Математические структуры и моделирование
2016. № 1(37). С. 59–65

УДК 004.934

ГИБРИДНЫЙ АЛГОРИТМ ВЫДЕЛЕНИЯ ЧАСТОТЫ ОСНОВНОГО ТОНА

О.А. Вишнякова
преподаватель, e-mail: olga@infotekorg.ru
Д.Н. Лавров
доцент, к.т.н., e-mail: dmitry.lavrov72@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В статье приведено описание алгоритма нахождения трека основного тона на базе смешанного алгоритма поиска в спектральной и временной областях для исходного сигнала и его нелинейного преобразования. Набор кандидатов формируется на выходе спектральной гармонической корреляционной функции и нормализованной взаимнокорреляционной функции. После итогового отсева кандидатов формируется конечный трек.

Ключевые слова: оценка основного тона, кросс-корреляционная функция.

Введение

В большинстве задач классификации речевых сигналов при параметрическом представлении речи значимым параметром является мгновенная частота основного тона F_0, определяемая как мгновенная частота колебаний голосовых связок диктора. Основными показателями качества оценки являются временное и частотное разрешения, то есть скорость реакции на изменение F_0 и величину отклонения, которое фиксирует алгоритм.

К настоящему времени предложен ряд алгоритмов оценки основного тона, в том числе использующих методы оценки как во временной, так и в частотной областях [1–5]. Наиболее популярными алгоритмами оценки являются RAPT [6], YIN [7] и SWIPE [8] и их модификации. Несмотря на низкий процент ошибок даже при наличии шумов (как фонового, так и обусловленного смешанным возбуждением речевого тракта), точность снижается при модуляции F_0.

Рассматриваемый алгоритм базируется на RAPT и использует в поиске кандидатов нормализованную кросс-корреляционную функцию (НККФ). Поставив перед собой задачу минимизировать чувствительность к модуляциям основного тона и степени зашумлённости сигнала, предложенный метод представляет собой комбинацию корреляционного метода и частотной селекции для оценки F_0. При этом, как было показано в [9, 10], добиваясь устойчивости к внешним помехам, оценка в спектральной области проводится как для исходного сигнала, так и для его нелинейного преобразования.

1. Описание алгоритма

Можно выделить основные шаги алгоритма, включающие предобработку, поиск кандидатов и итоговую постобработку. На рис. 1 приведена общая схема алгоритма.

Рис. 1. Схема алгоритма поиска трека основного тона

1.1. Предобработка

Фундаментальная частота F_0 проявляется при квадрировании сигнала даже при условии малой амплитуды либо отсутствия в исходных данных, как показано в [9], что характерно для телефонной речи. Таким образом, предобработка включает в себя создание копии исходного сигнала и его нелинейное преобразование (квадрирование), нормализацию, а также последующую фильтрацию полосовым фильтром с полосой пропускания (50–1500 Гц) исходного и

квадрируемого сигналов. Допустимый интервал на F_0 определяем 60–400 Гц. На рис. 2 приведён результат постобработки.

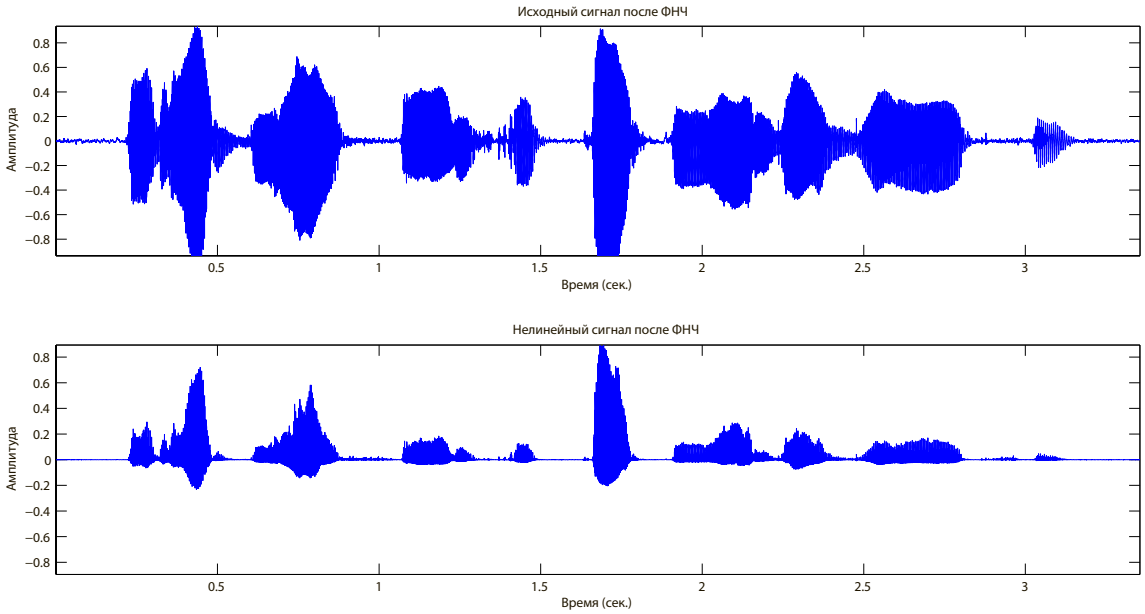

Рис. 2. Исходный сигнал после применения фильтра (сверху), нелинейно обработанный сигнал после применения фильтра (снизу)

1.2. Поиск кандидатов F_0 по максимумам SHC

Основа метода частотной селекции базируется на предположении, что при вокализованном возбуждении речевого тракта в спектре сигнала присутствуют пики на частотах, кратных частоте основного тона. Поиск выполняется на интервалах в 32 мс с перекрытием в 10 мс при частоте дискретизации в 16 кГц. Для лучшего частотного разрешения применяется интерполяция оконным sinc-фильтром, получая в итоге шаг по частоте в 7.8 Гц и ширине окна в 2048 отсчётов. Далее строится спектральная гармоническая корреляционная функция SHC, определяемая следующим соотношением:

$$SHC(n,f) = \sum_{f'=-WL/2}^{WL/2} \prod_{r=1}^{R} S(n, rf + f'),$$

где $S(t,n)$ — спектр сигнала для фрейма n, WL ширина спектрального окна, R число гармоник. Так как сигнал нормализован, максимальное значение функции 1.0. Выполняется поиск локальных максимумов только для спектра квадрируемого сигнала, при этом пороговое значение для отсеивания ложных экстремумов установлено в 0.6. На рис. 3 спектр и спектральная кросс-корреляционная функция.

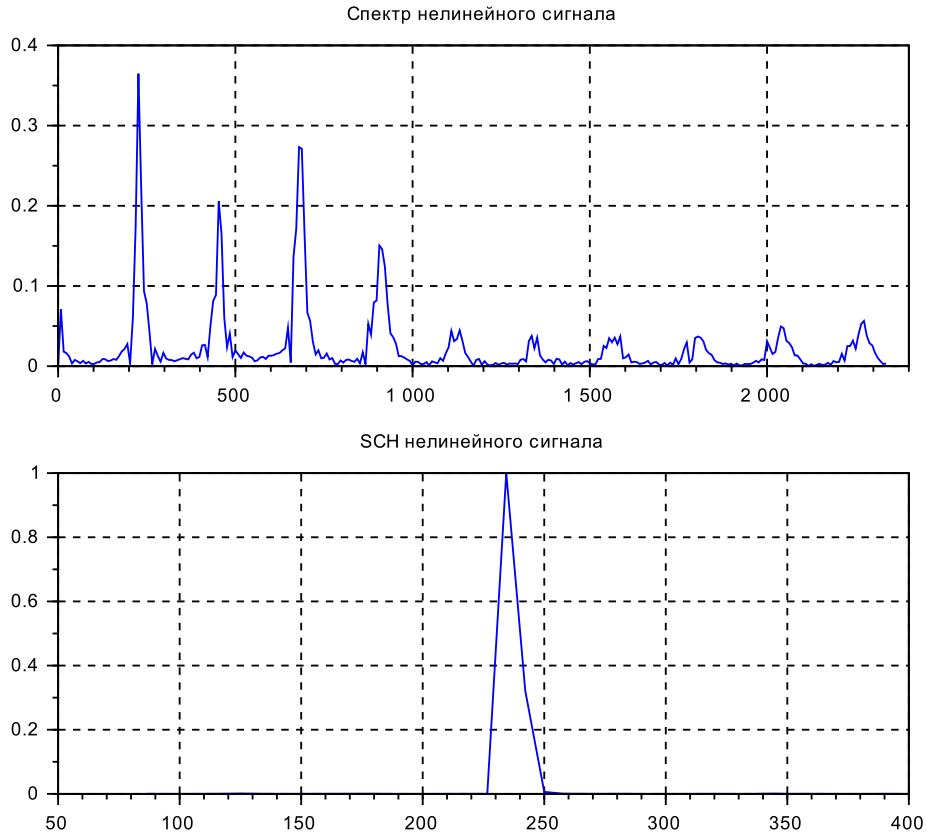

Рис. 3. Спектр фрейма нелинейного преобразования сигнала и его SCH

Для минимизации ошибок F_0 вычисляется на вокализованных участках. Для принятия решение о типе интервала используется нормализованное низкочастотное энергетическое соотношение NLFER, которое определяется отношением суммы спектральных компонент фрейма в диапазоне частот $F_{0max} - F_{0min}$ к среднему значению по всему сигналу.

$$NLFER(n) = \frac{\sum_{f=F_{0min}}^{F_{0max}} S(n,f)}{\frac{1}{N}\sum_{n=1}^{N}\sum_{f=F_{0min}}^{F_{0max}} S(n,f)}.$$

1.3. Поиск кандидатов F_0 по максимумам NCCF

Кандидаты вычисляются как для исходного, так и для нелинейно модифицированного сигнала, используя нормализованную кросс-корреляционную функцию NCCF (НККФ), определяемую следующим соотношением:

$$NCCF(k) = \frac{1}{\sqrt{e_0 e_k}} \sum_{n=1}^{N-K_max} s(n)s(n+k),$$

где

$$e_0 = \sum_{n=1}^{N-K_{max}} s(n)^2, e_k = \sum_{n=k}^{k+N-K_{max}} s(n)^2, K_{min} \leqslant k \leqslant K_{max}.$$

Локальные максимумы НККФ соответствуют задержке сигнала, равному периоду основного тона. В случае, когда имеется несколько локальных максимумов НККФ близких к единице, выбирается соответствующий наименьшему периоду. Так как значения на невокализованных участках значительно меньше 1, НККФ вычисляется только на вокализованных участках, определяемых NLFER.

1.4. Постобработка

На стадии постобработки выполняется поиск контура основного тона при помощи динамического программирования, соединяющий найденных кандидатов периода в спектральной и динамической областях, при этом накладывается ограничение, что частота основного тона изменяется медленно и, таким образом, значения частот смежных фреймов не должны сильно отличаться [11].

2. Результаты экспериментов

2.1. Речевая база данных

Тестирование алгоритмов поиска F_0 важно проводить на одних и тех же речевых базах данных. Существует несколько свободных баз, собранных различными исследовательскими лабораториями. В состав данных включают записи с ларингофона и значения эталонных частот основного тона, вычисленных по траекториям с ларингофона.

В качестве примера можно привести:

1. «The Pitch-Tracking Database». Включает 2342 предложений, произнесённых 10 мужскими и 10 женскими голосами [12].
2. «The fundamental frequency determination algorithm evaluation database». Включает по 50 предложений, произнесённых одним мужским и одним женским голосом [13].

В работе использовалась «The Pitch-Tracking Database». Эталонные частоты посчитаны при ширине окна в 32 мс и перекрытием в 10 мс.

2.2. Трек частоты основного тона

На рис. 4 приведён результат работы алгоритма — итоговый трек частоты основного тона.

Мерой ошибок считаем процент грубых ошибок (Gross Error — GE), вычисляемый как

$$GE = \frac{1}{N_{VF}} \sum_{k=1}^{N_V F} \delta(F_0^{ref}(t), F_0^{est}(t)),$$

$$\delta(F_0^{ref}(t), F_0^{est}(t)) = \begin{cases} 1, & \left| \frac{F_0^{ref}(t) - F_0^{est}(t)}{F_0^{ref}(t)} \right| > 0.2 \\ 0 \end{cases}$$

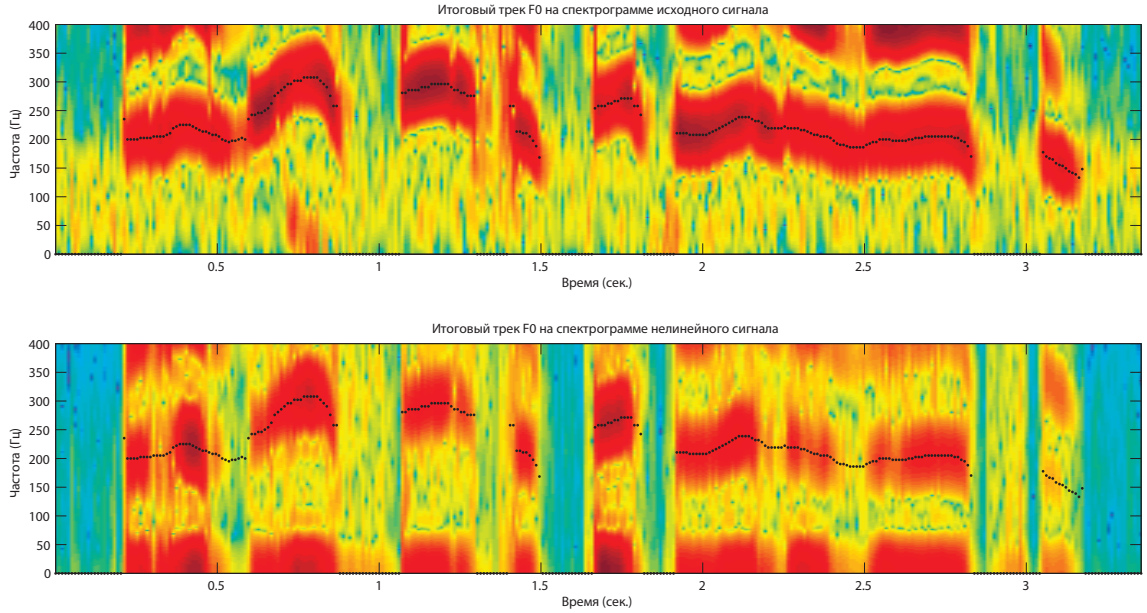

Рис. 4. Спектрограмма исходного сигнала и итоговый трек F_0 (сверху), спектрограмма нелинейного преобразования сигнала и итоговый трек F_0 (снизу)

где N_{VF} число вокализованных фреймов, F_0^{ref} эталонное значение F_0, F_0^{est} вычисленное значение. Таким образом, определяется число фреймов с отклонением полученной оценки более чем на 20%.

По результатам экспериментов для женских голосов GE = 4.1%, для мужских 3.7%.

3. Заключение

Предложенный метод нахождения трека основного тона реализован на базе смешанного алгоритма поиска в спектральной и временной областях для исходного сигнала и его нелинейного преобразования. Эффективность метода обусловлена использованием нелинейной версии сигнала для поиска кандидатов и объединением результатов поиска. Приведены результаты работы алгоритма.

Литература

1. Hess W.J. Pitch and voicing determination / Advances in Speech Signal Processing / edited by S. Furui, M.M. Sohndi. 1992. P. 3–48.
2. Hermes D.J. Pitch analysis / Visual Representations of Speech Signals / edited by M. Cooke, S. Beet, M.C. Wiley. 1993. P. 3–25.
3. Gerhard D. Pitch Extraction and Fundamental Frequency: History and Current Techniques. Technical report, Dept. of Computer Science, University of Regina, 2003.
4. Pavlovets A., Petrovsky A. Robust HNR-based closedloop pitch and harmonic parameters estimation // Proc. the 12th Annual Conference of the International Speech Communication Association (Interspeech-2011), Italy, Florence, 27-31 August 2011.

5. Zubrycki P. Petrovsky A. Quasi-periodic signal analysis using harmonic transform with application to voiced speech processing // ISCAS 2010: 2374-2377.
6. Talkin D. A Robust Algorithm for Pitch Tracking (RAPT) / Speech Coding and Synthesis / W.B. Kleijn, K.K. Paliwal eds. Elsevier, ISBN 0444821694. 1995.
7. Cheveigne A., Kawahara H. YIN, a fundamental frequency estimator for speech and music // Journal Acoust. Soc. Am. 2002. Vol. 111, № 4. P. 1917–1930.
8. Camacho A., Harris J.G. A sawtooth waveform inspired pitch estimator for speech and music // Journal Acoust. Soc. Am. 2008. Vol. 123, № 4. P. 1638–1652.
9. Zahorian S.A., Hu H. A spectral/temporal method for robust fundamental frequency tracking // The Journal of the Acoustical Society of America. 2008. № 123. P. 4559–4571.
10. Kavita K., Zahorian S. Yet another algorithm for pitch tracking // Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on. IEEE 2002. Vol. 1. P. 1—361.
11. Азаров И.С., Вашкевич М.И., Петровский А.А. Алгоритм оценки мгновенной частоты основного тона речевого сигнала // Цифровая обработка сигналов. 2012. № 4. С. 49–57.
12. Pirker G., Wohlmayr M., Petrik S., Pernkopf F. Database for multi-pitch tracking // Graz University of Technology, Signal Processing and Speech Communication Laborator. 2012. URL: `http://www2.spsc.tugraz.at/databases/PTDB-TUG/` (дата обращения: 06.02.2016).
13. Bagshaw P.C., Miller S.M., Jack M.A. Enhanced pitch tracking and the processing of the F0 contours for computer aided intonation teaching // Proceedings of EUROSPEECH, Berlin, Germany. 1993. 1003–1006. URL: `http://www.cstr.ed.ac.uk/research/projects/fda` (дата обращения: 06.02.2016).

THE HYBRID ALGORITHM OF EXTRACTION OF FUNDAMENTAL FREQUENCY

O.A. Vishnyakjova
Teacher, e-mail: olga@infotekorg.ru
D.N. Lavrov
Ph.D. (Eng.), Associate Professor, e-mail: dmitry.lavrov72@gmail.com

Omsk State University n.a. F.M. Dostoevskiy

Abstract. This paper presents a new algorithm for the estimation of the fundamental frequency of speech. It is based on the combination of time domain and frequency domain processing applied for the original and nonlinearly processed version of the signal. The set of candidates is formed by using a spectral harmonics correlation and the normalized cross-correlation function. Final F_0 track is calculated after candidates selection.

Keywords: estimation of the fundamental frequency of speech, cross-correlation function.