УДК 004.394.2

О.Н. Ладошко, аспирант
А.Н. Продеус, канд. техн. наук, доц.,
Национальный технический университет Украины
«Киевский политехнический институт», г. Киев, Украина
[1]ladoshko@gmail.com, [2]aprodeus@gmail.com

Оценка надежности выделителя частоты основного тона для акустического анализа речи

Проведено тестирование детектора частоты основного тона (ЧОТ). Рассмотрены критерии сравнения качества работы подобных алгоритмов. Полученные результаты могут быть использованы для оценки акустических характеристик спонтанной речи в задаче автоматического стенографирования.

Ключевые слова: надежность, частота основного тона, ларингограф, асинхронное сравнение, частотный контур, речевой сигнал.

Введение

Известно, что главной проблемой автоматического стенографирования речи является существенное отклонение разговорной речи человека от грамматических и лексических правил письменного языка [1]. Такое различие обусловлено наличием в спонтанной речи особенностей, классификация которых представлена в [2]. Одной из наиболее распространённой особенностью спонтанной речи человека является вокализованная пауза (ВП), представленная всевозможными вставками в виде протяженного произнесения звуков («э-э», «а-е» и др.) и слов. Эти вставки являются одной из причин ухудшения надёжности системы автоматического распознавания слитной речи, не адаптированной к распознаванию спонтанной речи. В работе [1] были проведены эксперименты по распознаванию спонтанной украинской речи, в которой показано, что ручная очистка спонтанной речи от её особенностей, включая и значительную часть ВП, позволяет существенно улучшить показатели надёжности распознавания речи в среднем от 1,25% до 6,45% для разных исследуемых выборок.

До настоящего времени способы улучшения надежности распознавания спонтанной украинской речи сводились к моделированию языковой модели спонтанной речи на основе биграммной модели речи и создания правил индивидуализированного транскрибирования [2]. Несмотря на некоторое улучшение надёжности распознавания при моделировании языковой модели, до настоящего времени не существует единого мнения о достаточном количестве индивидуализированных транскрипций для достижения требуемого показателя надежности распознавания спонтанной речи. Таким образом, может быть поставлена задача оценки характеристик речевых сигналов, не зависящих от лексического состава речи (слов и фонем), в реальных условиях записи спонтанной речи. Очевидно, для решения такой задачи необходим специальный программный инструментарий. В частности, для автоматического обнаружения ВП необходимо располагать «детектором ЧОТ» - системой автоматической оценки и анализа изменяющейся во времени частоты основного тона (траектории ЧОТ).

В данной работе представлены результаты исследований детектора ЧОТ, построенного с учетом рекомендаций работы [3]. Анализ результатов работы детектора ЧОТ состоит в оценке правильности определения вокализованности (voiced) или невокализованности (unvoiced) речи, а также определения процента грубых ошибок на участках, где отсутствуют ошибки voiced- unvoiced. Кроме того, можно предположить, что исследование траекторий ЧОТ станет дополнительным источником информации для обнаружения некоторых особенностей спонтанной речи.

Целью исследования является исследование детектора ЧОТ с целью получения робастного частотного контура речевого сигнала.

Задачей исследования является исследование оптимальных параметров детектора ЧОТ для оценки акустических характеристик речи, записанной в реальных условиях работы системы автоматического стенографирования.

Методы и средства проведения исследований

Предметом данного исследования является один из основных параметров устной речи – частота колебаний голосовых связок при произнесении вокализованной речи, называемая основным тоном – F_0 (величина обратная - период $T_0 = 1/F_0$) [4, 5, 6]. К настоящему времени разработано множество методов оценки ЧОТ [3, 4, 6], однако, каждый из них порождает наименьшее число ошибок лишь для области его дальнейшего использования. Как утверждается в [6], методы, работающие «во временной области», т.е. с выборками речевого сигнала, обладают наименьшей, по сравнению с другими методами (спектральными, кепстральными), ошибкой принятия решения о присутствии голоса в речи (voicing decision error rate) – не более 17%.

Кроме того, в работе [7] показано, что такие методы являются наиболее робастными в отношении принятия решения о вокализованности или невокализованности сегмента речи в условиях шума (voiced-unvoiced decision), искажений и побочных помех в сигнале. Поэтому для изучения акустических характеристик речи в данной работе программно реализован детектор ЧОТ, построенный с учетом рекомендаций работы [3].

Общая структурная схема моделируемого детектора ЧОТ приведена на рис. 1:

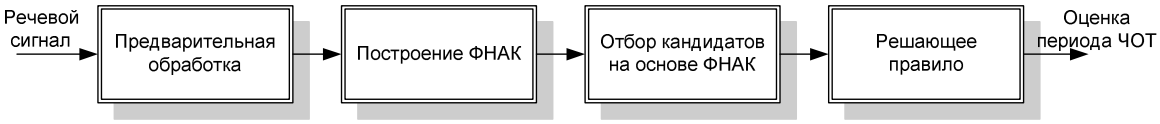

Рисунок 1 – Общая структурная схема выделителя ЧОТ

Метод, выбранный в качестве базового для построения детектора ЧОТ, основан на обработке речевого сигнала во временной области в окне анализа. Длина окна анализа выбирается достаточно длинной для надёжного обнаружения периодичности (минимум два периода ожидаемой ЧОТ), и достаточно малой для исключения преднамеренного сглаживания возможных изменений ЧОТ в окне.

Наилучшую стабильность анализа речевых сигналов обеспечивают интервалы анализа продолжительностью $30-50$ мс с шагом анализа $10-20$ мс [8].

Предварительная обработка речевого сигнала заключается в ручном вырезании всех длительных пауз и пропускании речевого сигнала через фильтр низких частот (ФНЧ) для упрощения временной структуры сигнала за счет устранения влияния высших формат ($F2, F3$ и $F4$). Такое упрощение временной структуры сигнала помогает алгоритму правильно классифицировать вокализованную речь.

Используемый ФНЧ, с частотой среза 1200 Гц, реализован на основе эллиптического фильтра (фильтр Кауэра). За счет возможности регулировки пульсаций амплитудно-частотной характеристики эллиптического фильтра в полосе пропускания и задержания удаётся обеспечить максимально возможную крутизну ската амплитудно-частотной характеристики.

Экспериментальные исследования показывают, что параметры фильтра предварительной обработки сигнала непосредственно не влияют на точность получения оценок ЧОТ. Этот вывод согласуется с данными работы [6].

Для преобразования сигнала в удобную для анализа форму используется один из оптимальных методов оценки ЧОТ, основанный на определении максимума автокорреляционной функции [9, 10] речевого сигнала в заданных границах его поиска.

Границы поиска ЧОТ задают реальные диапазоны изменения ЧОТ для мужских и женских голосов. В работе [6] были проведены эксперименты по измерению ЧОТ дикторов различного пола, с помощью устройства, фиксирующего сигнал смыкания и размыкания голосовых складок человека (ларингографа). Полученные статистические данные показали, что диапазон изменения ЧОТ для мужских голосов составляет 50-250 Гц, а женских – 120-400 Гц (рис. 2).

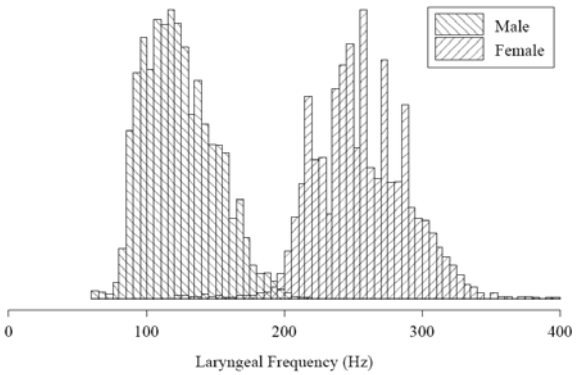

Рисунок 2 – Гистограммы ларингеальной частоты для мужской (male) и женской (female) речи [6]

Именно эти диапазоны изменения ЧОТ (от $p_{\min}$ до $p_{\max}$) служат границами поиска максимума функции нормированной автокорреляции (ФНАК) речевого сигнала.

Предполагая, что необходимо произвести оценку ЧОТ для сигнала $s(n)$, где n - индекс времени, а N - длина окна, вычисляется функция нормированной автокорреляции:

$$R(p)=\frac{\sum_{n=1}^{n-p} s(n)s(n-p)}{\sqrt{\sum_{n=1}^{n-p} s^2(n)s^2(n-p)}} \quad (1)$$

Диапазон сдвигов p выбирается таким образом, чтобы охватить диапазон реальных значений ЧОТ (рис.2).

Вычисляя значения ФНАК для полного набора $p\in[p_{\min},p_{\max}]$, находим такое значение сдвига p, при котором значение ФНАК максимально. Значение найденного таким образом сдвига и будет периодом (частотой) основного тона речевого сигнала $s(n)$.

Для построения сглаженной по кадрам анализа траектории ЧОТ, оценки положительных локальных максимумов ФНАК, полученные для отдельных кадров k, рассматриваем в виде набора возможных кандидатов $\{p_{m_k}\}$.

Отбор кандидатов на основе ФНАК в отдельном кадре анализа k проводиться путём умножения значения ФНАК на монотонно убывающую функцию

$$\exp(-a\cdot t), \quad (2)$$

с последующей сортировкой и отбором наибольших четырёх значений ФНАК.

Данная операция предназначена для увеличения вероятности выбора кандидата в ЧОТ с низшим значением ЧОТ и, тем самым, уменьшения вероятности выбора «ложных» пиков ФНАК на высоких частотах.

В качестве меры правдоподобия оценки ЧОТ значением p_{m_k}, при сравнении кандидатов $\{p_{m_k}\}$ для текущего кадра анализа выбраны величины локальных максимумов $R(p_m)$.

Решающее правило формулировалось как задача поиска пути методом динамического программирования. Таким образом, для получения наиболее вероятной траектории ЧОТ проводится анализ кандидатов в ЧОТ для группы смежных кадров:

$$\{k,k+1,...,k+K+1\}. \quad (3)$$

Наборы кандидатов смежных кадров образуют столбцы решетки, через узлы которой прокладывается траектория ЧОТ. Для каждого узла решетки m_k задается стоимость, которая выбирается пропорциональной величине локального максимума ФНАК:

$$d_N(m_k)=R_k(p(m_k)) \quad (4)$$

Исходя из того, что траектория изменения ЧОТ для вокализованных звуков считается плавной линией, мы учитываем эту особенность вокализованных звуков тем, что определяем вероятность изменения траектории ЧОТ путем введения функции стоимости пути, учитывающей расстояние между узлами решетки от кандидата $p(m_k)$ кадра k к кандидату $p(m_{k+1})$ кадра $k+1$.

Вероятность изменения траектории ЧОТ от кадра к кадру задаётся функцией стоимости пути, с регулируемым порогом α, который ограничивает возможные отклонения траектории ЧОТ для смежных кадров:

$$d_T(m_k,n_{k+1})=\begin{cases}0, & |p(m_k)-p(n_{k+1})|\le\alpha\cdot p(m_k)\\ -\infty, & |p(m_k)-p(n_{k+1})|>\alpha\cdot p(m_k)\end{cases} \quad (5)$$

В результате, оценка наиболее вероятной траектории ЧОТ через K кадров сглаживания, осуществляется выбором оптимального пути между узлами решетки

$$m_k, m_{k+1}...m_{k+K-1}, \quad (6)$$

максимизирующего функционал общей стоимости пути вида (7), где k – индекс кадра, m_k – индекс кандидата $p(m_k)$ на оценку ЧОТ для кадра k, K – количество смежных кадров анализа, по которым проводиться сглаживание траектории.

Исследование и выбор параметров детектора ЧОТ, а именно порогов a, α, K, которые задают основные особенности работе детектора ЧОТ для двух различных дикторов, будут рассмотрены далее.

$$D_L(m_k,m_{k+1},...m_{k+K-1})=\sum_{i=0}^{K-2}(d_N(m_{k+i})+d_T(m_{k+1},m_{k+i+1}))+d_N(m_{k+K-1}) \quad (7)$$

Результаты проведения экспериментов

Особенности предлагаемой реализации метода выделения ЧОТ и влияние плохо формализуемых факторов на работу выделителя ЧОТ приводит к тому, что оценку качества измерений частоты основного тона проводят экспериментально [3, 4, 7, 6]. Схема проведения экспериментов показана на рис. 3.

Для стандартизации и возможности сравнения результатов тестирования предложенного алгоритма использовались исходные коды для работы с тестовой базой [6]. Данная база эталонных сигналов состояла из 50 фраз, произнесенных одним диктором мужского и одним диктором женского пола без патологий в голосе. Речь была записана с помощью микрофона ближнего действия и ларингографа в заглушенной студии.

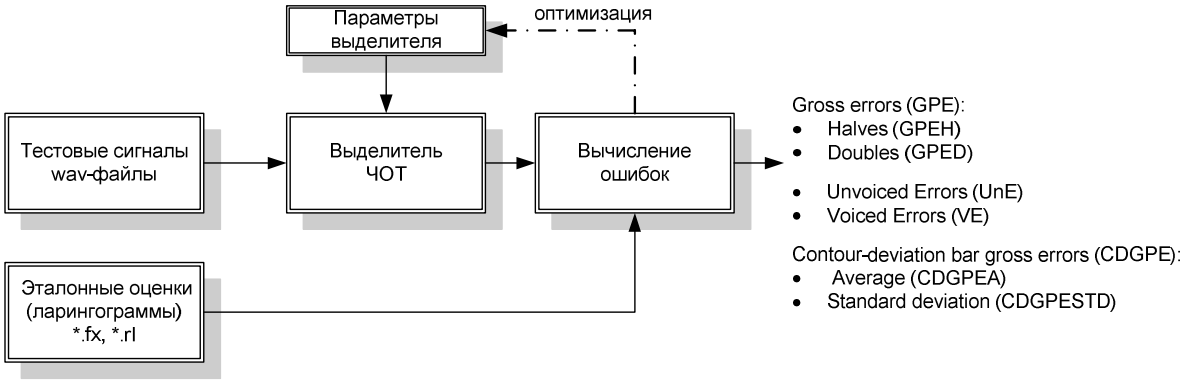

Рисунок 3 – Схема оценки надёжности выделителя ЧОТ для акустического анализа речи.

Обработанный сигнал, зафиксированный ларингографом [6], может служить эталоном для проверки надежности работы предлагаемой модели выделителя ЧОТ.

Для возможности сравнения результатов работы реализованного метода в виде контура ЧОТ и эталонных значений ЧОТ ларингографа [6], предопределенные параметры, детектора ЧОТ, были установлены такими: длинная окна анализа $38{,}4$ мс с шагом анализа $6{,}4$ мс.

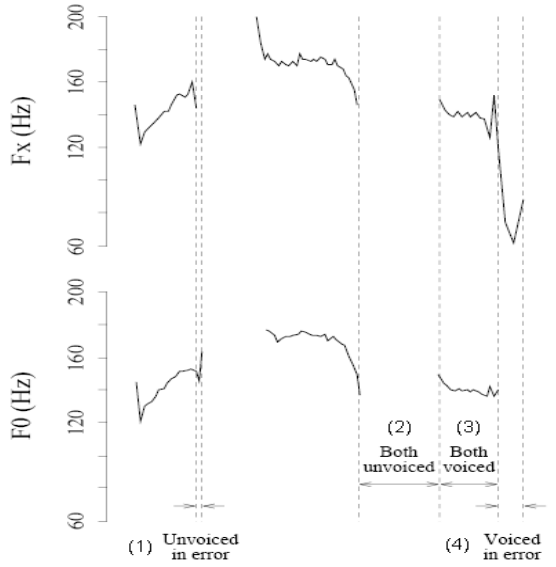

Рисунок 4 – Сравнение асинхронных контуров F_x - эталонный и F_0 - тестовый.

Ошибка квантования эталонного контура [6] F_x в среднем составила 0,80 Гц и 3,33Гц и стандартное отклонение генеральной совокупности 0,34 Гц и 0,86 Гц для мужчин и женщин соответственно. Эти ошибки не могут быть компенсированы.

Известные проблемы оценки надежности и нахождения объективных критериев сравнения работы выделителей ЧОТ [7], приводят к необходимости использовать некоторые из них, а именно: оценку точности в определении периода ОТ (ЧОТ) и точности принятия решения о наличии-отсутствии голоса (voiced-unvoiced decision) [7, 6].

Таким образом, сравнение эталонного контура F_x и тестируемого контура F_0 осуществляется по четырем критериям (рис. 4, п.1-4 [6, 7]):

1. Unvoiced Errors (UnE). Если $F_0 \neq 0$, а $F_x = 0$, то невокализованный (unvoiced) участок речи неверно классифицируется как вокализованный (voiced) тестируемым алгоритмом. Длительность области ошибки определяется путём поиска следующего момента времени, в котором либо контур $F_0 = 0$ либо $F_x \neq 0$.
2. Если оба контура $F_x = 0$ и $F_0 = 0$, то оба контура описывают тихий или невокализованный участок предложения. В таком случае ошибки не фиксируются.

3. Gross errors (GPE). Если $F_x \neq 0$ и $F_0 \neq 0$, то оба контура правильно описывают вокализованную речь. В этом случае вычисляется процент грубых ошибок – gross errors (GPE) из соотношения:

$$GPE = \frac{F_x - F_0}{F_x} \quad (8)$$

Halving Errors (GPEH). Если $GPE > 0{,}2$, то процент грубых ошибок выдаваемых выделителем ЧОТ составляет не менее 20%. При этом ошибку относят к категории Halving Errors, т.е. отклонения в виде уменьшения ЧОТ относительно эталонного значения F_x.

Doubling Errors (GPED). Если $GPE < -0{,}2$, то процент грубых ошибок выдаваемых выделителем ЧОТ относится к категории отклонения в виде увеличения значения ЧОТ относительно эталонного значения F_x.

В остальных случаях предполагается, что оценка ЧОТ получена с приемлемой точностью. Порог $\pm 20\%$ выбран из соображений учета ошибки квантования и конечного разрешения по частоте применяемых методов.

4. Voiced Errors (VE). Если $F_x \neq 0$, а $F_0 = 0$, то вокализованный участок речи неверно классифицируется как невокализованный (тихий). Длительность участка с этой ошибкой определяется путём поиска следующего момента времени, в котором либо $F_x = 0$, либо $F_0 \neq 0$.

Синхронизация временных отсчетов и соответствующих значений ЧОТ при сравнении асинхронных F_0 и F_x контуров осуществляется путем линейной интерполяции недостающих значений полиномом Лагранжа [6].

Длительности невокализованных и вокализованных участков речи, классифицированных как ошибки, суммировались по всей базе для каждого диктора в отдельности. Сумма ошибок, соответственно вокализованности и невокализованности выражалась как процент от общей длительности соответственно вокализованной и невокализованной речи в получаемом F_0 контуре. GPE, выраженная в процентах определялась для всех $F_x \neq 0$ и $F_0 \neq 0$. Конечная статистика также включает в себя оценку стандартного отклонения генеральной совокупности (population standard deviation (p.s.d.)) и средней (average), абсолютной девиации F_x и F_0 контуров, когда оба контура представлены вокализованной речью и выделитель ЧОТ не делает ошибок GPE в оценке ЧОТ.

Результаты сравнения каждой пары контуров усреднялись на наборе предложений каждого из дикторов.

При возникновении значительных ошибок (100% ошибок GPED или GPEH) в какой-либо из 50 фраз, приводящих к смещению результата усреднения, установленные параметры выделителя ЧОТ считались неоптимальными.

Таким образом, определялись такие оптимальные параметры выделителя ЧОТ, при которых не должно происходить смещение значений усреднённых по выборке ошибок в сторону наихудшей ошибки.

В таблице 1 и таблице 2 представлены результаты тестирования детектора ЧОТ по заданным критериям в зависимости от двух различных значений порога $\alpha = 1{,}1$ и $\alpha = 1{,}2$ (выражение (5)), задающего относительные границы поиска отклонения траектории ЧОТ для смежных кадров для мужского голоса. Сглаживание траектории производиться по $K = 5$ кадрам анализа с выдачей оценки ЧОТ для среднего кадра.

Из таблицы 1 и таблицы 2 видно, что наименьшие ошибки voiced-unvoiced и GPE соответствуют пороговому значению $\alpha = 1{,}1$. При этом все ошибки увеличиваются с увеличением порога до $\alpha = 1{,}2$. Поэтому в качестве оптимального значения порога α выбрано значение 1,1.

Наименьшая суммарная ошибка voiced-unvoiced для мужчин составляет 23,9%, а для женщин – 37,1%, при этом GPE для женщин достигает меньших значений – 0,8, чем у мужчин (табл.1). Данные показатели указывают на то, что проблема правильного определения вокализованных участков для женских голосов существеннее, чем для мужских голосов

Экспериментальным путём обнаружено, что увеличение порога изменения траектории ЧОТ для мужчин (табл.3) приводит к постепенному ухудшению количества ошибок принятия решения о наличии или отсутствии голоса во фрагменте речи с $\alpha = 1{,}1$ до $\alpha = 2{,}0$, в отличии от женских (табл.4) – с $\alpha = 1{,}1$ до $\alpha = 1{,}7$. Для женских голосов видно, что при $\alpha = 1{,}7$ происходит существенное увеличение ошибок правильной классификации невокализованных участков 101,2% (табл. 4).

Значение ошибок unvoiced (UnE – Unvoiced in Error) для мужского голоса изменяется в диапазоне от 9,4% до 92,1%, при этом ошибки типа voiced (VE – Voiced in Error) изменяются в диапазоне от 14,5% до 2,6%. Аналогичную связь наблюдаем и для женского голоса. Таким образом, увеличение количества ложно

определённых вокализованных (voiced) участков речи изменяется обратно пропорционально количеству неверно классифицированных невокализованных (unvoiced) участков речевого сигнала, как для мужского, так и для женского голоса (таблица 3 и таблица 4). Кроме того одновременно наблюдается увеличение процента грубых ошибок обоих типов GPED и GPEH.

Средняя абсолютная девиация эталонного и тестируемого контура при отсутствии выше описанных ошибок GPE двух типов не превышает 3,5 Гц для мужского голоса (таблица 3) и 6,5 Гц для женского голоса (таблица 4).

По результатам экспериментов для мужчин оптимальными являются значения $\alpha = 1{,}1$ и $a = 1{,}2$, а для женщин $\alpha = 1{,}1$ и $a = 0{,}1$. При данных значении достигаются наименьшие усредненные по речевой базе ошибки всех типов.

В таблице 1 и таблице 2 рассмотрены зависимости предустановленных порогов α и различных коэффициентов a регулировки монотонной убывающей кривой, для женского и мужского голоса. Данные для коэффициента a, представлены в таблице 1 и таблице 2 начиная со значения 1,0 (табл. 1) и 0,5 (табл. 2). Для остальных $a < 1{,}0$ и $\alpha \geq 1{,}1$ (табл. 1) и $a < 0{,}5$ и $\alpha \geq 1{,}1$ (табл. 2) экспериментально установлено, что значения ошибок увеличиваются по всем рассматриваемым критериям.

По результатам из таблицы 1 обнаружено, что ошибки становятся минимальными при $\alpha = 1{,}1$ и $a = 1{,}2$, при этом суммарное значение GPE = 1,0%, а суммарный процент ошибок voiced-unvoiced составил наименьшее значение 23,9%.

По результатам таблицы 2 видно, что наименьшие значения ошибок достигаются при значениях $\alpha = 1{,}1$ и $a = 0{,}1$ – GPE = 0,8%, а voiced-unvoiced – 37,1%. При дальнейшем уменьшении a до 0,0 наблюдается рост общего числа ошибок voiced-unvoiced до 39,4% и суммарного процента грубых ошибок GPED и GPEH до 1,2%.

Экспериментально выявлено, что оптимальное значение порога изменения траектории от кадра к кадру для получения минимального процента ошибок у мужчин и женщин совпадают – $\alpha = 1{,}1$.

Кроме того выявлено, что для женских голосов наблюдается большее абсолютное отклонение от эталонного контура (около 6 Гц), чем для мужчин (около 3 Гц) на участках контура, где отсутствуют ошибки GPE и voiced-unvoiced.

Рассматривая ошибки типа GPEH и GPED (табл. 1, 2, 3, 4) можно сказать, что исследуемый детектор имеет тенденцию к большему значению ошибок вида GPED:

$$GPED = \frac{F_x - F_0}{F_x} < -0{,}2\,, \qquad (9)$$

т.е. ошибок связанных с выбором пика на задержке меньшей, чем задержка, на которой значение ФНАК максимально.

Таким образом, существенным для уменьшения ошибок получения контура ЧОТ, в данном детекторе ЧОТ, является правильность выбора коэффициента a для мужчин и женщин. При этом для женских голосов необходимо подбирать a, при котором бы не происходило искажения амплитуд ФНАК на более высоких частотах, а для мужчин необходимо выбирать a так, что бы исключить возможность выбора ложных высокочастотных пиков. В результате для мужчин коэффициент a должен обеспечивать наибольшую скорость спадания экспоненты $\exp(-a \cdot t)$ для уменьшения вероятности выбора «ложного» пика на задержках кратных времени задержки истинного периода ЧОТ. Тем не менее, выбор ложного пика может быть обусловлен возможной нестационарностью амплитуды пиков ФНАК. Поэтому использование монотонно убывающей функции может также увеличивать и вероятность ошибочного выбора пиков на задержках, меньших периода ЧОТ.

Результаты таблицы 4 показывают, что длина окна сглаживания траектории ЧОТ существенно не влияет на оценку надежности её получения. Таким образом, учет большего контекста для проведения сглаживания траектории, при выбранных, для мужского и женского голоса, оптимальных параметрах выделителя приводит к незначительным изменениям правильного определения вокализованных и невокализованных участков речи, при относительно постоянных значениях GPE как для мужчин, так и для женщин. В данном случае оптимальное значение длинны окна сглаживания траектории необходимо подбирать согласно допустимой временной задержке получения ЧОТ при работе детектора.

Кроме того из результатов таблицы 5 видно, что для женщин преобладают ошибки типа unvoiced (24-27%), в то время как для мужчин основными являются ошибки типа voiced (14-15%).

Таблица 1. Оценки надежности ЧОТ при различных значениях порога α и a (мужской голос)

a	α	UnE, %	VE, %	GPE		CDGPE		$\sum$ GPE , %	$\sum$ V - Un , %
				D, %	H, %	average, Гц	p.s.d., Гц		

1,0	1,1	9,4	15,0	0,5	0,5	3,1	3,2	1,0	24,4
1,0	1,2	17,5	8,5	0,6	0,6	3,2	3,4	1,2	26,0
1,1	1,1	9,4	14,8	0,5	0,5	3,1	3,2	1,0	24,2
1,1	1,2	17,4	8,4	0,7	0,6	3,2	3,4	1,2	25,8
1,2	1,1	9,4	14,5	0,5	0,5	3,1	3,2	1,0	23,9
1,2	1,2	17,6	8,3	0,7	0,5	3,2	3,4	1,2	25,9
1,3	1,1	9,4	14,7	0,5	0,5	3,1	3,2	1,0	24,1
1,3	1,2	17,3	8,4	0,7	0,5	3,2	3,4	1,2	25,7
1,4	1,1	9,4	14,7	0,5	0,5	3,1	3,2	1,0	24,1
1,4	1,2	17,5	8,4	0,7	0,5	3,2	3,4	1,2	25,9
1,5	1,1	9,3	14,8	0,6	0,5	3,1	3,2	1,1	24,1
1,5	1,2	17,6	8,4	0,8	0,5	3,2	3,5	1,3	26,0

Таблица 2. Оценки надежности ЧОТ при различных значениях α и a (женский голос)

a	α	UnE, %	VE, %	GPE		CDGPE		$\sum$GPE, %	$\sum$V - Un, %
				D, %	H, %	average, Гц	p.s.d., Гц		
0,5	1,1	28,7	9,5	0,6	0,3	6,0	6,5	0,9	38,2
0,5	1,2	37,8	6,9	0,7	0,7	0,7	6,7	1,4	44,7
0,4	1,1	27,9	9,8	0,6	0,3	6,0	6,5	0,9	37,7
0,4	1,2	36,8	7,2	0,7	0,4	6,1	6,7	1,1	43,9
0,3	1,1	27,5	9,9	0,6	0,3	6,0	6,4	0,9	37,5
0,3	1,2	36,3	7,3	0,7	0,4	6,1	6,7	1,1	43,6
0,2	1,1	27,2	10,0	0,5	0,3	5,9	6,4	0,9	37,2
0,2	1,2	35,8	7,4	0,6	0,4	6,1	6,6	1,0	43,1
0,1	1,1	26,4	10,7	0,5	0,3	5,9	6,4	0,8	37,1
0,1	1,2	34,9	8,0	0,6	0,5	6,1	6,6	1,0	42,9
0.0	1,1	24,7	14,7	0,5	0,7	6,0	6,3	1,2	39,4

Таблица 3. Оценки надежности ЧОТ при различных значениях порога α и $a = 1{,}2$ (мужской голос)

α	UnE, %	VE, %	GPE		CDGPE		$\sum$GPE, %	$\sum$V - Un, %
			D, %	H, %	average, Гц	p.s.d., Гц		
1,1	9,4	14,5	0,5	0,5	3,1	3,2	1,0	23,9
1,2	17,6	8,3	0,7	0,5	3,2	3,4	1,2	25,9
1,3	25,4	6,5	0,8	0,7	3,3	3,6	1,5	31,8
1,4	33,5	5,6	0,9	0,7	3,3	3,6	1,6	39,2
1,5	41,9	4,9	1,0	0,8	3,4	3,6	1,8	46,8
1,6	51,8	4,4	1,0	0,9	3,4	3,6	1,9	56,2
1,7	60,6	3,9	1,1	0,9	3,4	3,7	2,0	64,5
1,8	71.00	3,5	1,2	0,9	3,4	3,7	2,1	3,5
1,9	80,7	3,2	1,3	0,9	3,4	3,7	2,3	83,9
2,0	92,1	2,6	1,5	1,0	3,4	3,7	2,5	94,7

Таблица 4. Оценки надежности ЧОТ при различных значениях порога α и $a = 0{,}1$ (женский голос)

α	UnE, %	VE, %	GPE		CDGPE		$\sum$GPE, %	$\sum$V - Un, %
			D, %	H, %	average, Гц	p.s.d., Гц		
1,1	26,4	10,7	0,5	0,3	5,9	6,4	0,8	37,1
1,2	34,9	8,0	0,6	0,5	6,1	6,6	1,0	42,9
1,3	45,7	6,2	0,7	0,6	6,2	6,9	1,2	51,8
1,4	59,1	4,7	0,9	0,6	6,3	7,0	1,5	63,8
1,5	70,4	4,1	1,0	0,7	6,4	7,1	1,6	74,4
1,6	83,0	3,5	1,2	0,8	6,4	7,2	1,9	86,5
1,7	98,0	3,2	1,2	0,8	6,5	7,2	2,0	101, 2

Таблица 5. Оценки надежности ЧОТ при различных значениях длины окна сглаживания траектории

Длина окна сглаживания, кол. кадров	UnE, %		VE, %		GPE D+H, %	
	М	Ж	М	Ж	М	Ж
3	9,6	27,1	14,4	10,5	0,9	0,8
5	9,4	26,4	15,0	10,7	1,0	0,8
7	9,5	25,6	14,5	10,9	1,0	0,8
9	9,4	25,4	14,8	11,1	1,0	0,8
11	9,4	24,3	14,8	11,7	1,0	0,8

Заключение

Экспериментально произведенная оптимизация параметров рассмотренного детектора ЧОТ позволяет получать качественные оценки траекторий ЧОТ речевых сигналов при допустимых ошибках их оценки.

Научная новизна работы заключается в оптимизации параметров рассмотренного детектора ЧОТ по заданным критериям качества.
Практическая значимость полученных результатов заключается в возможности повысить качество функционирования рассмотренного детектора ЧОТ за счет устранения грубых ошибок в определении вокализованных и невокализованных участков речи.

Дальнейшее направление исследований связано с устранением ошибок voiced-unvoiced классификации сегментов речи и последующим применением данного алгоритма для анализа записей устной речи, в условиях эксплуатации системы автоматического стенографирования.

Список использованной литературы

1. Ладошко О.Н. Аннотация и учет речевых сбоев в задаче автоматического распознавания спонтанной украинской речи / О.Н. Ладошко, В.В. Пилипенко // Искусственный интеллект. – Донецк – 2010. – № 3. – С. 238-248.
2. Ladoshko O.N. Annotation of Ukrainian Spontaneous Speech / O.N. Ladoshko, A.N. Prodeus // Proceedings of XXXI International Scientific Conference Electronics and Nanotechnology. 12-14 April, 2011, Kyiv, Ukraine. – K., 2011.
3. Бабкин В.В. Помехоустойчивый выделитель основного тона речи / В.В. Бабкин // 7-я Международная Конференция и Выставка Цифровая Обработка Сигналов и её Применение DSPA-2005. – М., 2005. – С. 175-178.
4. de Cheveigné A. YIN, a fundamental frequency estimator for speech and music / de Cheveigné A., Kawahara H. // J. Acoust. Soc. Am. 111. – 2002. – P. 1917–1930.
5. Gerhard D. Pitch extraction and fundamental frequency: history and current techniques / D. Gerhard // Technical report TR-CS 2003–06. 2003. University of Regina, Saskatchewan, Canada.
6. Bagshaw P. Automatic prosodic analysis for computer aided pronunciation teaching / P. Bagshaw // Univ. of Edinburgh, Edinburgh. – 1993, PhD Thesis. http://www.cstr.ed.ac.uk/projects/fda/Bagshaw_PhD Thesis.pdf
7. A comparative study of several pitch detection algorithms / Rabiner L.R., Cheng M.J., Rosenberg A.E. et al. // IEEE Trans. Acoust. Speech 24. – 1976. – P. 399-423.
8. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк // К.: Наук. Думка, 1987. – 264 с.
9. Rabiner L.R. On the use of autocorrelation analysis for pitch detection / L.R. Rabiner // IEEE Trans. Acoust. Speech Signal Process. 25. – 1977. – P. 24–33.
10. Баронин С.П. Автокорреляционный метод выделения основного тона речи. Пятьдесят лет спустя / С.П. Баронин // Речевые технологии. – 2008. – №2. – С. 3-12.

Надійшла до редколегії 15.03.2012

О.Н. ЛАДОШОКО, А.Н. ПРОДЕУС
Національний технічний університет України «Київський політехнічний інститут», м. Київ, Україна

ОЦІНЮВАННЯ НАДІЙНОСТІ ВИДІЛЮВАЧА ЧАСТОТИ ОСНОВНОГО ТОНУ ДЛЯ АКУСТИЧНОГО АНАЛІЗУ МОВЛЕННЯ

Виконано тестування запропонованої моделі виділювача частоти основного тону (ЧОТ). Розглянуті критерії порівняння якості роботи подібних алгоритмів. Отримані результати можуть бути використані для отримання акустичних характеристик мовлення в задачі автоматичного стенографування.

Ключові слова: надійність, частота основного тону, ларингограф, асинхронне порівняння, частотний контур, мовленнєвий сигнал.

O.N. LADOSHKO, A.N. PRODEUS
National Polytechnic University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine

PERFORMANCE STUDY OF PITCH DETECTION ALGORITHM FOR ACOUSTICAL ANALYSIS OF SPEECH

Performance study of pitch detection algorithm was conducted. It was considered quality's comparison criteria for similar algorithms. The obtained results can be used for purposes of extraction of acoustical spontaneous speech features in a challenge of automatic transcripts acquisition.

Keywords*: *performance, fundamental frequency, laryngograph, asynchronous comparison, frequency contour, speech signal.