

ОПРЕДЕЛЕНИЕ ЧАСТОТЫ ОСНОВНОГО ТОНА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ИМИТАЦИИ ОТЖИГА

Карижская А.Д., Жукова А.Б., Масленников А.Л.

Московский государственный технический университет имени Н.Э. Баумана, г.Москва

Ключевые слова: частота основного тона, распознавание речи, алгоритм имитации отжига, амплитудный спектр, выделение пиков.

Аннотация. В задачах распознавания речи зачастую требуется определение частоты основного тона, для чего существует несколько различных методов. В данной работе рассматривается модифицированный метод, основанный на поиске пиков амплитудного спектра речевого сигнала и последующего определения частоты основного тона. В предлагаемой модификации метода для решения оптимизационной задачи предлагается использовать алгоритм имитации отжига. В статье показывается принципиальная работа метода на различных по качеству речевых сигналах, а также, анализируется зависимость быстродействия решения задачи в зависимости от настраиваемых параметров алгоритма имитации отжига.

VOICE PITCH FREQUENCY DETECTION UTILIZING SIMULATED ANNEALING TECHNIQUE

Karizhskaya A.D., Zhukova A.B., Maslennikov A.L.

Bauman Moscow State Technical University, Moscow

Keywords: voice pitch frequency, speech recognition, simulated annealing, amplitude spectrum, peaks search.

Abstract. There is often necessary to determine the voice pitch frequency is speech recognition problems. There are different methods to solve this problem. The computation of the voice pitch frequency using spectrum peaks search method is considered in this work. As a method modification the solution to embedded optimization problem is done by simulated annealing (SA) technique. The results for different quality of speech signals are shown in this paper, as well as, the computational speed and its dependency on the SA technique parameters.

Введение

В задачах распознавания речи для определения конкретного звука или слова, произнесенного человеком, используются различные признаки, характеристики или параметры речевого сигнала или его части. Одной из таких характеристик является частота основного тона, то есть частота, на которую приходится максимум мощности речевого сигнала на частотах спектра кратных частоте основного тона [1, 2].

На данный момент существует несколько способов определения частоты основного тона: с использованием метода линейного предсказания и с непосредственным определением пиков амплитудного спектра и вычислением частоты основного тона согласно ее классическому определению [3, 4]. Отметим, что спектр речевого сигнала подвержен влиянию шумов различной природы, поэтому искомый пик, а следовательно, и частота основного тона, может определяться неверно [5]. В данной работе рассмотрен метод определения частоты основного тона, в котором для поиска соответствующего пика амплитудного спектра используется алгоритм имитации отжига.

Алгоритм имитации отжига

Алгоритм имитации отжига – один из методов глобальной оптимизации, основанный на стохастическом поиске. Его цель состоит в нахождении экстремума функции (рассматриваем случай максимума) $g(x_{\max})$ и определении соответствующего значения $x_{\max}$ аргумента [6, 7]. Алгоритм представляет собой итерационный процесс, на каждом шаге которого генерируется в заданном диапазоне случайным образом новая точка x_k , вычисляется значение $g(x_k)$, затем определяется вероятность принятия новой точки x_k в качестве решения на текущем k -ом шаге, согласно следующему уравнению (для случая поиска максимума)

$$P(x_k | x_{k-1}) = \begin{cases} 1 & , \quad g(x_k) - g(x_{k-1}) > 0 \\ e^{\frac{g(x_k) - g(x_{k-1})}{T_k}} & , \quad g(x_k) - g(x_{k-1}) \leq 0 \end{cases}; \quad (1)$$

где изменение температуры T , определяется как

$$T_k = h_T T_{k-1};$$

где $h_T = \text{const} < 1$ – коэффициент, характеризующий скорость уменьшения температуры.

Если рассчитанная вероятность в уравнении (1) больше заданного порога ε_p , то решение принимается, если нет, то процесс продолжается. В случае принятия решения происходит также понижение величины T , так называемой, температуры. Алгоритм останавливается, когда $T < \varepsilon_T$, где постоянная $\varepsilon_T > 0$, а последняя рассчитанная точка x_k принимается как $x_{\max}$.

Определение частоты основного тона

Алгоритм имитации отжига может быть применен для нахождения частоты основного тона речевого сигнала. Для этого в качестве параметра x необходимо выбрать частоту f , а за минимизируемую функцию принять функцию следующего вида:

$$g(f) = \sum_{i=1}^3 S(if); \quad (2)$$

где $S(if)$ – значения амплитуд спектра на соответствующих частотах кратных частотам f . Отметим, что согласно классическому определению частоты основного тона в уравнении (2) должно использовать произведение, однако на практике, это может привести к вычислительным ошибкам, а использование суммы, является более надежным [5].

Полагая, что диапазон, в котором находится частота основного тона, приблизительно соответствует интервалу от 50 до 300 Гц, а также учитывая, что на частотах выше 900 Гц могут присутствовать сильные помехи и шумы, количество элементов суммы в уравнении (2) принято равное трем. Температура T понижается с коэффициентом $h_T = \text{const} = 0.83$, который был подобран эмпирически, а критерий остановки $\varepsilon_T = 10^{-5}$. Начальное значение температуры принято равным 100, а частоты – 50 Гц. Критерий принятия решения $\varepsilon_p = 0.8$.

Пример нахождения пика амплитудного спектра, и соответствующей ему частоте основного тона с использованием алгоритма имитации отжига представлен на рис. 1 а) для речевого сигнала с минимальным уровнем помех и на рис. 1 б) и в) для речевого сигнала с явно выраженными помехами.

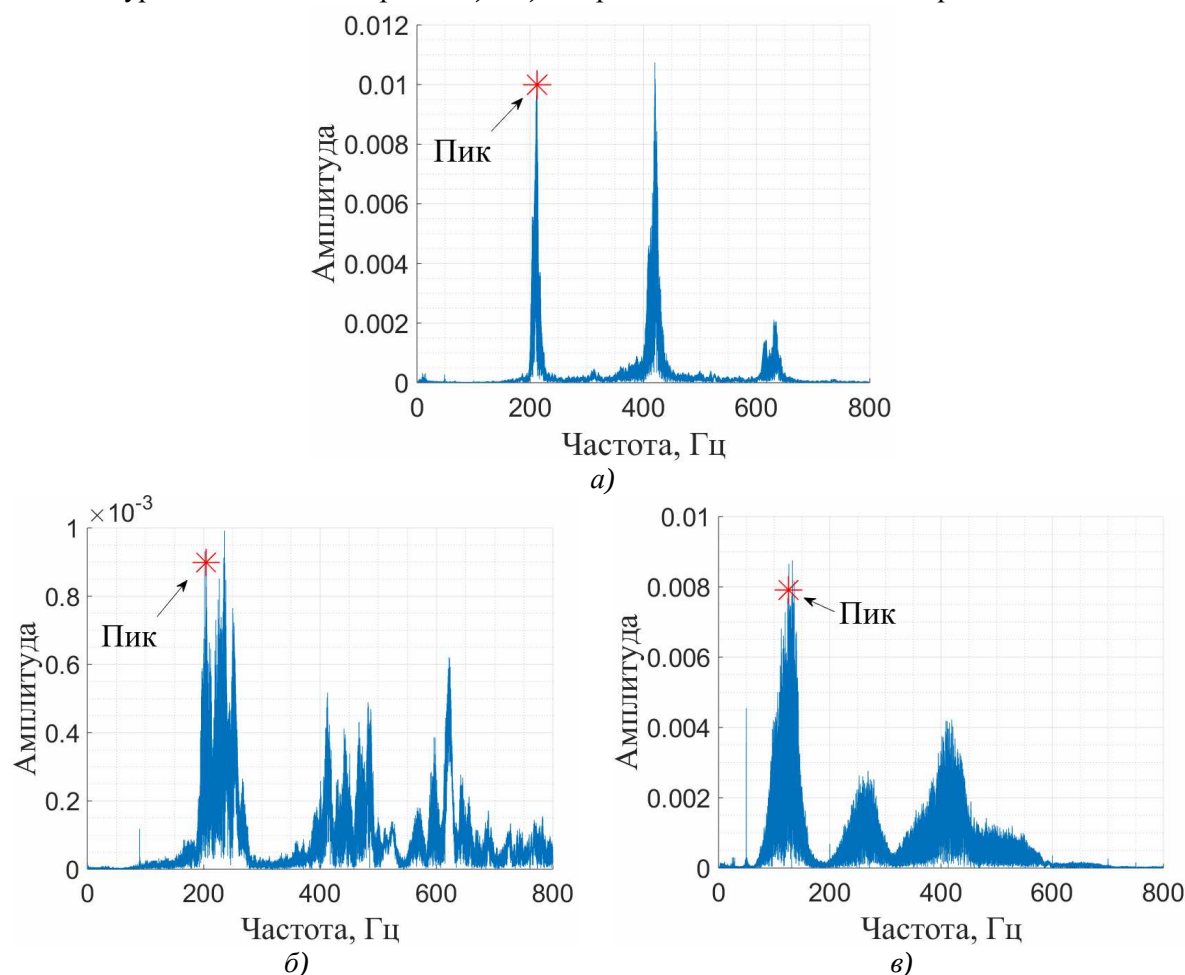


Рис. 1. Пример определения частоты основного тона при неидеальных условиях проведения эксперимента

Анализ быстродействия

Алгоритм имитации отжига – итерационный, поэтому актуален вопрос о его временной вычислительной сложности, на которую влияют три параметра алгоритма: задаваемая точность ε_T , порог принятия решения ε_p и величина h_T . Зависимость времени решения задачи от этих параметров показана на рис. 1.

Как видно, увеличение точности ε_T линейно увеличивает время решения задачи, а зависимость времени решения задачи от величины h_T является экспоненциальной. Отметим, что время решения

задачи в зависимости от ε_P носит линейный характер – с увеличением ε_P увеличивается время решения, однако это увеличение весьма незначительно, и можно считать, фактически инвариантным зависимость быстродействия от этого параметра.

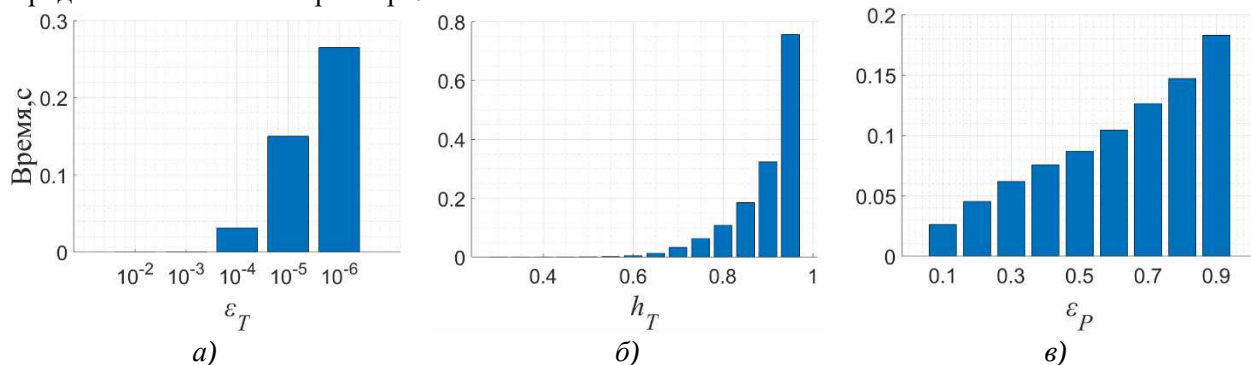


Рис. 2. Зависимость быстродействия параметров алгоритма

Заключение

В статье была рассмотрена задача определения частоты основного тона, в которой при решении вложенной оптимизационной задачи поиска соответствующего пика амплитудного спектра речевого сигнала, использовался алгоритм имитации отжига. В результате было показано, что при использовании данного алгоритма задача принципиально решается для различных по качеству речевых сигналов. Также было показано, что скорость решения задачи линейно зависит от задаваемой точности и экспоненциально от параметра понижения температуры, а зависимость от величины порога принятия решения можно считать инвариантной.

Список литературы

1. Плотников В.Н., Суханов В.А., Жигулевцев Ю.Н. Речевой диалог в системах управления. М.: Машиностроение, 1988. 224 с.
2. Бондаренко М.Ф., Работягов А.В., Щепковский С.В. Распознавание речи: этапы развития, современные технологии и перспективы их применения // Бионика интеллекта. 2010. Вып. 73, №2. С. 164-168.
3. Маркел Д.Д., Грэй А.Х. Линейное предсказание речи / Пер. с англ. под ред. Ю.Н. Прохорова, В.С. Звездина. М.: Связь, 1980. 308 с.
4. Rabiner L.R, Schafer R.W. Introduction to Digital Speech Processing. – Foundations and Trends in Signal Processing, 2007. DOI: 10.1561/20000000001.
5. Жукова А.Б., Масленников А.Л. Определение частоты основного тона методом поиска пиков в амплитудном спектре с добавлением весовых частотных функций // Политехнический молодежный журнал. 2019. №12. DOI: 10.18698/2541-8009-2019-12-556.
6. Джонс М.Т. Программирование искусственного интеллекта в приложениях / Пер. с англ. А.И. Осипова. М.: ДМК Пресс, 2011. 312 с.
7. Spall J.C. Introduction to stochastic search and optimization: estimation, simulation, and control. New Jersey: Wiley, 2003. 618 p.

Сведения об авторах:

Карижская Анна Дмитриевна – студентка, МГТУ им. Н.Э. Баумана, г.Москва;

Жукова Анастасия Борисовна – студентка, МГТУ им. Н.Э. Баумана, г.Москва;

Масленников Андрей Леонидович – старший преподаватель кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, г.Москва.