

УДК 004.394.2

МОДЕЛИРОВАНИЕ ВЫДЕЛИТЕЛЯ ЧАСТОТЫ ОСНОВНОГО ТОНА ДЛЯ ИССЛЕДОВАНИЙ СПОНТАННОЙ РЕЧИ

О.Н. Ладощко

Национальный технический университет Украины “КПИ”, Киев

Предложена программная реализация метода для оценки частоты основного тона для исследований спонтанной речи. Проведено тестирование помехоустойчивости реализованного метода на тестовых сигналах. Полученные результаты и методы могут быть использованы для исследования акустических характеристик спонтанной речи в задаче автоматического стенографирования.

Введение

Известно, что главной проблемой распознавания спонтанной речи (СР) является её существенное отклонение от грамматических и лексических правил языка [1]. Такое различие СР и подготовленной речи обусловлено наличием в ней спонтанных особенностей, классификация которых представлена в [2]. Эти особенности СР представлены вокализованными паузами (ВП) (вставки типа «э-э», «а-е»), невокализованными паузами (звуки дыхания, кашель), суржиком украинско-русской речи, редуцированием (искажением) звуков в словах слов, растягиванием, повторами и обрывами слов. Они являются одной из причин ухудшения надёжности системы автоматического распознавания слитной речи (САРСР) в режиме распознавания спонтанной речи. В работе [1] были проведены эксперименты по автоматическому распознаванию (АР) спонтанной украинской речи в задаче автоматизированного стенографирования, в которой показано, что ручная очистка спонтанной речи от её особенностей позволяет улучшить показатели надёжности распознавания речи в среднем от 1,25% до 6,45% для разных исследуемых выборок.

До настоящего времени способы улучшения надёжности 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, возникает интерес к моделированию и изучению независимых от лексического состава речи (слов и фонем), характеристик речевых сигналов в реальных условиях записи речи.

В данной работе производится моделирование выделителя частоты основного тона (ЧОТ) для исследования речевых сигналов и их акустических характеристик в реальных условиях записи. Предполагается, что в дальнейшем наличие знаний о траекториях ЧОТ станет дополнительным источником информации для обнаружения некоторых особенностей спонтанной речи.

2. Методы и средства проведения исследований

К настоящему времени разработано множество методов оценки ЧОТ [3, 4, 5], однако, каждый из них порождает наименьшее число ошибок лишь для области его дальнейшего использования. Известно, что при наличии сильной шумовой составляющей в сигнале, точность определения ЧОТ в нём снижается, поэтому для оценки ЧОТ был использован метод, предложенный в работе [3]. Общая структурная схема моделируемого выделителя ЧОТ [3] приведена на рис. 1:

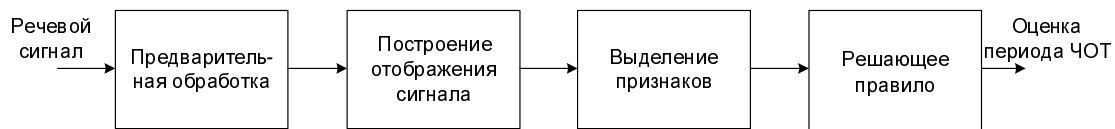


Рисунок 1 – Общая структурная схема выделителя ЧОТ

Разрабатываемый алгоритм основан на обработке речевого сигнала (РС) во временной области в окне, достаточно длинном для надёжного обнаружения периодичности (минимум два периода ожидаемой ЧОТ – 45 мс.), и достаточно малом для исключения преднамеренного сглаживания возможных изменений ЧОТ в окне. Таким образом, РС заменяется последовательностью квазистационарных участков РС со сдвигом (11,25 мс) на каждом шаге. В реализуемом выделителе ЧОТ предварительная обработка заключается в ручном вырезании всех пауз и пропускании РС через полосовой фильтр (ПФ). ПФ с полосой пропускания 70-1200 Гц реализован на основе эллиптического фильтра. Известно, что одним из оптимальных методов оценки ЧОТ является метод, основанный на определении максимума автокорреляционной функции (АКФ) РС в заданных границах его поиска (50-500 Гц). Построение *отображения сигнала* строится для индексов границ поиска максимумов p от

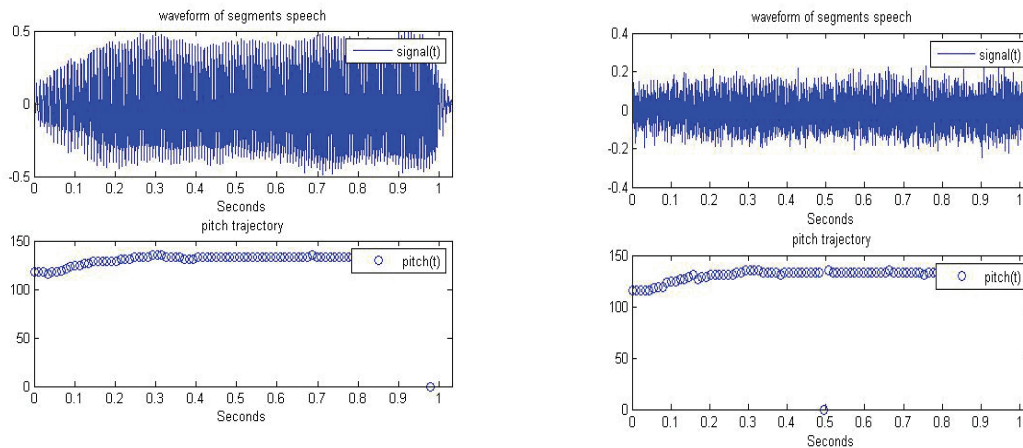
$p_{\min} = 16$ до $p_{\max} = 160$ для тестових сигналів [3] ($F_d = 8\text{кГц}$) на основі
расчета функции нормированной автокорреляции (ФНАК) $R(p)$:

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

Для реальных речевых сигналов задача поиска глобального максимума ФНАК сводиться к оценке *положительных локальных максимумов* ФНАК отдельного кадра k , образующих набор возможных кандидатов $\{p_{m_k}\}$ с дальнейшим применением правил нахождения локальных максимумов ФНАК кандидатов в ЧОТ [3]. Для исключения ошибок выбора кандидата в кадре при увеличении амплитуды пиков ФНАК на задержках кратных периоду ЧОТ значения ФНАК умножаются на монотонно убывающую функцию $\exp(-a \cdot \tau)$. Кроме того, для зашумлённых сигналов достоверное определение ЧОТ по одному кадру анализа оказывается недостаточным [3, 6]. В связи с этим учитывается информация о возможной траектории ЧОТ для каждого кадра анализа на основе оценок ЧОТ смежных кадров. Таким образом, проводится совокупный анализ группы смежных кадров – $K = \{k, k+1, \dots, k+K+1\}$, учитывающий вероятности возможных кандидатов на оценку ЧОТ, пропорциональные величине локального максимума ФНАК $d_N(m_k) = R_k(p(m_k))$ для каждого из кадров путём поиска оптимального пути, максимизирующего общую вероятность появления $D_L(m_k, m_{k+1}, \dots, m_{k+K-1}) = \sum_{i=0}^{K-2} (d_N(m_{k+i}) + d_T(m_{k+i}, m_{k+i+1})) + d_N(m_{k+K-1})$ оценок ЧОТ для группы кадров с условием незначительного изменения траектории ЧОТ для гласных звуков [3].

3. Результаты экспериментов

Для оценки помехоустойчивости разработанного на основе [3] выделителя ЧОТ использовались тестовые сигналы v и $v0$ [7], состоящие из изолированного произнесения 91 гласного звука, исключая паузы. Выбор параметров для тестирования алгоритма осуществлялся из [3, 8]. Недостающие параметры (пороги α , a и M_k – количество кандидатов) определялись экспериментально.



а – сигнал v - без шуму

б – сигнал $v0$ - белый шум, $SNR = 0$ дБ

Рис. 2. Осциллограмма одного звука из тестовых сигналов и его траектории ЧОТ.

Количественная оценка точности выдаваемых оценок ЧОТ (см. табл. 1) осуществлялась на v -файле (без шума), путём сравнения оценок, получаемых ЧОТ на выходе реализуемого метода и алгоритма программного продукта Praat [9] в виде процента грубых ошибок [5]

$$GPE = \frac{1}{M} \sum_{m=1}^M \left| \frac{p(k) - F0(k)}{F0(k)} \right|, \text{ где } M - \text{количество измерений, } p(k) - \text{оценка,}$$

$F0(k)$ – эталонное значение. Усреднение GPE проводилось для всех значений измеряемых оценок ЧОТ на гласных звуках присутствующих в тестовых звуковых файлах v и $v0$.

Таблица 1. Оценки выделителей ЧОТ (величина GPE , %)

Точность в сравнении с Praat, файл- v	Помехоустойчивость разработанного выделителя файл- $v0$, $SNR = 0$ дБ	Помехоустойчивость Бабкин В.В. файл $v0$, $SNR = 0$ дБ
2,96	1,36	1,4

Для определения помехоустойчивости реализованного метода () эталонные значения ЧОТ были взяты в виде оценок ЧОТ для файла- v (без шума) на выходе реализованного выделителя ЧОТ.

Таким образом, для оценки помехоустойчивости использовались два звуковых тестовых сигнала [7] – v без шума и $v0$ с $SNR = 0$ дБ. Предварительно исключались явные ошибки оценки ЧОТ, связанные с разрывом траектории. В нашем случае разрывы траектории ЧОТ связаны с равенством ЧОТ на этих участках нулю при произнесении гласных звуков.

Заклучение

Представлены результаты моделирования оценки ЧОТ. Проведенное тестирование помехоустойчивости реализованного метода на тестовых сигналах показало достаточно высокую помехоустойчивость (процент грубых ошибок $GPE = 1,36\%$) и точность реализованного выделителя ЧОТ (2,96%). Тем не менее, проведенные исследования по тестированию программной реализации метода показали необходимость дальнейших исследований для устранения некоторых типов ошибок, выдаваемых выделителем ЧОТ, связанных с возможными нерегулярностями голоса в исследуемом материале, а так же особенностями разработанного алгоритма.

Список литературы

1. Ладоско О.Н., Пилипенко В.В. Аннотация и учет речевых сбоев в задаче автоматического распознавания спонтанной украинской речи // Искусственный интеллект. – Донецк – № 3. – 2010. – С. 238-248.
2. Ladoshko O. N., Prodeus A. N. Annotation of Ukrainian Spontaneous Speech. Proceedings of XXXI International Scientific Conference Electronics and Nanotechnology. 12-14 April, 2011, Kyiv, Ukraine.
3. Бабкин В. В. Помехоустойчивый выделитель основного тона речи. // 7-я Международная Конференция и Выставка Цифровая Обработка Сигналов и её Применение DSPA-2005. – Москва 16-18 марта. – 2005г.– С.175-178.
4. Gerhard D. Pitch extraction and fundamental frequency: history and current techniques // Technical report TR-CS 2003–06. 2003. University of Regina, Saskatchewan, Canada.
5. Rabiner L.R., Cheng M.J., Rosenberg A.E., Mc-Gonegal C.A. A comparative study of several pitch detection algorithms // IEEE Trans. Acoust. Speech 24, pp. 399-423. – 1976.
6. Баронин С. П. Автокорреляционный метод выделения основного тона речи. Пятьдесят лет спустя // Речевые технологии. – М. – №2. – 2008. – С. 3-12.
7. Бабкин В.В. Тестовые файлы для оценки помехоустойчивости выделителей ОТ. Звуковые файлы v, v0. СПб: Центр ЦОС СПб ГУТ, 2005.
8. Adam P. Vogel, Paul Maruff, Peter J. Snyder, James C, Mundt. Standartization of pitch-range settings in voice acoustic analysis. // Behavior Research Methods. – 41(2)/ – 2009. – pp. 318-324.
9. P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer", Institute of Phonetic Sciences of the University of Amsterdam", pp. 132-182, 1999.

Получено 12.09.2011