим пример: дан фрейм длиной 256 отсчетов (выборок), частота звука 16 кГц. Пусть человеческая речь сосредоточена в диапазоне частот от 300 Гц до 8 кГц. Наиболее часто используемое количество Mel-коэффициентов равно десяти, его и будем использовать.

Сначала необходимо рассчитать гребенку фильтров, чтобы представить спектр в формате mel-шкалы. Мел-фильтр является треугольным окном, которое суммирует энергию на своем диапазоне частот и вычисляет mel-коэффициенты. Поскольку мы знаем количество коэффициентов, то сможем построить набор из десяти фильтров (рис. 5).

В области низких частот (те частоты, которые нам наиболее интересны) количество окон больше, что обеспечивает высокое разрешение. Это позволяет существенно повысить качество распознавания.

Для того чтобы найти энергию сигнала, перемножим вектор спектра сигнала и функцию окна, в результате чего получим вектор коэффициентов. Если их возвести в квадрат, представить в виде логарифма и получить из них кепстральные коэффициенты, то получим искомые mel-коэффициенты. Кепстральные

коэффициенты можно получить как с помощью Фурье-преобразования, так и с помощью дискретного косинусоидального преобразования [4].

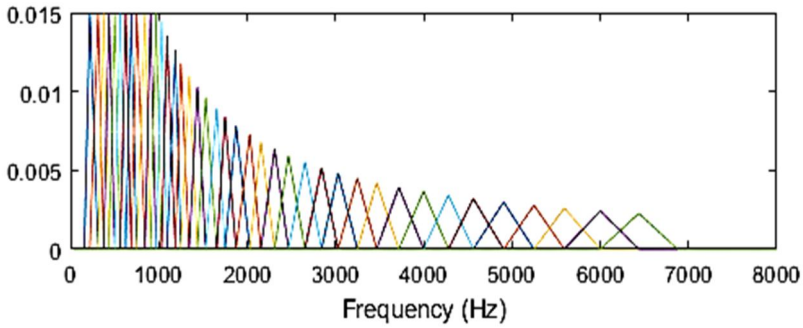

Рис. 5. Mel-частотные кепстральные коэффициенты

Диапазон частот составляет от 300 Гц до 8 кГц. На mel-шкале этот диапазон соответствует от 401,25 до 2834,99. Теперь строим двенадцать опорных точек для постройки десяти треугольных фильтров (Мел-шкала и шкала в герцах):

$$m[i] = [400.25; 622.50; 1066.00; 1286.25; 1507.50; 1727.74; 1939.98; 2161.25; 2392.49; 2612.73; 2835.98] \quad (17)$$

$$h[i] = [310; 517.35; 782.90; 1105.97; 1499.04; 1973.32; 2554.33; 3261.62; 4122.63; 5160.75; 6445.70; 8010] \quad (18)$$

Как мы уже говорили, длина фрейма составляет 256 отсчетов сигнала, частота 16 кГц (откладывается по оси OX). Наложим рассчитанную шкалу на спектр сигнала.

$$f(i)=floor((frameSize+1)*h(i)/sampleRate) \quad (16)$$

что соответствует

$$f(i)=4;8;12;17;23;31;40;52;66;82;103;128 \quad (17)$$

По опорным точкам построим фильтры:

$$H_m = \begin{cases} 0 & k < f[m-1]; \\ \frac{(k-f[m-1])}{(f[m]-f[m-1]} & f[m-1] \le k < f[m]; \\ \frac{(f[m+1]-k)}{(f[m+1]-f[m])} & f[m] \le k < f[m+1]; \\ 0 & k > f[m+1], \end{cases} \quad (18)$$

Фильтр перемножается со спектром:

$$S[m] = \log(\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k]), 0 \leq \text{m} < \text{M} \quad (19)$$

Мел-фильтры применяются к энергии спектра, затем полученные значения логарифмируются.

Дискретное косинусоидальное преобразование (ДКТ) применяется для получения кепстральных коэффициентов, оно сжимает полученные результаты, повышает вклад первых коэффициентов и понижает вклад последних.

$$C[l] = \sum_{m=0}^{M-1} S[m] cos\left(\pi l \left(m + \frac{1}{2}\right)/M\right), 0 \leq 1 < \text{M} \quad (20)$$

Получается, что у нас имеется 12 коэффициентов (рис. 5). В итоге небольшой конечный набор значений (например, двенадцать коэффициентов в нашем случае) позволяет заменить использование огромного числового массива отсчетов сигнала, либо спектра сигнала, либо периодограммы сигнала. Каждому слову конечной длины соответствует набор мел-частотных кепстральных коэффициентов. Затем необходимо найти наиболее близкую модель для определенного набора мел-частотных кепстральных коэффициентов. Для этого мы ищем евклидово расстояние между вектором мел-частотных кепстральных коэффициентов и вектором исследуемой модели. Искомой является та модель, у которой рассчитываемое расстояние наименьшее.

Набор MFCC коэффициентов для одного и того же слова может отличаться, например, в том случае, если слово произносится двумя разными людьми, либо скорость произношения отличается. Для этих целей используется алгоритм динамической трансформации времени. Он рассчитывает оптимальную деформацию времени между сравниваемыми временными последовательностями [5].

	-2	**10**	**-10**	**15**	**-13**	**20**	**-5**	**14**	**2**
3	5	12	25	36	53	70	78	89	90
-13	16	28	15	43	37	70	78	105	104
14	32	20	39	16	43	43	62	62	74
-7	37	37	23	38	21	49	45	66	71
9	48	38	42	29	44	33	47	50	57
-2	48	50	46	46	40	55	36	52	55

Рис. 6. Результаты расчетов

III. Архитектуры система автоматического распознаваия речи

Архитектура системы распознавания речи показана на рис. 7. Такие блоки из базовой системы как *вычисление признаков* и *акустических моделей* используются

перед первым проходом алгоритма. Также на втором проходе используется обычное *сравнение образов* в условиях ограниченного словаря.

Изменения касаются дополнительного первого прохода алгоритма, где *фонетический стенограф* используется для получения последовательности фонем. Затем процедура *выборки информации* создает ограниченный словарь для второго прохода алгоритма

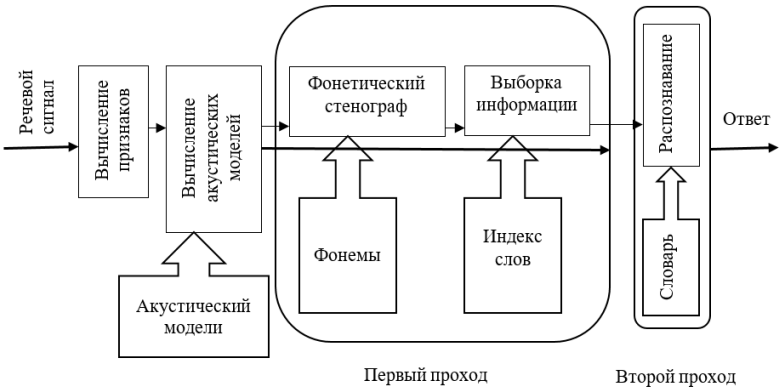

Рис. 7. – Архитектура двухпроходной системы распознавания речи из сверхбольших словарей

Распознавание слитной речи представляет собой многоуровневый процесс. После предварительной обработки речевого сигнала и выделения из него информативных признаков выполняется выделение лексических элементов речи. Это первый уровень распознавания. На втором уровне выделяются слоги и морфемы, на третьем -слова, предложения и сообщения. На каждом уровне сигнал кодируется представителями предыдущих уровней. То есть слоги и морфемы составляются из фонем и аллофонов, слова -из слогов и морфем, предложения и сообщения из слов.

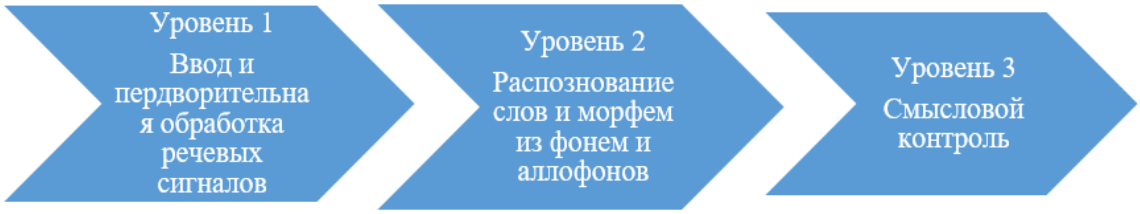

Рис. 8. Процесс распознавания речи

Мы будем использовать искусственные нейронные сети для распознавания речи. В таком случае, наша система будет выглядеть следующим образом.

Рис. 9. Структура системы распознавания речи

Речевой сигнал может поступать как из файла, так и в реальном времени с микрофона. Для того, чтобы звук можно было подать на вход нейросети, необходимо осуществить над ним некоторые преобразования. Очевидно, что представление звука во временной форме неэффективно. Оно не отражает характерных особенностей звукового сигнала. Гораздо более информативно спектральное представление речи. Для получения спектра используется набор полосовых фильтров, настроенных на выделение различных частот, или дискретное преобразование Фурье. Затем полученный спектр подвергается различным преобразованиям, например, логарифмическому изменению масштаба (как в пространстве амплитуд, так и в пространстве частот). Это позволяет учесть некоторые особенности речевого сигнала – понижение информативности высокочастотных участков спектра, логарифмическую чувствительность человеческого уха, и т.д.

Как правило, полное описание речевого сигнал только его спектром невозможно. Наряду со спектральной информацией, необходима ещё и информация о динамике речи. Для её получения используются дельта-параметры, представляющие собой производные по времени от основных параметров. Полученные таким образом параметры речевого сигнала считаются его первичными признаками и подаются на вход нейронной сети, на выходе которой будут соответствующие сигналу фонемы. Затем фонемы собираются в слова и предложения. Наиболее медленным участком этой цепи является нейронная сеть, т.к. для точного распознавания требуется достаточно большое число нейронов (в одном только входном слое для извлечения лишь информативной части спектра, без учета интонации, необходимо около 50-100 нейронов). В то же время, чтобы найти выход одного нейрона необходимо вычислить взвешенную сумму всех его входов и значение пороговой функции. Для вычисления взвешенной суммы необходимо произвести число умножений и сложений в соответствии с числом входных каналов. После этого вычислить значение пороговой функции, а это, в зависимости от сложности самой функции, от одной операции сравнения, до нескольких математических операций. Но так как все нейроалгоритмы являются высоко параллельными, выходы нейронов одного слоя могут вычисляться одновременно.

Следовательно, нейронная сеть может быть эффективно распараллелена, что позволит достичь высокого быстродействия.

Нейрон получает сигналы (импульсы) от других нейронов через дендриты и передает сигналы, сгенерированные телом клетки, вдоль аксона, который в конце разветвляется на волокна, на окончаниях которых находятся синапсы [6;7]. Математическая модель нейрона описывается соотношением [8]:

$$y = f(s), s = \sum_{i=1}^{n} x_i \omega_i + b$$

где w_i – вес синапса, b – значение смещения, s – входной сигнал, y – выходной сигнал нейрона, n – число входов нейрона, f – функция активации. Техническая модель нейрона представлена на Рисунке 1:

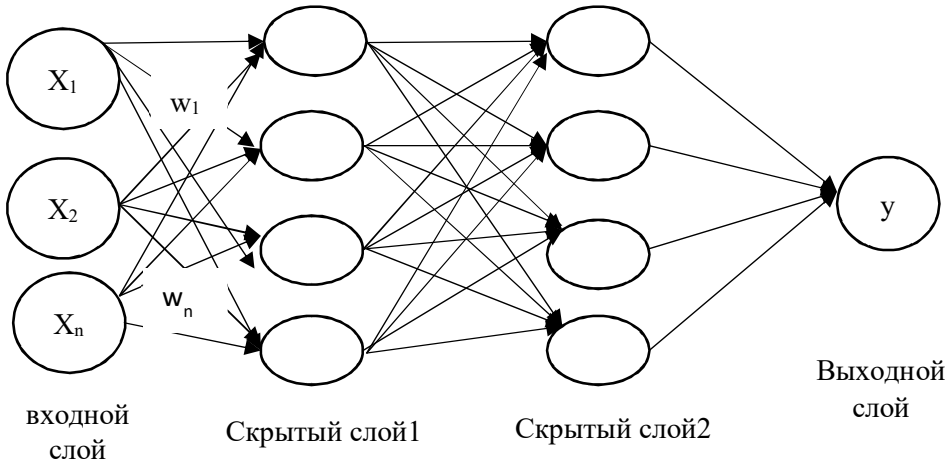

Рис. 10. Структурная схема нейрона: x_1, x_2, ..., x_n – входной сигнал нейрона; w_1 ,w_2 ,…,w_n – набор весовых коэффициентов; y – выходной сигнал

Как видно, нейроподобный элемент выполняет несложные операции взвешенного суммирования, обрабатывая результат нелинейным пороговым преобразованием. Особенность нейросетевого подхода заключается в том, что структура из простых однородных элементов позволяет решать нетривиальные задачи благодаря сложной организации связей между элементами. Структура связей определяет функциональные свойства сети в целом. Процесс функционирования сети зависит от величин синаптических связей. Задав определенную структуру сети (на этапе проектирования), находят оптимальные значения весовых коэффициентов w_1, w_2,…,w_n и смещений b всех нейронов. Этот этап называют обучением нейронной сети. Решить поставленную задачу распознавания речи с помощью нейронной сети заданной структуры – значит путем обучения по выборке, заданной k парами

значений входных и выходных векторов (X^i, Y^i), i=1,…,k, найти такую конфигурацию, чтобы обеспечить наиболее оптимальное в определенном смысле ее функционирование. Различают два подхода к обучению нейронной сети: обучение с учителем и обучение без учителя (самообучение). При обучении с учителем на вход сети подается один из векторов X^i обучающей выборки, выход сети Y сравнивается с выходом Y^i обучающей выборки, и при необходимости делаются поправки в весовые коэффициенты и смещения нейронов сети. В алгоритмах обучения без учителя подстройка весов синапсов производится на основании информации о состоянии нейронов и уже имеющихся весовых коэффициентов по одному из правил обучения. Процесс повторяется, пока выходные значения сети не стабилизируются с заданной надежностью.

Экспериментальная часть

Мною была написана программа в среде Matlab и проведены экспериментальные исследования алгоритма распознавания речевых сигналов стояла задача собрать данные по распознаванию слов: «Torequl .....» Каждое слово было произнесено третью, людьми по три раз. Был установлен шумовой порог, т.к. шумы хоть и были незначительны, но все же могли повлиять на результаты. Результаты статистических данных приведены в таблицах.

Таблица 1. Данные по слову «Torequl....»

	Попытка 1	Попытка 2	Попытка 3	Попытка 4	Попытка 5
Человек 1	+	-	+	+	-
Человек 2	-	+	+	+	+
Человек 3	-	+	+	-	+
Человек 4	+	+	-	-	+
Человек 5	-	+	-	-	-
Человек 6	+	+	-	+	+
Человек 7	-	+	+	+	+

Итого получается, что процент распознавания слова «Torequl…» равен порядка 75%.

Алгоритм программы

В начале работы на экран выводится главное окно программы. После этого на динамик микрофона подается звуковое сообщение, за который отвечает модуль ввода речевого сигнала. Затем на главном окне пользователь выбирает режим работы программы. Если выбран режим создания эталона, за который отвечает модуль создания базы данных (БД) эталонов, то программа обрабатывает и сохраняет входной сигнал с микрофона и выводит спектр на экран. Если же выбран режим распознавания, то программа обрабатывает результаты и сравнивает с заранее записанным эталоном в БД, сохраняет входной сигнал и переходит к его

распознаванию с помощью вычисления первой и второй конечной разности полной фазовой функции, т.е. определяем количество звуков в данном слове, что видно из проделанного ранее моделирования, Определяем начало и конец слова с помощью выделения огибающей. Результат распознавания выводится на дисплей. На рис.11 представлен схематический вид программы. Схема изображена ниже.

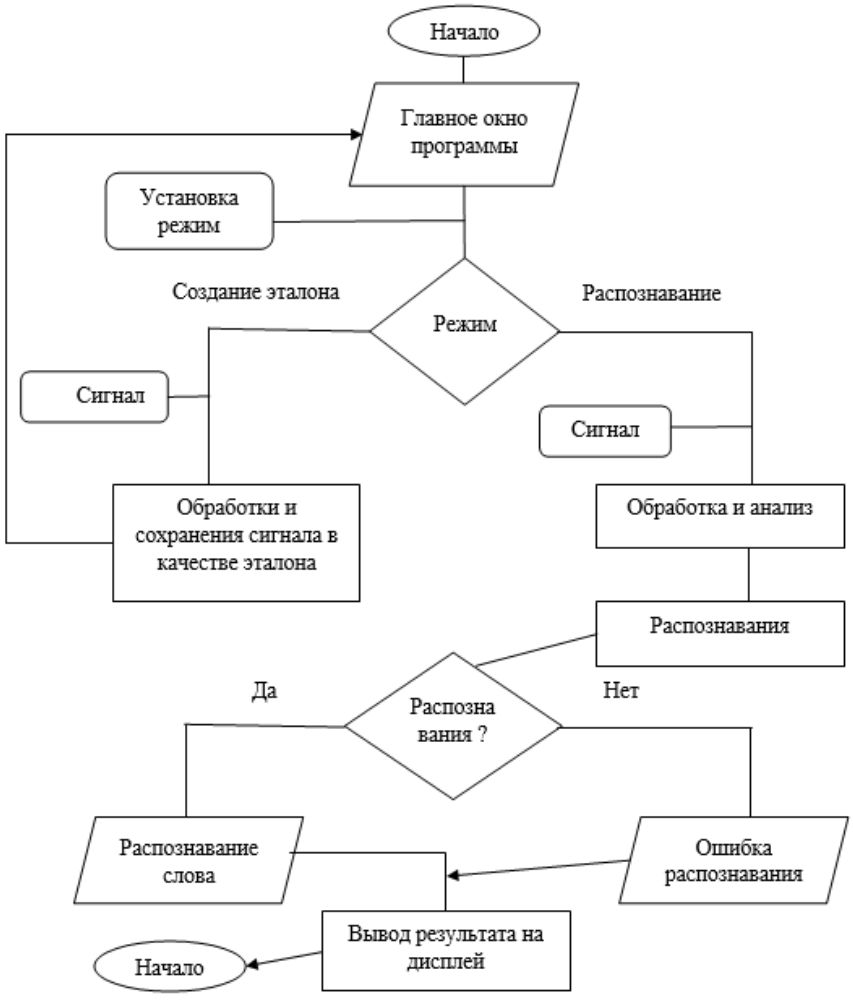

Рис.11. Алгортим программы

Заключение

Было отмечено, что MFCC для каждого отдельного пользователя уникальны. Определенные вариации наблюдались из-за различий в локальности области записи. Эти MFCC затем сравниваются, то есть MFCC шаблона и ввода в реальном времени сравниваются для каждого пользователя. В программировании евклидово

расстояние используется для сравнения шаблона и ввода в реальном времени. Таким образом, алгоритм MFCC используется для распознавания речи.

Обсуждаемый подход был опробован на голоса разных людей. База данных была создана для голоса 7 разных лиц, соответствующих цифрам от 0 до 9 и некоторым управляющим словам. Коэффициенты MFCC, соответствующие этим обучающим голосовым образцам, сохранялись вместе с вычисленными весами. MFCC вычисляется для тестового образца. Статья подготовлена на основе проекта ИРН-AR05131207 «Разработка технологии мультиязычного автоматического распознавания речи с использованием глубоких нейронных сетей».

Список литературы

1. Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi “Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) Techniques” Issue 3, March 2010.

2. Watcher, M. D., Matton, M., Demuynck, K., Wambacq, P., Cools, R., “Template Based Continuous Speech Recognition”, IEEE Transaction on Audio, Speech, & Language Processing, 2007.

3. Gupta, R., and Sivakumar, G., “Speech Recognition for Hindi Language”, IIT BOMBAY, 2006.

4. Ingle V., Proakis J. Digital Signal Processing Using Matlab V4 – Boston: ITP, 1997.

5. Rabiner, L. Juang, B. H., Yegnanarayana, B., “Fundamentals of Speech Recognition”, Pearson Publishers, 2010.

6. Barsky A.B., Neural networks: recognition, management, decision-making.

7. Cheong Soo Yee and abdul Manan ahmad, Malay Language Text Independent Speaker Verification using NN-MLP classifier with MFCC, 2008 international Conference on Electronic Design.

8. Wu Junqin, Yu Junjun, “*An Improved Arithmetic of MFCC in Speech Recognition System*,” *IEEE Transaction on Audio Speech processing*, *and Language, pp.719-722, 2011.*

9. Gurpreet Kaur, Harjeet Kaur, “*Multi Lingual Speaker Identification on Foreign Languages Using Artificial Neural Network with Clustering*”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, 2013.

10. L. Rabiner and G. Juang, “Fundamentals of Speech Recognition,” Prentice-Hall, 1993.

11. H. Hasegawa, M. Inazumi, “Speech Recognition by Dynamic Recurrent Neural Networks,” Proceedings of 1993 International Joint Conference on Neural Networks.

12. Furui, S., 50 years of progress in speech and speaker recognition. SPECOM 2005, Patras, 2005: p. 1-9.

Бердибеков А.Т., Доля А.В.	Особенности совершенствования информационных систем управления	71
Джаксылыкова А.Б., Амиржан С., Пазовский Н.С.	Сравнительный анализ методов выделения коллокации	76
Елеусинов А., Бурибаев Ж., Мажитов Ш.	Моделирование многозвенных роботизированных манипуляторов, используя Sim-Mechanics	82
Ергалиев Е., Мухамедиев Р., Якунин К., Сымагулов А., Кайрбеков А., Дуйсенбаева А.	Использование облачных платформ для решения задач машинного обучения	87
Жомартова Л.М., Мусаев М.С., Рахимова Д.Р.	Семантический поиск на основе модели векторного представления слов	95
Касенов Д.Д.	Некоторые аспекты применения беспилотных летательных аппаратов в интересах Вооруженных сил	103
Мамырбаев О.Ж., Мекебаев Н.О., Турдалыулы М.	Алгоритмы и архитектуры систем распознавания речи	107
Смагул С.С.	Исследование методов распознавания эмоционального оттенка сообщения пользователей социальной сети Twitter	122
Уалиева И.М., Красовицкий А.М., Мейрамбеккызы Ж., Мусабаев Р.Р.	Распознавание генерализации в текстах сми на основе программно-экспертного подхода	130
Хайрова Н.Ф., Мамырбаев О.Ж., Мухсина К.Ж., Пилипенко А.А.	Моделирование грамматических способов выражения семантики факта в английском предложении	136