Математическое и компьютерное моделирование:
материалы III Международной научной конференции
(Омск, 12 ноября 2015 г.). Омск, 2015. С. 168–170. УДК 004.934

О.А. Вишнякова, Д.Н. Лавров

Омский государственный университет им. Ф.М. Достоевского, г. Омск

ГИБРИДНЫЙ АЛГОРИТМ ОЦЕНКИ ОСНОВНОГО ТОНА

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. Рассматриваемый алгоритм основывается на RAPT [1] и представляет собой комбинацию корреляционного метода и частотной селекции для оценки F_0, устойчивой к внешним помехам.

Можно выделить основные шаги алгоритма:

1. Предобработка. Фундаментальная частота F_0 проявляется при квадрировании сигнала даже при условии малой амплитуды либо отсутствия в исходных данных, как показано в [4], что характерно для телефонной речи. Таким образом, предобработка включает в себя создание копии исходного сигнала и его нелинейное преобразование (квадрирование), нормализацию, а так же последующую фильтрацию полосовым фильтром с полосой пропускания (50–1500 Гц) исходного и квадрируемого сигналов. Допустимый интервал на F_0 определяем 60–400 Гц.

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. 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, мак-

симальное значение функции 1.0. Выполняется поиск локальных максимумов только для спектра квадрируемого сигнала, при этом пороговое значение для отсеивания ложныжных экстремумов установлено в 0.6.

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

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

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, определяемую следующим соотношением:

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

где

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

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 ЧОТ смежных фреймов не должны сильно отличаться [5].

Литература

1. *Talkin D.* A Robust Algorithm for Pitch Tracking (RAPT) // Speech Coding and Synthesis / ed. by W.B. Kleijn, K.K. Paliwal. Elsevier, 1995.
2. *Zahorian S.A., Hu H.* A spectral/temporal method for robust fundamental frequency tracking // The Journal of the Acoustical Society of America. 2008. Vol. 123. P. 4559–4571.

3. *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. I–361.
4. *Zahorian S.A., Dikshit P., Hu H.* A spectral-temporal method for pitch tracking // Proceedings of the Ninth International Conference on Spoken Language Processing, Interspeech 2006. Pittsburgh, 2006. P. 1710–1713.
5. *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.