

Санкт-Петербургский Государственный Университет

Математико-механический факультет

Кафедра системного программирования

**Сегментация речи по источнику первичного возбуждения,
определение артикуляционных классов сегментов**

Курсовая работа студента 445 группы

Меламуда Александра Евгеньевича

Научный руководитель А.Е. Булашевич
Кандидат технических наук

Санкт-Петербург

2010

Оглавление

Введение. История вопроса	3
Постановка задач	5
Выбор методологии решения	6
Анализ фонетических классов	8
Предлагаемое решение	10
Алгоритм реализации	15
Результаты	18
Направление дальнейшей работы	19
Список литературы	20

Введение. История вопроса

Область применения речевых технологий охватывает широкий спектр задач и постоянно расширяется. Основные работы ведутся в направлении автоматизации распознавания человеческой речи и ее транскрибации. Проблемой начали заниматься с 60-х годов. Исторически первыми были попытки распознавания изолированно произнесенных или выделенных из потока дикторской речи вручную гласных или изолированных слов, причем только по достаточно длинным стационарным участкам [4], [7]. В качестве признаков использовались данные о формантах участка. Задача автоматического распознавания даже не ставилась.

В начале 90-х годов данный подход был заброшен ввиду широкого распространения статистических методов, основанных на применении скрытых марковских моделей (hidden markov model (HMM)). Применение подобного подхода позволило полностью решить задачу распознавания читаемой речи лабораторного качества. Ошибка на уровне слов не превышала нескольких процентов [1], [6]. Автоматы, построенные по данному принципу, приемлемо справляются с задачей распознавания дикторской речи в реальных условиях (офис, автомобиль), особенно при ограниченном словаре или при наличии настройки на конкретного диктора. В плюсы подобного подхода можно также отнести тот факт, что фонемы распознаются с учетом динамики их произнесения. Как правило, марковская модель состоит из трёх состояний, соответствующих стационарному участку фонемы (центральное состояние) и двум переходным (начальное и конечное состояния).

Современные системы распознавания речи построены преимущественно по следующему принципу: исходный звуковой сигнал проходит стадию предобработки (нормализация, приведение к некоему общему уровню громкости, иногда – подъем высоких частот), потом параметризуется (господствующая параметризация MFCC и ее различные вариации), затем сигнал передается на обработку автомату, реализующему HMM. В теории и диссертациях сетка кадров привязана на вокализованных участках к импульсам основного тона. На практике же господствует жесткая сетка кадров ввиду как отсутствия алгоритмов расстановки импульсов ОТ с требуемой надежностью, так и сравнительной сложности программной реализации гибкой сетки кадров.

На данный момент задачи распознавания и сегментации потока фонем решаются одновременно с помощью HMM. Хотя по своей природе это разные задачи. Границей

между фонемами является переходный участок звука, где речевой тракт перестраивается. В сигнале это проявляется как сравнительно быстрое изменение характеристик сигнала. В рамках такого понимания границы между фонемами задача сегментации является независимой от конкретного языка (от набора фонем), в то время как задача распознавания фонем, очевидно, зависит от языка. Сложность заключается в том, что большинство современных исследований направлено на улучшение математико-статистического аппарата. Но практически нет исследований, задачей которых является разработка «лингвистических» методов решения задачи. Это выливается в то, что исчерпывающего ответа на вопрос, изменение каких характеристик, а главное в каких масштабах, является достоверным признаком наличия границы фонемы в слитном речевом потоке, нет.

Глава 1. Постановка задачи

Задачей данной курсовой работы является задача сегментации речевого потока на основании источника первичного возбуждения в двух вариантах постановки: на жесткой сетке кадров и с определением границ точных сегментов.

В соответствии с источником первичного возбуждения (голос, шум, взрыв) выделяются 5 артикуляционных типов звукового сегмента:

- взрыв
- физическая пауза (глухая смычка)
- высокочастотный шум при отсутствии периодов частоты основного тона (глухая щель)
- гармонические колебания при наличии высокочастотного шума (звонкая щель)
- гармонические колебания при наличии формантной структуры (вокализованные участки)

Фонетисты выделяют ещё один тип звукового сегмента, называемый звонкой смычкой. Он характеризуется наличием небольшого числа гармоник основного тона (обычно 2-3, редко 4) без формантной структуры. Взрыв занимает малое время (единицы миллисекунд) и не имеет самостоятельного значения, так как взрыв всегда завершает смычку (и является, собственно говоря, раскрытием смычки).

Поэтому ставится задача обнаружения границ пяти типов звуковых сегментов: глухая смычка, глухая щель, звонкая щель, вокализованные участки, звонкая смычка. Заметим, что при физически осмысленном подходе к сегментации проявляется близость этой задачи к задаче собственно распознавания, так как различным артикуляционным типам сегмента соответствуют чётко определенные группы фонем (гласные, звонкие согласные, глухие согласные и смычные). В результате успешной сегментации может быть существенно сокращено число ошибок распознавания, так как устоявшийся подход на основе НММ имеет заметную долю спутывания фонем из разных артикуляционных классов.

Глава 2. Выбор методологии решения

На данный момент при решении задачи сегментации существует несколько общих подходов: анализ спектра сигнала, анализ неких временных статистик, вероятностные методы

Вероятностные методы

Суть метода заключается в том, чтобы спроектировать некую НММ, задачей которой было бы распознавание 5 литералов (которые и соответствовали бы звуковым сегментам). Был поставлен следующий эксперимент: распознающему русскому языку автомату подавались на вход речевые сигналы. После распознавания фонемы, принадлежащие одному классу по источнику первичного возбуждения, сливались в один класс. Подобный автомат сегментации показал точность порядка 70%. Причем каких-либо методов существенно улучшить данный результат на данный момент не известно. Еще одним минусом является тот факт, что для успешной работы алгоритма необходима значительная по объему обучающая выборка (не менее 3-4 часов), размеченная вручную. В то же время следует отметить, что корректная ручная разметка и распознавание 1 минуты сигнала занимает порядка 8-10 часов работы квалифицированного лингвиста.

Анализ сигнала в спектральной области

Суть метода заключается в представлении сигнала в спектральной области. В дальнейшем анализируется изменение градиента спектра. Моменты значительного изменения градиента как раз и принимают за переходные участки. Проводились эксперименты по сегментации, использующий данный подход. Например, [3]. Были получены результаты на уровне 60-70% правильных границ, причем граница считалась правильной, если находилась в пределах 15 мс от границы, поставленной при ручной сегментации. При наличии в потоке речи среднего темпа фонем с длительностью порядка 30 мс такая точность определения места границы представляется явно недостаточной.

Анализ временных статистик

Суть метода заключается в том, чтобы по изменению неких временных характеристик распознать текущее состояние речевого тракта, а так же наличие или отсутствие того или иного источника первичного возбуждения. Преимуществом данного подхода является очень хорошее временное разрешение. В то время как спектральные методы подобного преимущества лишены. Сложности заключаются в том, что активное внедрение НММ в начале 90-х годов, решившее проблему распознавания лабораторной речи, стало причиной

того, что исследования проблемы во временной области практически прекратились. На данный момент исчерпывающего ответа, анализ какой из характеристик сигнала поможет с достаточной точностью выявить наличие того или иного артикуляционного признака, нет.

Выбор методологии решения

Исходя из постановки задачи данной работы, наиболее перспективным представляется решение данной задачи путем исследования временных закономерностей и зависимостей сигнала. Единственным его существенным недостатком является слабая теоретическая база. В то же время он может обеспечить отличное временное разрешение, несравнимое со спектральными методами. Использование же неких вероятностных методов нежелательно по следующей причине: явно или неявно алгоритмы, построенные на основе НММ, с большой долей вероятности уже будут учитывать результаты, полученные при сегментации. Из-за того, что существенно новых данных при обработке сигнала извлекаться не будет, с большой долей вероятности существенного улучшения результатов распознавания ожидать не приходится.

Глава 3. Анализ фонетических классов

Гласные

Гласные — тип звуков, при артикуляции которых потоку воздуха не создаётся существенных препятствий, соответственно, нигде над гортанью не создаётся сколько-нибудь существенного воздушного давления. При артикуляции гласных звуков ротовая полость играет роль резонатора. Акустически это отражается как периодические колебания по всему спектру. Основным источником возбуждения являются голосовые связки. Других источников возбуждения нет. Для гласных характерна большая энергия сигнала и наличие формантной структуры (существенных подъемов в определенных частях спектра) [5]

Звонкие согласные

Звонкие согласные характеризуются наличием как голосового, так и шумового возбуждения. Наличием шумового возбуждения и вызвано отличие от гласных звуков. Однако отличить подъем спектра в высокочастотной части вследствие наличия реально высокочастотного шума от подъема спектра вследствие наличия шлейфа убывающих гармоник довольно трудно. [4],[5] Дополнительной сложностью является тот факт, что вариативность различных представителей данного класса фонем большая (можно видеть на рис. 1 и рис. 2).

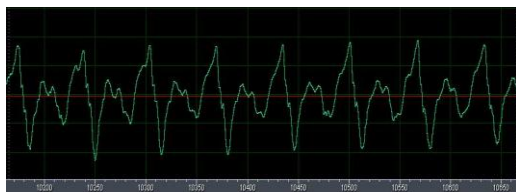


Рис 1
Звук 'n'
Гармоническая структура просматривается вплоть до старших частот. Спектрограмма сильно напоминает спектрограмму гласных

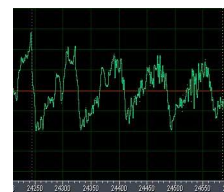
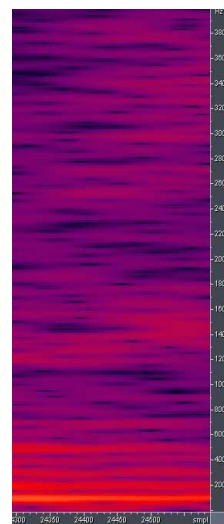
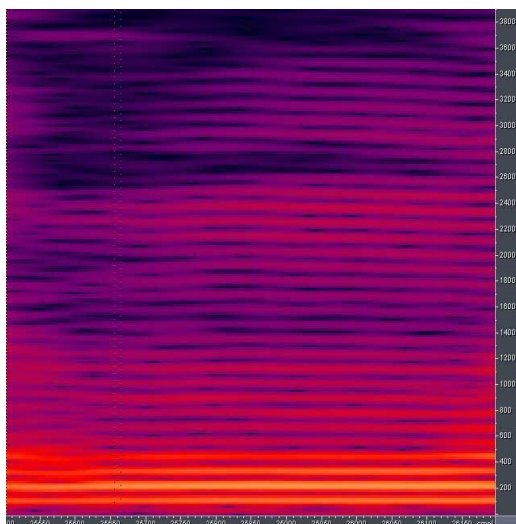


Рис 2
Звук 'z'
Гармоническая структура просматривается только на низких частотах. Явно видно наличие высокочастотного шума



Глухая щель

Согласные, характерным признаком которых является шум от сближения органов произношения, который составляет практически всё содержание звука. Голос при произношении подобных согласный отсутствует. Какая-либо гармоническая структура отсутствует. Центр тяжести спектра расположен на высоких частотах. [4], [5]

Смычные согласные

Смычные согласные — согласные звуки, при артикуляции которых органы речи находятся в таком положении, что поток воздуха из легких полностью блокируется с помощью смычки, создаваемой в полости рта или в гортани. Смычки делятся на звонкие и глухие. Глухие смычки очень похожи на физические паузы. Они характеризуются отсутствием энергии по всей ширине спектра. Звонкая смычка характеризуется наличием небольшого числа гармоник основного тона (обычно 2-3, редко 4) без формантной структуры. Взрыв занимает малое время (единицы миллисекунд) и не имеет самостоятельного значения, так как взрыв всегда завершает смычку (и является, собственно говоря, раскрытием смычки). В частности взрыв является характерным признаком смычки. После взрыва может проявляться фаза аспирации, внешне напоминающая глухую щель (каковой физически и является). Может достигать 15-20 мсек.

Глава 4. Предлагаемое решение

Первичное определение наличия или отсутствия голосового возбуждения

Определение наличия или отсутствия голосового возбуждения является довольно важным признаком, который сразу же позволяет с довольно большой точностью разделить файл на гласные/звонкие согласные и шумные/смычные.

За основу принят метод, предложенный Кочаровым Д.А. в своей диссертационной работе [1]. Необходимость его модификации и изменения некоторых параметров было вызвано тем фактом, что Кочаров работал с более чистым и лучше оцифрованным сигналом.

Суть модифицированного метода заключается в следующем: наличие вокализации в кадре и точные границы вокализованных сегментов в исходном сигнале $S(t)$ вычисляется на основе вычисления мгновенной энергии предварительно отфильтрованного сигнала $SF(t)$.

Для фильтрации применялся полосный фильтр, который оставляет полосу 400-700 Гц.

Это было сделано для того, чтобы по возможности заглушить все частоты, на которых возможно шумовое возбуждение (как низкочастотные шумы паузы, так и высокочастотные шумы шумового возбуждения речи). На озвонченных участках остается мощная составляющая. На неозвонченных участках энергия будет практически равна нулю. После фильтрации вычислялась энергия E сигнала $SF(t)$ по небольшому участку (в поккадровой постановке задачи целиком по кадру в 92 отсчета, при определении точной границы по окну длиной 10 мс с шагом в 1 мс). Далее по большому участку сигнала (в пределах 2 с) вычислялось среднее арифметическое значение мгновенной энергии EE . Решение о том, вокализовано ли окно (кадр) или нет принималось на основании вычисления отношения между мгновенной энергией E и средней мгновенной энергией EE .

Окно (кадр) считается вокализованным, если $E > T * EE$

Окно (кадр) считается невокализованным, если $E < T * EE$, где T - коэффициент, определяющий пороговое значение.

Выявленные проблемы

Проведенные эксперименты выявили довольно сильную вариативность оптимального порогового значения (от 0.01 до 0.1). Возможным объяснением подобного феномена может служить следующее: перед началом сегментации сигнал приводится к одному уровню громкости. Учитывая то, что уровень паузы и так сильно варьируется от файла к файлу, приведение к общему уровню только усиливает эту вариативность. Для

вокализованных участков, которые, как правило, являются высокоэнергетичными, подъем уровня энергии паузы существенного влияния не оказывает, а вот на участки, где голоса нет (и соответственно общий уровень энергии не сильно велик) подъем уровня энергии паузы существенное влияние оказывает.

Разделение смычных и глухих щелевых

После первичного разделения голоса на звонкие и глухие участки, в глухих участках остались щелевые и смычные согласные. Отличительной особенностью щелевых согласных является наличие высокоэнергетичного высокочастотного шума. Подобным свойством смычные не обладают [3]. Для того, чтобы зафиксировать высокочастотные колебания анализировалась статистика среднего количества переходов сигнала через ноль. Предварительно сигнал центрировался. Для того, чтобы подавить низкоэнергетический шум паузы сигнал центральноограничивался (отсчетам, значение которых меньше чем некое пороговое значение T присваивалось значение 0). Таким образом на смычных участках отсеивались низкочастотные шумы паузы. На подобным образом модифицированном сигнале считается статистика переходов через ноль. Отсчеты считаются щелевыми, если количество переходов через ноль больше некоторого порога T .

Выявленные проблемы



Рис 3. Явление аспирации после взрыва

Отсчет 17940 – 18400 – глухой смычный звук ‘k’, 18400-18520 – начало гласного ‘o’.

18210 – 18230 – взрыв, 18230 – 18400 – затухание взрыва, фаза аспирации (170 отсчетов \approx 21 мсек.)

Проблему при разделении смычных (особенно глухих смычек) и щелевых представляет так называемая фаза аспирации (рис. 3), которой может заканчиваться взрыв после смычки. Физически фаза аспирации очень похожа на тихую щелевую. В отличие от смычки, где практически ничего нет (по сути уровень паузы), на фазе аспирации явно просматривается наличие энергии на высоких частотах шумового характера. По длительности аспирация тоже может достигать длительности щелевой. Были

предположения, что определение наличия или отсутствия взрыва является второстепенной задачей, в силу того, что любая смычка обязана закончиться взрывом. Однако подобный фонетический феномен как фаза аспирации показал, что без определения наличия взрыва (что не является тривиальной задачей) определить, является ли шум после паузы разрешением взрыва или уже другой согласной, не представляется возможным

Разделение глухой и звонкой смычки

Данные классы довольно сильно отличаются. Если глухая смычка по своей сути практически является паузой (если не брать во внимание фазу взрыва), то звонкая смычка характеризуется наличием небольшого числа гармоник основного тона (обычно 2-3, редко 4) без формантной структуры. Причем практически вся энергия сосредоточена ниже 500 герц. Различать подобные классы предлагается путем сравнения энергии в полосах 0-500 и 500-4000 Гц. На отсчетах звонких смычек энергия в полосе до 500 Гц должна значительно преобладать над энергией в полосе 500-4000 Гц. На участках паузы глухих смычных энергии нет практически по всей полосе 0-4000 Гц. На участках взрыва энергия распределена равномерно так же по всей полосе.

Выявленные проблемы

На рис.4 продемонстрировано явление остаточного озвончения на глухих звуках после гласной. Физика данного процесса заключается в том, что после гласной голосовые связки перестают колебаться, но резонаторы затухнуть не успевают. Это вызывает формирование гармонических колебаний, которые постепенно уходят под уровень паузы. Проблема заключается в том, что колебания, получаемые таким образом очень похожи на картину, характерную для звонкой смычки (1-3 гармоники, основная энергия на частотах до 500 Гц). Как правило, подобные участки опознаются как звонкие смычные, хотя таковыми не являются. Следует отметить, что комбинация звонкого смычного, за которым следует глухой смычный или глухой щелевой крайне редка, особенно в русском языке. Это позволяет записывать подобные «звонкие смычки» для русской речи в последующий глухой сегмент, пренебрегая производимой ошибкой.

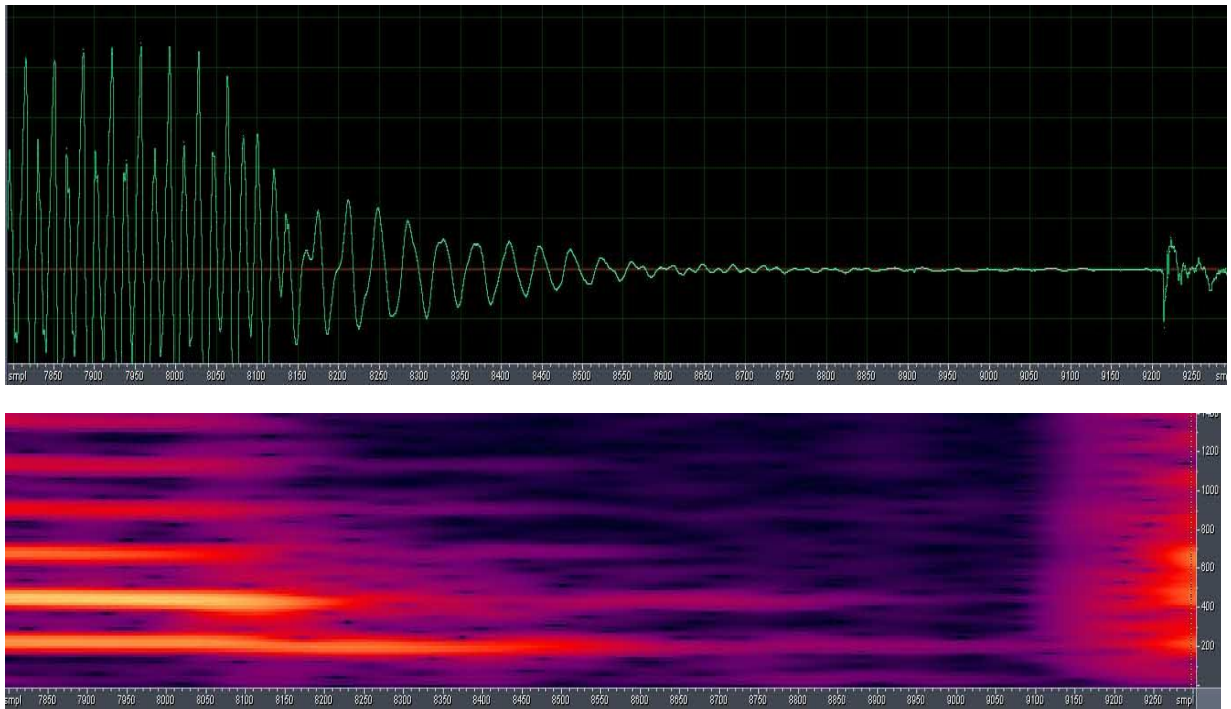


Рис 4. Явление остаточного озвончения

До 8150 отсчета – гласный ‘а’, 8150 – 8900 – глухой смычный ‘т’, 8900 – 9250 – глухой смычный ‘р’

8150 – 8600 – остаточное озвончение после гласной. На спектрограмме явно видно шлейф первой гармоники, который затухает довольно медленно. Остальные гармоники затухают довольно быстро

Разделение гласных и звонких согласных

Перед началом данной курсовой работы предполагалось, что удастся однозначно разделить эти два фонетических класса. Впоследствии, после начала работ и консультаций с фонетистами, оказалось, что класс звонких согласных не является однородным. По сути, оказалось, что звонкие согласные – всё звонкое, что не является гласными. Причем некоторые представители (например, ‘L’) практически неотличимы от гласных, а некоторые (например, ‘Z’) отличаются крайне сильно (можно видеть на рис. 5).

На данный момент единого признака, позволяющего с приемлемой точностью отличить гласные и звонкие согласные, найдено не было. После нескольких поставленных экспериментов была выявлена статистика, позволяющая практически безошибочно определить согласные звуки, которые слева и справа были окружены гласными.

Метод заключался в следующем: считалась статистика «глубины» максимумов. Отсчету i , где фиксировался максимум, присваивалось следующее значение:

$$s[i] = |s[i] - s[k1]| + |s[i] - s[k2]|$$

где $k1$ – отсчет, на котором фиксировался минимум, предшествующий максимуму i , $k2$ – отсчет, на котором фиксировался минимум, следующий за максимумом i . Остальным отсчетам присваивалось значение 0.

Оказалось, что локальные минимумы подобной статистики как раз и расположены на согласных, зажатых между гласными. Подобным образом были выделены довольно точно приблизительно треть согласных

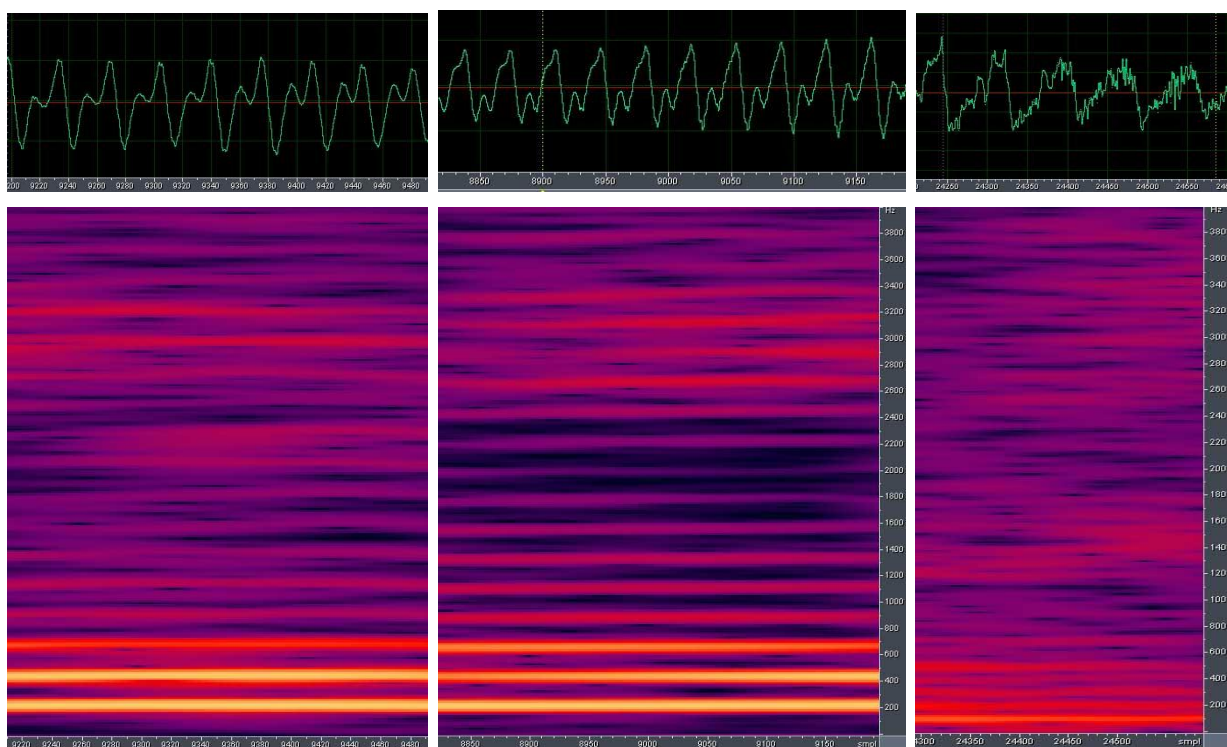


Рис 5. Слева на право: гласная фонема 'i', согласная фонема 'l', согласная фонема 'z'

Глава 5. Алгоритм реализации

1. Первичное разделение на вокализованные и невокализованные участки.

Модифицированный алгоритм Кочарова

АЧХ фильтра :

Значение	0	0.5	1	1	0.5	0
Частота(Гц)	0-300	400	500	600	700	800-4000

400-700 Гц – полоса 1-й форманты. Этим обусловлен ее выбор.

Порог - 0.015. Получен эмпирическим путем. На нем показаны наилучшие результаты при обработке большого числа файлов (2100 файлов, 10 дикторов). Результатом является разделение сигнала на шумные и звонкие кадры

2. Выделение глухих щелевых согласных из шумных

У исходного сигнала все значения, которые по модулю $< 1/50$ максимального значения сигнала объявляются нулями (центральная фильтрация). По этому сигналу считается статистика среднего кол-ва переходов через ноль в кадре (zerosStat1).

Исходный сигнал фильтруется фильтром с АЧХ:

Значение	1	0.5	0
Частота(Гц)	0-300	400	500-4000

У отфильтрованного сигнала все значения, которые по модулю $< 1/30$ максимального значения отфильтрованного сигнала объявляются нулями. По этому сигналу считается статистика среднего кол-ва переходов через ноль в кадре (zerosStat2). Шумный кадр считается щелевой согласной, если по кадру $\text{zerosStat2} / \text{zerosStat1} < 0.7$. Смысл в том, что отрезая верхи фильтром, мы добиваемся существенного уменьшения кол-ва переходов через ноль именно на щелевых. Этот факт мы и отлавливаем. Результатом есть разделение шума на смычные и глухие щелевые.

3. Разделение глухой и звонкой смычки

Исходный сигнал фильтруется фильтром с АЧХ:

Значение	0	0.5	1	1	1	0.5	0
Частота(Гц)	0	100	200	300	400	500	600-4000

Шумный кадр считается глухим смычным, если отношение средней энергии отфильтрованного сигнала к средней энергии всего сигнала < 0.65 . Таким образом мы отлавливаем тот факт, что у звонких смычных практически вся энергия сосредоточена ниже 500 Гц. Таким образом, фильтрация практически не затронет звонкие смычные.

4. Выделение звонких смычных из глухих смычных и щелевых

Необходимость подобного действия заключается в том, что небольшая часть смычек оказалась не в своем классе. Вызвано это жесткими значениями порогов.

Исходный сигнал x фильтруется фильтром с АЧХ:

Значение	0	0.5	1	1	0.5	0
Частота(Гц)	0	100	200	300	400	500-4000

Щелевой или гл. смычный кадр считается звонким смычным, если отношение энергии отфильтрованного сигнала к энергии всего сигнала > 0.5

5. Выделение из голоса глухих смычных

Первичное разделение происходило по энергетическому признаку. Незначительная часть смычных оказалась в голосе.

Исходный сигнал фильтруется фильтром с АЧХ:

Значение	1	1	1	1	0.5	0
Частота(Гц)	0	100	200	300	400	500-4000

У отфильтрованного сигнала все значения, которые по модулю $< 1/50$ максимального значения отфильтрованного сигнала объявляются нулями. По этому сигналу считается статистика среднего кол-ва переходов через ноль в кадре.

Звонкий кадр, у которого кол-во переходов через ноль < 2 – глухая смычка

6. Блок уточнения:

- 1) Если $i-1$ й кадр = $i+1$ й кадру, а i -й кадр от них отличается, то i -й = $i+1$
- 2) Если i -й кадр не равен ни $i-1$ ни $i+1$ и $i+1 = i+2$ и $i-1$ – голос – то $i = i+1$
- 3) Если i -й кадр не равен ни $i-1$ ни $i+1$ и $i-1 = i-2$ и $i+1$ – голос – то $i = i-1$

4) Если за звонкой смычкой следует глухая смычка – то все кадры звонкой смычки становятся глухими смычками (борьба с остаточным озвончением)

7. Выделение звонких согласных из голоса(частично)

По кадру считается «статистика максимумов»

Кадр, где наблюдается локальным минимум и его окрестность считается звонкой согласной.

Результаты

Были поставлены эксперименты по сегментации большого числа файлов с помощью реализованного автомата (2150 файлов от 10 дикторов, 5 мужчин и 5 женщин). Точность сегментации составила чуть меньше 70% (табл. 1).

Эталон\Автомат	Гласные	Зв.Согласные	Гл.Щелевые	Зв.Смычные	Гл.Смычные
Гласные	28.8000	1.3000	0	0.2000	1.3000
Зв.Согласные	11.8000	5.0000	0.2000	0.6000	0.9000
Гл.Щелевые	1.0000	0.1000	6.4000	0	3.7000
Зв.Смычные	1.4000	0.3000	0	3.2000	0.7000
Гл.Смычные	1.9000	0.3000	4.2000	0.7000	25.8000

Табл. 1 «Матрица спутывания» работы автомата сегментации. Ячейка [n,m] означает, что k% отсчетов, размеченных вручную как класс n, были распознаны автоматом как класс m.

Переход класс в класс: 69.2000%

Представлены усредненные результаты обработки 2150 файлов от 10 дикторов, 5 мужчин и 5 женщин.

Основные ошибки были сконцентрированы в гласных/звонких согласных, а так же щелевых/смычных. Следует отметить, что эксперименты, поставленные по сегментации с помощью НММ, показали приблизительно такие же результаты. Суть эксперимента сводилась к следующему: существующему распознающему автомату подавался сигнал. После распознанные фонемы сливались в классы. Точность подобной «сегментации» оказалась так же порядка 70%. Примечательно то, что сегментацию, полученную с помощью НММ, существенно улучшить не представляется возможным. Предложенный же в рамках данной курсовой работы подход может быть подвергнут улучшению сравнительно легко: как за счет отказа от фиксированных порогов – попытки адаптивной настройки порогов, так и за счет проведения дополнительных исследований, направленных на выявление дополнительных статистик, которые позволили бы снизить количество ошибок в определенном классе.

Направление дальнейшей работы

Выбранная методология решения, а так же сама структура предлагаемого автомата открывает большой простор для улучшения полученных результатов. Возможности для улучшения:

- Привлечение новых данных, исследование новых статистик, улучшение старых
 - Первичное выделение пауз. Учет средней энергии паузы в энергетических критериях.
 - Выделение и опознание взрыва: возможность более точно определять смычные, возможность опознания «фазы аспирации»
 - Более глубокий анализ вопроса гласные/звонкие согласные
- Адаптивная настройка пороговых значений статистик: попытка настройки порогов путем анализа, например, гистограммы значений статистик по файлу
- Наличие автомата, который сможет с необходимой точностью (до 1 мсек) выделять моменты прихода импульсов основного тона в том числе и на зашумленных участках (например, звонкие согласные), даст возможность существенно уточнить границы. Основная идея заключается в том, что принято считать, что границы фонем проходят по импульсам основного тона. Таким образом, зная отсчеты прихода импульсов, будет возможно корректировать границы фонем, сдвигая их к ближайшему импульсу ОТ.

Предполагается, что подобным образом удастся улучшить результаты до уровня 90%.

Список литературы

1. Кочаров Д.А. Автоматическая интерпретация звуков речи // Диссертационная работа СПбГУ 2008
2. Карабалаева М.Х., Шарипбаев А.А. Алгоритмы пофонемного распознавания казахской речи в амплитудно-временном пространстве // Знания-Онтологии-Теории (ЗОНТ-09)
3. В.Н. Сорокин, А.И. Цыплихин. Сегментация и распознавание гласных // Информационные процессы, Том 4 , № 2, стр. 202-220
4. Г. Фант. Акустическая теория речеобразования // Издательство «Наука», 1964 г.
5. М.Ф. Деркач, Р.Я. Гумецкий, Б.М. Гура, М.Е. Чабан. Динамические спектры речевых сигналов // Издательство при львовском государственном университете издательского объединения «Вища школа», 1983 г.
6. В.В. Митянок. Метод аппроксимации для определения числовых характеристик некоторых низкочастотных звуков человеческой речи // Электронный журнал «Техническая акустика» <http://www.ejta.org> 2008, 15
7. В.И. Галунов, Ю.Л. Голованова, М.И. Орлова, Б.Н. Форш, Н.Н. Ягунова. Использование различных способов первичного описания речевых сигналов в задачах автоматического распознавания речи // Автоматическое распознавание слуховых образов, Том 3, стр. 301-303, Киев 1982 г.