На правах рукописи

Мельников Александр Алексеевич

БЫСТРЫЕ АЛГОРИТМЫ ОЦЕНКИ ПАРАМЕТРОВ ПОЛИГАРМОНИЧЕСКОЙ МОДЕЛИ ГОЛОСОВОГО СИГНАЛА

Специальность 01.01.09 —
«Дискретная математика и математическая кибернетика»

Автореферат

диссертации на соискание учёной степени
кандидата физико-математических наук

Санкт-Петербург — 2016

Работа выполнена в Санкт-Петербургском государственном университете

Научный руководитель: доктор физико-математических наук, профессор
Барабанов Андрей Евгеньевич

Официальные оппоненты: **Соколов Виктор Федорович,**
доктор физико-математических наук, профессор,
Федеральное государственное бюджетное учреждение науки Коми научный центр Уральского отделения Российской академии наук (Коми НЦ УрО РАН),
отдел математики, ведущий научный сотрудник

Пацко Валерий Семёнович,
кандидат физико-математических наук, старший научный сотрудник,
Федеральное государственное бюджетное учреждение науки Институт математики и механики им. Н.Н.Красовского Уральского отделения Российской академии наук (ИММ УрО РАН),
отдел динамических систем, заведующий сектором

Ведущая организация: Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук

Защита состоится «16» ноября 2016 г. в 16 часов на заседании диссертационного совета Д 212.232.29 на базе Санкт-Петербургского государственного университета по адресу: 199178, Санкт-Петербург, 10 линия В.О., д.33/35, ауд. 74.

С диссертацией можно ознакомиться в библиотеке им. М. Горького Санкт-Петербургского государственного университета по адресу: 199034, Санкт-Петербург, Университетская наб., 7/9 и на сайте http://spbu.ru/science/disser/dissertatsii-dopushchennye-k-zashchite-i-svedeniya-o-zashchite.

Автореферат разослан «___» _____ 2016 г..

Ученый секретарь
диссертационного совета
Д 212.232.29, проф., докт. физ.-мат. наук В.М. Нежинский

Общая характеристика работы

Актуальность темы. Математические модели речевых сигналов применяются для решения разнообразных задач анализа, синтеза, распознавания и кодирования речи [Springer, Handbook of speech processing, 2007]. Основной моделью звонкого речевого сигнала на коротком промежутке времени является периодическая функция, которая полностью определяется своим периодом, а также амплитудами и фазами всех гармоник, входящих в ряд Фурье. Идентификация этих параметров по измеряемым отсчётам с определённой частотой дискретизации и в условиях шумов составляет задачу анализа речевого сигнала. Наибольшую сложность представляет оценка периода модели или обратной ему величины - частоты основного тона (ЧОТ), которая характеризует высоту слышимой речи. Комплексные амплитуды оцениваются по методу наименьших квадратов из условия минимума квадратичной невязки сигнала и модели.

Многие исследователи занимались вопросом определения ЧОТ голосового сигнала. Первые шаги в этом направлении были сделаны ещё в 19 веке. Гельмгольц [1912] с помощью специальных резонаторов выделял гармоники, содержащиеся в речевом сигнале.

Первыми математическими подходами можно считать семейство методов, основанных на выборе маркеров ЧОТ с последующей их обработкой. Простейший из них — zero-crossing rate — выделяет последнее пересечение нуля перед каждым максимумом на периоде. Целое семейство аналогичных методов освещено в литературе [Dologlou I., Carayannis G., 1989; Hess W. J, 1976; Ananthapadmanabha T., Yegnanarayana B., 1975, Dolansky L. O., 1955; Howard I. S., Walliker J., 1989; Hess W. J., 1976].

На следующем этапе развития алгоритмов можно выделить автокорреляционные методы и методы, основанные на вычислении функции расстояния. Один из распространённых типов алгоритма, «average magnitude difference function», AMDF [M. Ross, 1974; Sobolev V., Baronin S., 1968; Moorer J. A., 1974; Cheveigne A. de, Kawahara H., 2002], основан на предположении о том, что средняя величина сигнала будет оставаться практически одинаковой от одного периода к другому.

Shimamura и Kobayashi [2001] в своей работе совместили корреляционные методы (ACF) и AMDF методы путём взвешивания ACF обратной к AMDF величиной.

Hirose [1992] и Talkin [1995] показали, что автокорреляционную функцию можно вычислять, используя дополнительную нормализацию.

Terez [2002] описал иной подход. Основная идея приводит к функции расстояния в многомерном пространстве состояний. Идея использования многомерной репрезентации сигнала для определения ЧОТ начинается с 1950-ых годов [Hess W. J., 1982].В 1964 году Рейдер опубликовал векторный алгоритм поиска ЧОТ, где он использовал выходные сигналы от фильтров (см. Yaggi, 1962) и их преобразование Гильберта для формирования многомерного вектора, по которому производилась оптимизация.

Автокорреляционные методы описаны в работах [Rabiner L., 1977; Sondhi M. M., 1968; Markel J. D., 1972].

Часть алгоритмов нацелена на определение параметров полигармонической модели речевого сигнала. Общая идея здесь такова: оценить параметры модели речевого сигнала так, чтобы модель наилучшим образом аппроксимировала анализируемый сигнал (см. [Griffin D. W., 1988; Noll A. M., 1969; McAulay R. J., Quatieri T. F., 1990].

Ещё одно семейство методов переносит обработку речевого сигнала в частотную область. Тут можно отметить такие методы, как детектирование пиков спектра, методы спектральной корреляции, harmonic product spectrum, методы на основе кепстра (см. [McLeod P., Wyvill G., 2003; Dziubinski M., Kostek B., 2004; Kondoz A. M., 2005; Lahat M., Niederjohn R. J., Krubsack D. A., 1987; Schroeder M. R., 1968; Martin P. A., 1987; Brown J. C., 1992; Hermes D. J., 1988; Duifhuis H., Willems L. F., Sluyter R., 1982; Noll A. M., 1967; Indefrey H., Hess W. J., Seeser G., 1985; Martin P., 1982]).

Отдельно можно выделить методы, являющиеся наиболее успешными и популярными в наше время: YAAPT [Zahorian S. A., Hu H., 2008], MBSC [Tan L. N., Alwan A., 2013], SWIPE [Camacho A., Harris J. G., 2008], WU [Wu M., Wang D., Brown G. J., 2003], YIN [De Cheveigne A., Kawahara H., 2002], PEFAC [Gonzalez S., Brookes M., 2011], IRAPT [Azarov E., Vashkevich M., Petrovsky A., 2012]. Они вобрали в себя идеи из разных типов предшествующих алгоритмов.

Целью диссертационной работы является создание алгоритма оценки параметров полигармонической модели голосового сигнала, который превосходит аналогичные алгоритмы по точности и имеет эффективную вычислительную сложность.

Для достижения поставленной цели необходимо было решить следующие **задачи**.

1. Получить алгоритм оценивания всех параметров стационарной и аффинной полигармонической модели речевого сигнала.
2. Получить алгоритм приближённого расчёта ЧОТ для стационарной полигармонической модели речевого сигнала на коротких фреймах, имеющий сложность $N \log N$ и установить взаимосвязь между точностью и скоростью работы алгоритма.
3. Провести сравнение с существующими алгоритмами оценивания ЧОТ.

Основные положения, выносимые на защиту.

1. Получен способ оценивания комплексных амплитуд для аффинной полигармонической модели речевого сигнала (Теорема 1). На основе утверждения для аффинной модели утверждение теоремы распространяется для стационарного случая.
2. Получен критерий для оценивания ЧОТ аффинной полигармонической модели речевого сигнала (Следствие 1). На основе утверждения для аффинной модели следствие распространяется для стационарного случая.
3. Получена формула явного вычисления значений функционала качества $J_{\min}$ для стационарной модели голосового сигнала (Теорема 2).
4. Получен метод приближённого вычисления значений функционала качества (Лемма 1, Теоремы 3, 4, Следствие 2).
5. Получена аналитическая оценка точности аппроксимации функционала качества (Теоремы 5, 6).

Научная новизна. Все результаты, выносимые на защиту, являются новыми.

Практическая значимость Полученные результаты обеспечивают высокую точность оценки параметров полигармонических моделей голосово-

го сигнала на коротких промежутках времени, содержащих около двух периодов для стационарной модели и около трёх периодов для аффинной модели. Они позволяют эффективно моделировать короткие аллофоны, а также переходные процессы с высокой точностью.

Достоверность полученных результатов подтверждается доказательством всех сформулированных утверждений, а также сравнением результатов с известными алгоритмами на опубликованной открытой базе данных.

Апробация работы. Основные положения диссертационной работы доложены на Международной конференции Speech and Computer (2015), Международной конференции Image Analysis and Processing (2015), Всероссийском совещании по проблемам управления (2014) и на семинарах кафедры теоретической кибернетики математико-механического факультета СПбГУ.

Работы [1—3] написаны в соавторстве. В работе [1] автору принадлежат формулировки и доказательства основных теорем, выносимых на защиту. В работе [2] автором сформулирован критерий выбора периода основного тона, близкий к методу максимума правдоподобия. Продемонстрировано сравнение с известными алгоритмами оценки ЧОТ на общедоступной базе, тем самым продемонстрирована прикладная значимость результатов. В [3] диссертантом алгоритм оценивания ЧОТ был применён к прикладной задаче сегментации речи.

Работа поддержана Санкт-Петербургским государственным университетом, проект номер 6.37.349.2015.

Основные результаты работы внедрены при выполнении прикладных научных исследований по теме «Разработка методов лингвистического и семантического анализа для интеллектуальной обработки текстов, полученных в результате автоматического распознавания звучащей спонтанной русской речи» в рамках Соглашения с Министерством образования и науки РФ №14.579.21.0008 от 05.06.2014 (ID проекта RFMEFI57914X0008)

Публикации. Основные результаты по теме диссертации изложены в 4 печатных изданиях [1—4], 3 из которых изданы в журналах, рекомендованных ВАК [1—3].

Содержание работы

Во **введении** формулируются понятие частоты основного тона, описывается область исследований.

Первая глава диссертационной работы содержит основные понятия и описание наиболее популярных алгоритмов оценивания ЧОТ. Приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, сформулированы научная новизна и практическая значимость представляемой работы.

Во второй главе диссертационной работы представлены основные алгоритмы оценивания параметров полигармонических моделей речевого сигнала, оптимизация сложности которых составляет основное содержание всей работы.

Сформулирован и доказан способ выбора периода основного тона, близкий к методу максимума правдоподобия, который обобщает несмещённый критерий оценки периода основного тона из работы [Griffin D. W., 1988] на случай коротких промежутков времени и аффинных моделей.

Пусть $s = (s_t)_{t=-N/2}^{N/2-1}$ — анализируемый участок голосового сигнала длины N отсчётов. Аффинная модель сигнала определяется как

$$\widehat{s}_t = \sum_{k=-M}^{M} \left(A_k e^{\frac{2\pi i}{P} kt} + B_k \frac{t}{N} e^{\frac{2\pi i}{P} kt} \right), \qquad -\frac{N}{2} \le t \le \frac{N}{2} - 1,$$

где P — период основного тона, M — число гармоник, A_k, B_k — комплексные амплитуды, $A_k = \bar{A}_{-k}$, $B_k = \bar{B}_{-k}$ для всех k. Полный набор параметров модели содержит значение P и векторы $A = (A_k)_{k=0}^{M}$ и $B = (B_k)_{k=0}^{M}$. При $B = 0$ модель называется стационарной.

Величина $F = N/P$ есть ЧОТ, выраженная в количестве периодов сигнала на выбранном промежутке времени. Число гармоник M может быть различным, но максимальная частота гармоники, равная FM, не должна превосходить частоту Найквиста, равную $N/2$. Поэтому обычно выбирают $M = [P/2]$. В дальнейшем асимптотические формулы при $N \to \infty$ и при фиксированном F соответствуют неограниченному увеличению частоты дискретизации сигнала на фиксированном промежутке времени.

Требуется оценить все параметры модели по вектору измерений s. Естественным критерием качества модели речевого сигнала $\widehat{s}_t(P,A,B)$ выглядит среднеквадратичная невязка

$$E(P,A,B) = \sum_{n=-N/2}^{N/2-1} |w_t(s_t - \widehat{s}_t(P,A,B))|^2,$$

где $w_t = (1+\cos(2\pi t/N))/2$ — подходящее окно. Минимизацию этой функции можно провести последовательно:

$$\min_{P,A,B} E(P,A,B) = \min_{P} \left[\min_{A,B} E(P,A,B) \right] = \min_{F} J_{\min}(F),$$

где $F = N/P$. Результат минимизации $J_{\min}(F)$ функционала $E(P,A,B)$ по векторам комплексных амплитуд A и B вычисляется по методу наименьших квадратов (МНК).

Точка минимума функционала $E(P,A,B)$ в действительности плохо подходит для оценки параметров данной модели. В частности, множество моделей с фиксированным значением P и произвольными A, B содержится в классе аналогичных моделей с удвоенным периодом $2P$. Отсюда минимум E всегда достигается на удвоенном периоде независимо от сигнала s, что не соответствует принятому понятию высоты звука.

Эффективный подход к оценке величины P был предложен в [Griffin D. W., 1988] и назван несмещённым критерием оценки периода основного тона. Идея состоит в оценке дисперсии белого шума, входящего в модель сигнала, что приводит к оценке сигнала, часто совпадающей с методом максимума правдоподобия.

В статье [Griffin D. W., 1988] нет доказательств, много допущений и нестрогих переходов, а формулировки верны только для достаточно больших временных интервалов. В следующей теореме сформулирован и доказан более

общий результат. Введём обозначения.

$$\begin{aligned}\Big(r_{P,N}(f) \;\; r_{PQ,N}(f) \;\; r_{Q,N}(f)\Big) &= \frac{1}{N}\sum_{t=-N/2}^{N/2-1} w_t^2 e^{-\frac{2\pi i}{N}ft}\left(1 \;\; \tfrac{t}{N} \;\; \tfrac{t^2}{N^2}\right),\\ \Big(\omega_{P,N}(f) \;\; \omega_{PQ,N}(f) \;\; \omega_{Q,N}(f)\Big) &= \frac{1}{N}\sum_{t=-N/2}^{N/2-1} w_t^4 e^{-\frac{2\pi i}{N}ft}\left(1 \;\; \tfrac{t}{N} \;\; \tfrac{t^2}{N^2}\right).\end{aligned}$$

Столбец $(r_{P,N}(kF))_{k=0}^{2M}$ порождает самосопряжённую тёплицеву матрицу R_N^P размера $2M+1$. Аналогично порождаются самосопряжённые тёплицевы матрицы R_N^{PQ}, R_N^Q. Самосопряжённые тёплицевы матрицы W_N^P, W_N^{PQ}, W_N^Q размера $2M+1$ определяются соответствующими функциями $\omega_N(f)$. Определим блочно тёплицевы матрицы

$$R_N = \begin{pmatrix} R_N^P & R_N^{PQ} \\ R_N^{PQ} & R_N^Q \end{pmatrix}, \qquad W_N = \begin{pmatrix} W_N^P & W_N^{PQ} \\ W_N^{PQ} & W_N^Q \end{pmatrix}.$$

Функции $r_{P,N}, r_{PQ,N}, r_{Q,N}, \omega_{P,N}, \omega_{PQ,N}, \omega_{Q,N}$ имеют поточечные пределы, и по ним определяются матрицы R_∞, W_∞ того же размера $2(2M+1)$. Пусть при фиксированном F задана асимптотически линейная функция $M = M(N)$. Предположение регулярности состоит в существовании и равенстве пределов

$$\lim_{N\to\infty}\frac{1}{2M+1}\operatorname{tr}(R_N^{-1}W_N) = \lim_{N\to\infty}\frac{1}{2M+1}\operatorname{tr}(R_\infty^{-1}W_\infty) = h_\infty(F).$$

Теорема 1. *Пусть сигнал* $s = (s_t)_{t=-N/2}^{N/2-1}$ *является случайным вектором:*

$$s_t = \sum_{k=-M}^{M}\left(a_k e^{\frac{2\pi i}{N}Fkt} + b_k\frac{t}{N}e^{\frac{2\pi i}{N}Fkt}\right) + v_t, \qquad -N/2 \le t \le N/2-1,$$

где F *- количество периодов на выбранном промежутке времени,*
$M = [N/(2F)]$ *целая часть числа, а* v_t *— белый шум с дисперсией* σ^2*.*

Пусть оценка сигнала задаётся моделью

$$\widehat{s}_t = \sum_{k=-M}^{M}\left(A_k e^{\frac{2\pi i}{N}Fkt} + B_k\frac{t}{N}e^{\frac{2\pi i}{N}Fkt}\right), \qquad -N/2 \le t \le N/2-1,$$

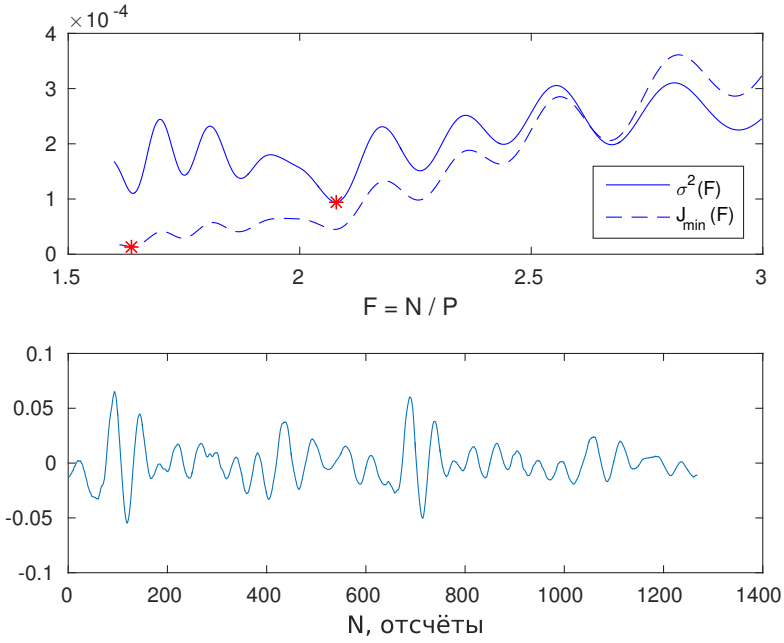

Рис. 1 — Нижний график: сигнал, содержащий $F_0 = 2.08$ периода. Верхний график: нормированная функция $J^s_{\min}(F)$ и функция $\widehat{\sigma}^2(F)$.

в которой амплитуды гармоник A_k, B_k рассчитываются по МНК с функционалом качества $E(P,A,B)$.

Тогда в предположении регулярности

$$\mathsf{E}J_{\min}(F) = \frac{3}{8}\,\sigma^2\left(1 - \frac{h_\infty(F)}{F}\right) + o(1) \qquad (N \to \infty).$$

Данное утверждение можно использовать для сравнения качества различных моделей, отличающихся только количеством F периодов в выбранном промежутке времени.

Следствие 1. *Пусть выполнены условия теоремы 1 и, в частности, фиксировано число F в уравнениях сигнала и модели. Тогда величина*

$$\widehat{\sigma}^2 = \frac{8}{3}\,\frac{J_{\min}}{1 - \frac{h_\infty(F)}{F}}$$

является асимптотически несмещённой оценкой дисперсии белого шума в сигнале.

В стационарном случае верно аналогичное утверждение. Для иллюстрации рассмотрим сигнал на нижнем графике рис. 1. На верхнем графике

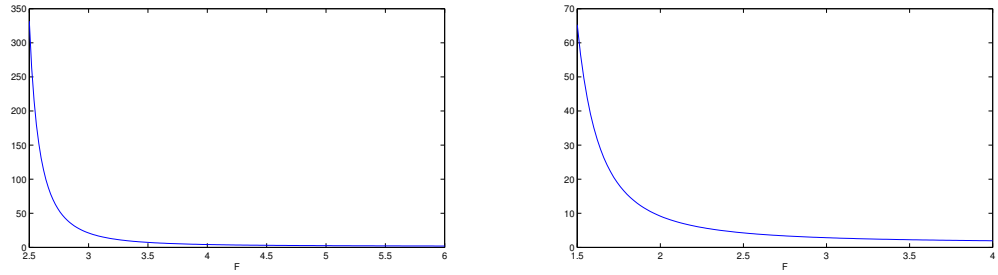

Рис. 2 — Множитель $H(F)$: слева для аффинной, справа для стационарной модели.

показана масштабированная функция $\alpha\, J_{\min}(F)$ и функция $\widehat{\sigma}^2(F)$. Минимумы этих функций достигаются при разных F и значение $F_0 = 2.08$ правильное. Из нижнего графика видно, что множитель

$$H(F) = \frac{1}{1 - \frac{h_\infty(F)}{F}},$$

связывающий величины $J_{\min}(F)$ и $\widehat{\sigma}^2(F)$, играет существенную корректирующую роль. Графики функций $H(F)$ для стационарной и аффинной моделей представлены на рис. 2.

Из графиков видно, что аффинная модель с частотой $F < 2.8$ практически не может быть идентифицирована. Проведённые расчёты показали, что на интервале $F \in [2.8, 3.6]$ функцию $h_\infty(F)$ можно приблизить многочленом

$$h_\infty(F) \approx -2.1967 + 2.8434 \cdot F - 0.3863 \cdot F^2, \qquad 2.8 \le F \le 3.6.$$

При $F \ge 3.6$ достаточно хорошим приближением является
$\lim_{F\to\infty} h_\infty(F) = 35/216\,(48\,\pi^2 - 385)/(2\,\pi^2 - 15) \approx 3.034$.

Идентификация стационарной модели требует не менее полутора периодов сигнала. На интервале $F \in [1.6, 3.0]$ функцию $h_\infty(F)$ для стационарной модели можно приблизить многочленом
$h_\infty(F) \approx -1.2635 + 3.0399 \cdot F - 0.9621 \cdot F^2 + 0.1018 \cdot F^3$, $1.6 \le F \le 3$. При $F \ge 3$ достаточно хорошим приближением является
$\lim_{F\to\infty} h_\infty(F) = 35/18 \approx 1.944$.

В разделе 2.3.2 доказано, что формула "несмещённого критерия периода основного тона" из [Griffin D. W., 1988] совпадает с критерием $\widehat{\sigma}^2$ для стационарной модели, в котором вместо $h_\infty(F)$ подставлено предельное зна-

чение 1.944. Поэтому этот критерий не изменился при $F > 3$. На коротких промежутках времени, включая пример из рис. 1, результат оценивания ЧОТ в соответствии со следствием 1 точнее, чем в [Griffin D. W., 1988].

Функция $\widehat{\sigma}^2(\cdot)$ может иметь большое количество локальных минимумов, поэтому необходимо провести перебор по некоторой сетке аргументов. Вычислительная сложность расчёта показателей $J_{\min}(F)$ в этом случае может оказаться недопустимо высокой.

В качестве такой сетки аргументов в [Griffin D. W., 1988] для стационарной модели были удачно выбраны все целые значения $P = N/F$. Был предложен алгоритм одновременного вычисления всех показателей $J_{\min}$ для всех целых P со сложностью порядка $N \log_2 N$. Как и весь подход, этот алгоритм предназначен для длинных интервалов времени с $F > 3$, что доказано в разделе 3.1.2. Эффективные численные алгоритмы расчёта $J_{\min}$ для целых P в общем случае основаны на следущей теореме.

Теорема 2. *Пусть P — целое. В классе стационарных моделей, при $F \geq 1.6$, минимум среднего квадрата невязки сигнала и модели равен*

$$J_{\min} = \frac{1}{N}\left(\sum_{t=-N/2}^{N/2-1} w_t^2 s_t^2 - \frac{1}{F}\sum_{m=0}^{P-1}\frac{|y_m|^2}{C_m}\right),$$

где при $0 \leq m \leq P-1$

$$y_m = \sum_{n=N_P^-(m)}^{N_P^+(m)} \widetilde{s}_{m+nP}, \quad C_m = \frac{8}{3F}\sum_{n=N_P^-(m)}^{N_P^+(m)} w_{m+nP}^2,$$

где $\widetilde{s}_t = w_t^2 s_t$ *при* $-N/2 \leq t \leq N/2-1$ *и* $N_P^+(m) = \left[\frac{N/2-1-m}{P}\right]$, $N_P^-(m) = \left[\frac{-N/2-m}{P}\right] + 1$.

В третьей главе диссертационной работы сформулированы результаты, позволяющие для стационарной полигармонической модели речевого сигнала получить быстрый вычислительный алгоритм оценивания ЧОТ. Также устанавливается связь между скоростью работы этого алгоритма и и точностью вычисления функционала качества $J_{\min}$.

Задача поиска значений $J_{\min}$ для целых периодов основного тона P сводится к расчёту значений

$$\phi(P) = \sum_{m=0}^{P-1} \frac{|y_m(P)|^2}{C_m(P)}$$

по всем целым значениям P в промежутке допустимых значений $P \in [P_{\min}, P_{\max}]$ где $P_{\min}$ это небольшое число, определяемое частотой дискретизации, а $P_{\max} \approx 5/8N$. Числа $C_m(P)$ можно считать затабулированными. Для вычисления значения $\phi(P)$ по его определению требуется порядка N^2 операций, что требуется сократить до $N \log N$.

В частном случае, когда $P \leq N/3$, можно считать, что $C_m \approx 1$. В этом случае алгоритм из [Griffin D. W., 1988] рассчитывает значения $\phi(P)$ для всех целых P со сложностью, пропорциональной $N \log N$. В общем случае при $F < 3$ потребуется баланс между скоростью и точностью.

Определим функцию вещественной переменной

$$\widehat{\eta}_0(x) = \frac{4}{3\pi} \int_{-\pi}^{\pi} \cos^4 \frac{t}{2}\, e^{itx}\, dt = \frac{4 \sin(\pi x)}{\pi x(x^2-1)(x^2-4)}.$$

Это непрерывная функция и $\widehat{\eta}_0(x) = \mathcal{O}(x^{-5})$ при $x \to \infty$.

Теорема 3. *1. При всех $F > 0$*

$$C_m(P) = d_P(z_P^m), \qquad z_P = e^{-\frac{2\pi i}{P}}, \quad 0 \leq m \leq P-1,$$

где функция d_P определяется рядом Лорана

$$d_P(z) = \sum_{n=-\infty}^{\infty} \widehat{\eta}_0(nF) z^n, \qquad |z| = 1.$$

2. Для любого $F \geq 1.6$ функция $d_P(z)$ приближается следующим образом:

$$\left| d_P(z) - |K_F|^2 \, |1 - \alpha_F z|^2 \right| \leq 0.01, \qquad |z| = 1,$$

где

$$K_F = \frac{1}{2}\left(\sqrt{1 + 2\widehat{\eta}_0(F)} + \sqrt{1 - 2\widehat{\eta}_0(F)} \right), \qquad \alpha_F = -\frac{\widehat{\eta}_0(F)}{K_F^2}.$$

В соответствии с теоремой 3 коэффициент $1/C_m(P)$ приближается геометрической прогрессией:

$$\frac{1}{C_m(P)} \approx \frac{8}{3F|K_F|^2}\,|g_P(z_P^m)|^2, \qquad g_P(z) = \frac{1}{1-\alpha_F z} = \sum_{k=0}^{\infty} \alpha_F^k\, z^k.$$

Заменим исходную функцию $\phi(P)$ соответствующей аппроксимацией:

$$\phi(P) \approx \frac{8}{3F|K_F|^2}\phi_0(P), \qquad \phi_0(P) = \sum_{m=0}^{P-1} |g_P(z_P^m) y_m|^2.$$

Введём обозначение $v_P(t) = s_t w_t^2 g_P(z_P^t)$ при $-N/2 \le t \le N/2-1$.

Лемма 1. *Пусть P — целое число, $1 \le P < N/2$. Тогда функция ϕ_0 может быть вычислена по формуле*

$$\phi_0(P) = r_P(0) + 2\sum_{k=1}^{[N/P]} r_P(kP),$$

где $r_P(t)$ — корреляционная функция сигнала $v_P(t)$, дополненного нулями при $|t| \ge N/2$.

Количество слагаемых в суммах из леммы 1 по всем P пропорционально $N \log N$. Поэтому сложность расчёта функции ϕ_0 определяется сложностью расчёта значений $r_P(kP)$ с целыми k.

Введём обозначения для сигнала $\widetilde{s}_t = s_t w_t^2$ и, соответственно, для ДПФ

$$\widetilde{S}_n = \sum_{t=-N/2}^{N/2-1} \widetilde{s}_t e^{-\frac{2\pi i}{2N} tn}, \quad F_{n,j} = \sum_{t=-N/2}^{N/2-1} \widetilde{s}_t \left(\frac{2t}{N}\right)^j e^{-\frac{2\pi i}{2N} tn}$$

при $-N \le n \le N-1$ и $j \ge 0$. Введём также обозначение

$$\rho_{\ell,j}(\tau) = \frac{1}{2N}\sum_{n=-N}^{N-1} \widetilde{S}_n^* F_{n+\ell,j} e^{\frac{2\pi i}{2N}\tau n}, \qquad j \ge 0, \quad \ell \ge 0, \quad |\tau| < N,$$

причём $F_{n,j}$ продолжается периодически по n, так что последняя свёртка циклическая. При $k \ge 0$ округлим число $2kF$ до ближайшего целого числа $2kF = \ell_{kF} + x_{kF}$, где $|x_{kF}| \le 0.5$ и ℓ_{kF} целое.

Теорема 4. *При* $|\tau| < N$

$$r_P(\tau) = \frac{1}{1-\alpha_F^2}\left[\frac{1}{2N}\sum_{n=-N}^{N-1}|\widetilde{S}_n|^2 e^{\frac{2\pi i}{2N}\tau n} + 2\,\mathrm{Re}\sum_{k=1}^{\infty}\alpha_F^k\sum_{j=0}^{\infty}\frac{(-\pi i x_{2Fk}/2)^j}{j!}\rho_{\ell_{2Fk},j}(\tau)\right].$$

Приближение функции r_P основано на выборе конечного множества пар неотрицательных индексов $M = \{(\ell, j)\}$, для которых вычисляются величины $\rho_{\ell,j}(\tau)$.

Следствие 2. *Общее количество преобразований Фурье, необходимых при вычислении приближения*

$$\widehat{r}_{P,M}(\tau) = \frac{1}{1-\alpha_F^2}\left[\rho_{0,0}(\tau) + 2\,\mathrm{Re}\sum_{(k,j):(\ell_{2Fk},j)\in M}\alpha_F^k\frac{(-\pi i x_{2Fk}/2)^j}{j!}\rho_{\ell_{2Fk},j}(\tau)\right].$$

для $-N \le \tau < N$, *равно*

$$N_{\mathrm{fft}} = |M| + J_{\max} + 1.$$

Для повышения качества аппроксимации есть смысл включать в множество M все пары (ℓ, j) при $0 \le j \le J(\ell) - 1$. Далее будем считать, что это условие выполнено.

Максимальное число ℓ, для которого $(\ell, 0) \in M$, обозначим через L.

Пусть P — целое число и $F = N/P$. Максимальное целое число k, для которого $2Fk \le L + 0.5$, обозначим $K(P) = \lfloor P(2L+1)(4N)\rfloor$.

Теорема 5. *Погрешность приближения функции* ϕ_0 *оценивается сверху следующим образом:*

$$|\phi_0(P) - \widehat{\phi}_0(P)| \le \sum_{t=-N/2}^{N/2-1}\gamma_P(t)|\widetilde{s}_t|\left(|\widetilde{s}_t| + 2\sum_{q=1}^{\lfloor\frac{t}{P}+\frac{F}{2}\rfloor}|\widetilde{s}_{t-qP}|\right),$$

где

$$\gamma_P(t) = \frac{2}{1-\alpha_F^2}\left[\sum_{k=1}^{K(P)}|\alpha_F|^k\frac{1}{J(\ell_{2Fk})!}\left|\frac{\pi t}{N}x_{2Fk}\right|^{J(\ell_{2Fk})} + |\alpha_F|^{K(P)+1}\frac{1+|\alpha_F|}{|1+\alpha_F z_P^t|^2}\right]$$

u $z_P^t = e^{-\frac{2\pi i}{2N}2Ft}$.

Для каждого вычета $k = 0,1,\ldots,P-1$ минимальное число из промежутка $-N/2+1 \le t \le N/2-1$, сравнимое с k, обозначим t_k^0, а количество чисел из этого промежутка, сравнимых с k, обозначим N_k.

Теорема 6. *Для каждого* $P \in [P_{\min},\, P_{\max}]$

$$|\phi_0(P) - \widehat{\phi}_0(P)| \le \lambda \|s_w\|^2,$$

где $s_w = (s_{w,t})_{t=-N/2}^{N/2-1}$, $s_{w,t} = s_t w_t$,

$$\lambda = \max_{0\le k\le P-1} \lambda_k, \quad \lambda_k = \|A_k\|, \quad a_{i,j,k} = c_{,k} i c_{j,k} d_{\max\{i,j\},k}, \quad 0 \le i,j \le N_k - 1,$$
$$c_{n,k} = w_{t_k^0+nP}, \quad d_{n,k} = \gamma_P(t_k^0 + nP).$$

В четвёртой главе диссертационной работы продемонстрирована работа алгоритма оценивания ЧОТ из главы 3, а также проведено сравнение качества оценивания ЧОТ с наиболее известными существующими алгоритмами, такими как: YAAPT [Zahorian S. A., Hu H., 2008], SWIPE [Camacho A., Harris J. G., 2008], RAPT [Talkin D., 1995], PEFAC [Gonzalez S., Brookes M., 2011], и YIN [De Cheveigne A., Kawahara H., 2002]. В качестве тестовой была выбрана база PTDB-TUG [Pirker G., 2011]. Извлечено 300 голосовых участков из разных аудиозаписей, содержащихся в базе. Затем, были получены оценки ЧОТ всеми вышеуказанными алгоритмами (их реализации на Matlab могут быть найдены в открытом доступе). После оценки ЧОТ были получены значения комплексных амплитуд и вычислены значения функции невязки $J_{\min}(P_0)$ для кандидатов ЧОТ, нормированы на энергию сигнала и усреднены по всем экспериментам. В таблице 1 представлены значения $J_{\min}(P_0)$ для разных вариантов применения высокочастотных фильтров к исходному сигналу и к смоделированному. Через PTDB отмечен результат, полученный по разметке ЧОТ, имеющийся в базе данных PTDB-TUG.

Предложенный в диссертационной работе алгоритм оценивания ЧОТ назван «метод минимизации дисперсии шума» (МДШ), или «Noise Variance Minimization» (NVM).

Частота среза	PTDB	YAAPT	SWIPE	RAPT	PEFAC	YIN	**NVM**
0 Гц	0.3318	0.6025	0.2900	0.3147	0.3046	0.2781	**0.2709**
200 Гц	0.3531	0.6281	0.3092	0.3328	0.3221	0.2958	**0.2885**
400 Гц	0.3763	0.6467	0.3311	0.3534	0.3429	0.3163	**0.3095**
600 Гц	0.3895	0.6561	0.3445	0.3682	0.3577	0.3304	**0.3241**
1000 Гц	0.4082	0.6675	0.3635	0.3887	0.3795	0.3504	**0.3450**
2000 Гц	0.4300	0.6793	0.3861	0.4131	0.4049	0.3735	**0.3693**

Таблица 1 — Среднее значение для $J_{\min}(P_0)$ после высокочастотной фильтрации сигнала

Заключение.

В рамках этой диссертационной работы получены результаты, позволяющие эффективно оценивать параметры полигармонических моделей речевого сигнала. Результаты представляют практическую ценность в задачах, где критична точность локализации старших гармоник речевого сигнала, а также в условиях коротких временных интервалов и быстрого изменения параметров сигнала.

Основные результаты работы заключаются в следующем.

1. Рассчитаны асимптотические коэффициенты в алгоритме оценивания комплексных амплитуд для стационарной и аффинной полигармонических моделей речевого сигнала.
2. Получены алгоритмы оценивания частоты основного тона для стационарной и аффинной полигармонических моделей речевого сигнала на основе несмещённой оценки дисперсии шума модели.
3. Получен алгоритм приближённого расчёта частоты основного тона для стационарной полигармонической модели речевого сигнала на коротких фреймах, имеющий асимптотическую сложность $N \log N$, и установлена связь между его точностью и скоростью работы.
4. Проведено сравнение разработанного алгоритма расчёта частоты основного тона с существующими алгоритмами.

Публикации автора по теме диссертации

1. *Melnikov A.*, *Barabanov A.* **Guaranteed estimation of speech fundamental frequency with bounded complexity algorithm** // Cybernetics and Physics. — 2016. — Т. 5, № 1.
2. *Barabanov A.*, *Melnikov A.*, *Magerkin V.*, *Vikulov E.* **Fast Algorithm for Precise Estimation of Fundamental Frequency on Short Time Intervals** // Lecture Notes in Computer Science. Т. 9319. — Springer, 2015. — С. 217—225.
3. *Melnikov A.*, *Akhunzyanov R.*, *Kudashev O.*, *Luckyanets E.* **Audiovisual Liveness Detection** // Lecture Notes in Computer Science. Т. 9280. — Springer, 2015. — С. 643—652.
4. *Мельников А.* Быстрый алгоритм идентификации параметров модели голосового сигнала // Сборник трудов XII Всероссийского совещания по проблемам управления (ВСПУ-2014). — 2014. — С. 3090—3101.