

Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

В.Б. Гитлин,
доктор технических наук

Д.А. Лузин



Дано описание структурной схемы алгоритма автоматического выделителя основного тона (ОТ). В алгоритме реализована совместная работа локального и интегрального выделителей ОТ. Локальный выделитель ОТ использует метод GS, интегральный — автокорреляционную функцию спектра. Решение ТОН / НЕ ТОН принимается на втором проходе алгоритма по признакам, вычисляемым синхронно с ОТ. Приведены результаты сопоставительных испытаний предлагаемого алгоритма с работой кепстрального, автокорреляционного, фильтрового методов, метода Рабинера-Голда и метода ЛЛК.

Abstract

Fully automatic speech pitch extraction algorithm is described. The algorithm is composition of the team-working local and integral pitch extracting methods. The local pitch extracting method has used the method GS, the integral method is founded on autocorrelation function of speech spectrum. Voice / Unvoice decision is made by means of the speech features, calculated synchronous with pitch, at the second pass of the algorithm. Testing results of peak-peaking, cepstral, autocorrelation, filter, Gold-Rabiner, LLK methods and the proposed algorithm are presented.

Потребность в разработке автоматического выделителя мгновенных значений периода (частоты) ОТ

В работах [1]–[3] дан обзор современного состояния проблем, связанных с автоматическим распознаванием, обработкой и синтезом речи. Авторы указанных работ подчёркивают чрезвычайно сложный характер проблем, связанных с речевой тематикой. Дальнейшее



Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

решение как научных, так и практических задач в этой области требует решения более конкретных, возможно, даже более «узких» задач, от которых предшествующие отличались широкой неопределённой постановкой [4], [5]. К числу задач, требующих первоочередного решения, можно отнести задачи выделения первичных признаков речевого сигнала (таких, как основной тон) в условиях реальной речевой обстановки [6].

Основной тон является одним из основных параметров речевого сигнала. В нём представлена информация об интонационной структуре произнесения [7]–[11], об индивидуальности голоса диктора [12] и его эмоциональном состоянии [13], [14], о возрастных [15] и патологических изменениях голосового аппарата. В мгновенных значениях частоты ОТ, в моментах включения и отключения голосового источника содержится информация о фонемном составе и скорости произнесения [8], [16], об окружающей обстановке [17], о способе произнесения (нормальная, форсированная, несвязная речь) [18]. В тональных языках (китайский, вьетнамский и т.п.) ОТ играет роль смысловозначительного признака [19]. Характерным признаком живой речи, улучшающим её восприятие, служат микровариации ОТ [20]. Диапазон ОТ влияет на качество воспринимаемого голоса и оценку эмоционального состояния диктора [13].

Через интонацию передаётся информация о достаточно высоких уровнях лексико-синтаксической структуры текста, которая наиболее близка к целостному смысловому образу сообщения [7]–[11], [21]. Просодические характеристики, одной из составляющих которых служит контур ОТ, могут быть использованы для определения коммуникативной направленности высказывания (утверждение, вопрос), синтаксических особенностей (перечисление, противопоставление), модальности (приказ, удивление), для выделения главного и общего, для получения информации об окружающей диктора среде. Интонационное членение в значительной степени определяет стратегию анализа текста слушателем, делает распознавание более помехоустойчивым и оптимальным с точки зрения понимания целостного смысла сообщения [22], [23].

Информация, содержащаяся в контуре ОТ, находит своё отражение в значениях как мгновенных частот ОТ на коротких интервалах времени, особенно в начале вокализации, после пауз, на границах синтаксических единиц, так и в относительно медленных изменениях мелодии ОТ, имеющих глобальный характер. Переносимая ОТ информация есть существенное отличие акустического сигнала речи от письменной речи. Письменную речь можно рассматривать как некоторую упрощённую модель акустического речевого сообщения.

Статистические параметры основного тона: диапазон значений, скорость изменения, стандартное отклонение и т.д. — определяют структуру и параметры устройств выделения основного тона. Эти устройства входят в большинство систем распознавания речи, устройств компрессии речи и являются одним из основных и наиболее сложных блоков. Например, синтез речи для интервалов, соответствующих периодам ОТ, позволил повысить качество синтезированной речи [24].

Модель генерации интонационного контура нельзя считать окончательно сформированной. Взаимодействие факторов, влияющих на контур ОТ, нелинейно и имеет довольно сложные зависимости, с трудом поддающиеся описанию. Пока не удалось разработать удовлетворительной классификации этих данных, объяснить все случаи их подобия и расхождения.

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

Выделитель частоты ОТ, который позволил бы извлекать максимум информации из траектории частоты ОТ, должен определять мгновенные значения частоты ОТ на всём интервале работы голосового источника, включая моменты начала и окончания вокализации. Знание мгновенных значений частоты ОТ позволяет сохранить всю информацию, содержащуюся в контуре ОТ, включая изменения, связанные с мелодией ОТ. Этот алгоритм должен работать в автоматическом режиме для произвольного речевого сигнала в условиях реальной речевой обстановки. Алгоритм должен быть достаточно прост и удобен для встраивания в другие программы по обработке речи. Интерфейс алгоритма должен быть дружелюбным по отношению к произвольному, в том числе незнакомому с особенностями алгоритма, пользователю.

Описание алгоритма

Среди локальных методов выделения ОТ следует отдать предпочтение пиковым методам [25], [26]: они нечувствительны к смещению нулевого уровня, не требуют нормализации по амплитуде, простановка марок начала возбуждения речевого тракта по максимумам сигнала технически проще. Однако пиковые методы обладают рядом существенных ограничений [24], [25]: 1. Подвержены сильному влиянию помех. 2. Чувствительны к подавлению низкочастотного диапазона. 3. Плохо работают на переходных участках между вокализованными и невокализованными звуками, когда интенсивность речевого сигнала значительно изменяется. 4. Адаптация параметров пиковых выделителей затруднена.

Предложенный в работах [27], [28], [29] GS — метод выделения ОТ в существенной мере позволяет подавить влияние мешающих факторов, присутствующих пиковым методом. **Рисунок 1** поясняет принцип действия GS-метода. Сплошной линией показана функция логарифма положительной полуволны речевого сигнала $S(t)$, штриховой линией — функция решений $g(t)$, марками M_1, M_2, M_3 — главные максимумы речевого сигнала. Функцию решений $g(t)$ генерируют на каждом интервале ОТ. Конкретную форму речевого сигнала между марками не рассматривают. Возможность перестройки функции решений в ходе выделения траектории ОТ является существенным достоинством метода GS.

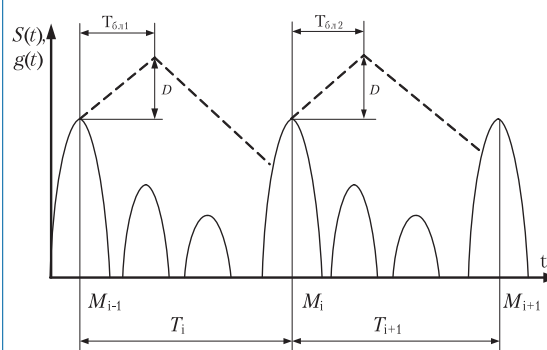


Рис. 1. Принцип работы метода GS

На интервале нарастания $g(t)$ (интервале блокировки) решение о новом импульсе ОТ не принимают. Спадающая часть функции $g(t)$ должна обеспечить выделение момента возбуждения следующего периода ОТ. Спадающую часть функции решения строят по формуле:

$$g(t) = S(M_i) + D - \frac{DA}{QT_{ож}(i)} \cdot (t - t_j), \quad (1)$$

где $S(t)$ — логарифм положительной полуволны речевого сигнала; t — текущее время; D — приращение максимума функции решения, оцениваемое исходя из статистики речи; t_j — момент времени, соответствующий началу спадающей части функции решений; A — коэффициент асимметрии функции решения относительно нарастающей её части; $T_{ож}$ — ожидаемое значение текущего периода ОТ; Q — коэффициент заполнения периода ОТ (отношение времени блокировки к длительности периода ОТ). После того, как момент пересечения $g(t)$ и $S(t)$ найден, выполняют процедуру поиска максимума лепестка сигнала $S(t)$, с которым произошло



Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

пересечение. Координата максимума соответствует началу следующего периода ОТ, т.е. марке M_{i+1} . Далее находят длительность текущего периода ОТ T_i :

$$T_i = M_i - M_{i-1} \quad (2)$$

Решение «ТОН / НЕ ТОН» в работе [29] принимают на основе следующих признаков, вычисляемых синхронно с ОТ на периоде основного тона [30]:

- 1) Логарифм частоты пересечения нулевого уровня сигналом (ЛПЧН)

$$Z = K_z \lg(M/T_0), \quad (3)$$

где K_z — нормирующий коэффициент; M — количество пересечений нулевого уровня на периоде основного тона.

- 2) Логарифм энергии сигнала:

$$E = K_e \cdot \lg\left(\varepsilon + \sum_{i=1}^N x_i^2\right), \quad (4)$$

где K_e — нормирующий коэффициент; N — длительность периода основного тона в отсчётах; x_i — отсчёты исходного речевого сигнала; ε — малая величина, предотвращающая логарифмирование нулевой суммы.

- 3) Нормализованный коэффициент корреляции с единичной задержкой:

$$R(1) = K_c \cdot \left(\frac{\sum_{i=1}^{N-1} x_i \cdot x_{i+1}}{\sum_{i=1}^N x_i^2} \right), \quad (5)$$

где K_c — нормирующий коэффициент; N — длительность периода основного тона в отсчётах; x_i — отсчёты исходного речевого сигнала.

Из признаков Z , E , $R(1)$ формируют обобщённый признак $G(i)$, где i — номер текущего периода ОТ:

$$G = K_G \frac{R(1) \cdot E}{Z_{cr}}. \quad (6)$$

Область значений обобщённого признака $G(i)$ при помощи порогов g_0 , g_1 , g_2 разбивают на четыре подобласти: «Уверенно НЕ ТОН», «Неуверенно НЕ ТОН», «Неуверенно ТОН» и «Уверенно ТОН». Окончательное решение Т/НТ принимают на этапе постобработки в соответствующем блоке.

Начальное значение периода ОТ $T_{ож}$ на переходе от невокализованного интервала к вокализованному, значения параметров Q , A , g_0 , g_1 , g_2 в работе [28] оператор определял экспериментально и оптимальные значения подбирал индивидуально для каждого диктора. Кроме того, в алгоритме, предложенном в работе [28], полярность речевого сигнала, поступающего на обработку, не оценивали. Отсутствие оценки полярности сигнала, поступающего на обработку, снижает точность оценки частоты ОТ. В целом основные недостатки алгоритма, изложенного в работах [27]–[30] (алгоритм GS1), можно представить следующим образом.

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

Перед каждым новым произнесением путём использования априорных сведений о входном речевом сигнале алгоритм GS1 требует:

1. Установки начального значения периода (частоты) ОТ на переходах от невокализованного сигнала к вокализованному.
2. Настройки параметров алгоритма, таких как асимметрия функции решения.
3. Коррекции значений порогов принятия решения Т/НТ, которые необходимо подстраивать при смене условий записи речевого сигнала.

В ходе дальнейших исследований были предложены способы оценки полярности входного речевого сигнала и способы поиска оптимального значения коэффициента асимметрии функции решений [31], [32]. Было постановлено принимать решение Т/НТ окончательно на втором проходе алгоритма после оценки статистических параметров сигнала. Наиболее существенное изменение состоит в следующем. Для поиска начальной оценки значения периода (частоты) ОТ и для коррекции ошибочных решений локального выделителя ОТ организована совместная работа локального и интегрального алгоритмов выделения ОТ [33]–[35]. Структурная схема модифицированного составного алгоритма (алгоритм GS2) приведена на рисунке 2. Алгоритм состоит из четырёх основных блоков: блок предварительной обработки, блок простановки марок, блок начальной оценки частоты ОТ и блок принятия решения «ТОН / НЕ ТОН».

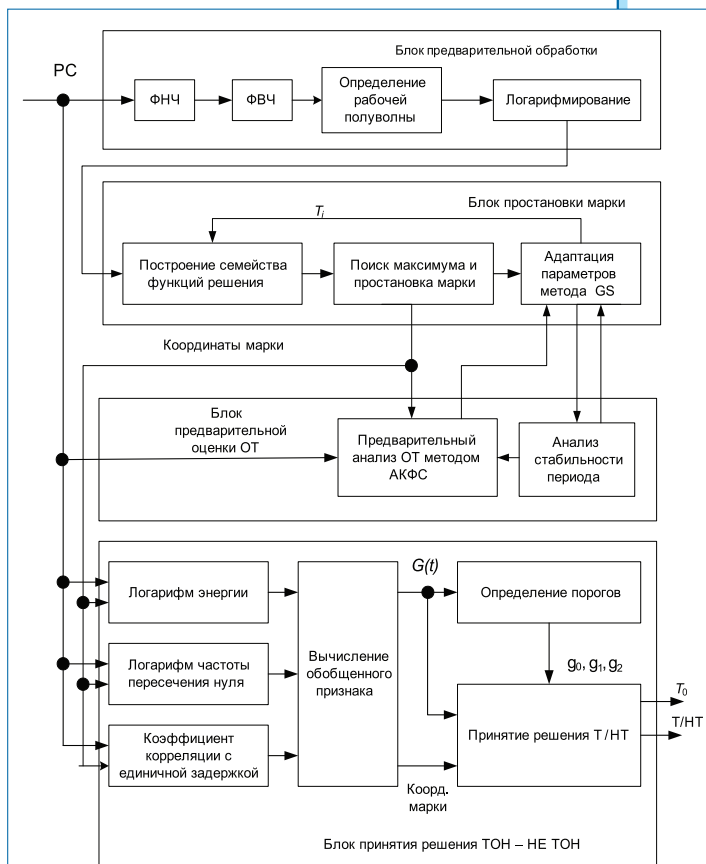


Рис. 2. Структурная система составного алгоритма выделения основного тона

Блок предварительной обработки. В состав блока предварительной обработки входят: фильтр нижних частот (ФНЧ) [36], фильтр верхних частот (ФВЧ), определитель рабочей полуволны сигнала и логарифматор. В отличие от работы [28], добавлен субблок определения полуволны речевого сигнала.

Работа субблока основана на оценке асимметрии речевой волны [31]. В качестве критерия, по которому определяют асимметрию речевой волны, используют суммарную разность A площадей положительной и отрицательной полуволн центрально ограниченного речевого сигнала (рис. 3). При вычислении суммарной разности используют не менее пяти уровней центрального ограничения.

Блок простановки марок состоит из следующих субблоков: субблок построения семейства функций решения; субблок поиска максимума и простановки марки; субблок адаптации параметров алгоритма GS2 к значению нового периода ОТ. В отличие от работы [28], в субблоке построения функций решения для каждого нового периода ОТ строят *семейство* функций решения (рис. 4) [32].

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра



Рис. 3. Осциллограмма центрально-ограниченного речевого сигнала

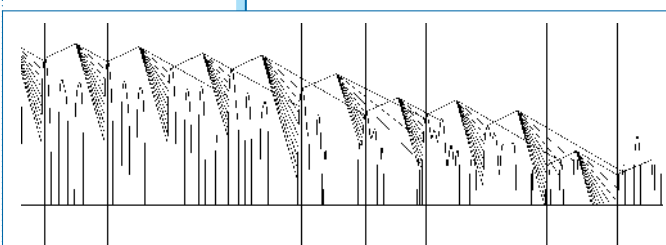


Рис. 4. Семейство функций решений для мужского голоса на переходе от вокализованного звука к невокализованному



Рис. 5. Структурная схема метода оценки частоты основного тона речи с использованием автокорреляционной функции спектра

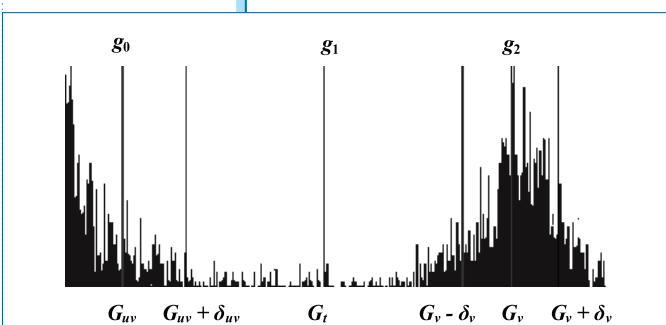


Рис. 6. Гистограмма обобщённого признака Т/НТ всех произнесений диктора АЮ

Каждая из функций решения определяет положение максимума на новом лепестке сигнала, который рассматривают в качестве кандидата в марки нового периода ОТ. В субблоке поиска максимума и простановки марки для каждого из максимумов, найденного семейством функций решения, вычисляют длительность предполагаемого нового периода ОТ и отношение d этой длительности к длительности предшествующего периода ОТ.

Блок предварительной оценки периода ОТ. Из множества максимумов, найденных функциями решений, выбирают тот [32], который имеет наименьшее значение коэффициента d . В месте расположения этого максимума ставят марку нового периода ОТ. Если отношение длительности нового периода ОТ к длительности предшествующего периода ОТ превышает порог $d_{\text{пор}} = 0,25$ или порог изменения амплитуды найденных соседних периодов ($K_L = 0,3$), то инициализируется работа алгоритма начальной оценки частоты ОТ с использованием автокорреляционной функции спектра (методом АКФС). Структурная схема этого алгоритма показана на рисунке 5 [33]. Длительность нового периода ОТ, соответствующую проставленной марке, передают в субблок адаптации параметров (рис. 2).

В блоке принятия решения Т/НТ на первом проходе производят расчёт обобщённого признака G (формула (6)) для каждого найденного периода ОТ. На втором проходе выполняют построение гистограммы распределения обобщённого признака G и определение интервалов гистограммы: уверенно «ТОН», уверенно «НЕ ТОН», неуверенно «ТОН», неуверенно «НЕ ТОН» — с последующим отнесением каждого периода к тому или иному интервалу (рис. 6). Экспериментально установлено, что минимальная ошибка принятия решения Т/НТ обеспечивалась при условии выбора порогов:

$$g_0 = G_{uv}, g_1 = G_t, g_2 = G_v,$$

где G_{uv} — среднее значение обобщённого признака невокализованных интервалов произнесения, G_v — среднее значение обобщённого

признака вокализованных интервалов произнесения, G_t — значение обобщённого признака, разделяющего вокализованные и невокализованные области гистограммы. На последнем шаге алгоритма принимают окончательное решение Т/НТ и корректируют ошибки для всего произнесения в целом.

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

Экспериментальная оценка надёжности выделения ОТ совместным алгоритмом (алгоритм GS2)

Проведены сопоставительные испытания алгоритма выделения ОТ, работающего в полностью автоматическом режиме (алгоритм GS2), и алгоритмов выделения ОТ, основанных на пиковом, фильтровом, автокорреляционном, кепстральном методах, методе Рабинера-Голда и методе ЛЛК, реализованных в системе SIS 5.0 [37]. Все методы, кроме ЛЛК, требовали ручной подстройки параметров алгоритма выделения ОТ и параметров алгоритма принятия решения «ТОН / НЕ ТОН». В методе ЛЛК подбирали ручным способом только параметры алгоритма принятия решения Т/НТ. Параметры всех алгоритмов (кроме алгоритма GS2) устанавливали таким образом, чтобы обобщённая ошибка выделения ОТ была минимальной. Ошибки определяли путём сравнения измеренного контура ОТ с эталонным [38]. Эталонный контур получали ручным способом по осциллограммам. Отклонения от эталонного контура, не превышающие 5%, не учитывали. Отклонения в диапазоне 5–15% относили к малым ошибкам. Ошибки, превышающие 15%, — к большим.

Результаты испытаний алгоритмов выделения ОТ для общей группы голосов (15 дикторов, 38 произнесений)

Выделитель ОТ	Ошибка Т/НТ, %	Ошибка НТ/Т, %	Т/НТ _{сп} , %	Большие ошибки, %	Малые ошибки, %	Обобщённая ошибка, %	Отношение Т/НТ _{сп} / обобщённая ошибка
Чистый сигнал							
GS2	1.97	2.37	2.17	1.70	6.16	2.76	0.79
Пиковый	0.62	7.27	3.94	1.23	10.06	4.13	0.95
Кепстральный	3.89	5.50	4.70	3.76	19.95	6.02	0.78
АКФ	1.67	21.44	11.55	2.98	10.00	11.93	0.97
Рабинер-Голд	1.93	8.52	5.23	2.87	9.59	5.96	0.88
Фильтровой	0.11	14.83	7.47	1.16	9.02	7.56	0.99
ЛЛК	0.64	6.29	3.47	1.08	6.63	3.63	0.95
Сигнал с аддитивным шумом С/Ш = 5 дБ							
GS2	2.01	18.18	10.10	15.58	20.72	18.56	0.54
Пиковый	1.47	36.22	18.84	7.03	24.26	20.11	0.94
Кепстральный	1.95	37.90	19.92	11.85	36.75	23.18	0.86
АКФ	2.63	36.30	19.46	3.35	13.56	19.75	0.99
Рабинер-Голд	2.10	32.57	17.33	4.00	17.23	17.79	0.97
Фильтровой	1.38	41.44	21.41	2.81	22.55	21.59	0.99
ЛЛК	1.01	45.93	23.47	2.13	20.61	23.57	1.00
Сигнал ограничен полосой телефонного канала							
GS2	0.61	14.74	7.67	6.19	6.38	9.86	0.78
Пиковый	0.90	19.50	10.20	12.08	9.89	15.81	0.65
Кепстральный	4.98	15.29	10.14	2.91	19.03	10.55	0.96
АКФ	1.06	46.60	23.83	5.36	11.06	24.42	0.98
Рабинер-Голд	2.37	19.00	10.69	28.76	5.22	30.68	0.35
Фильтровой	0.10	37.56	18.83	10.10	6.81	21.37	0.88
ЛЛК	0.10	37.56	18.83	10.10	6.81	21.37	0.88

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

Обобщённую ошибку вычисляли по формуле:

$$\delta_{об} = \sqrt{\delta_{бол}^2 + T/HT_{cp}^2},$$

где $\delta_{об}$ — обобщённая ошибка, $\delta_{бол}$ — большая ошибка, T/HT_{cp} — средняя ошибка T/HT.

Результаты испытаний представлены в таблице. Примеры контуров ОТ (диктора-мужчины VBG и диктора-женщины GNA), полученных для сигнала с аддитивным шумом при соотношении сигнал / шум 5 дБ, представлены на [рис. 7](#) и [рис. 8](#). Отметим что, несмотря на меньшее значение обобщённой ошибки, метод Голда-Рабинера менее предпочтителен по сравнению с методом GS2. Указанное несоответствие объясняется тем, что большее количество ошибок HT/T в методе Голда-Рабинера привело к снижению количества больших ошибок из-за уменьшения длины вокализированных интервалов при худшем отображении общего хода траектории ОТ. В целом алгоритм GS2, работающий в полностью автоматическом режиме, оказался предпочтительным по сравнению с остальными методами, использующими ручную подстройку. Нам известен ещё один полностью автоматический метод выделения ОТ [39]. Однако во время экспериментов мы не располагали возможностью сравнения с этим методом.

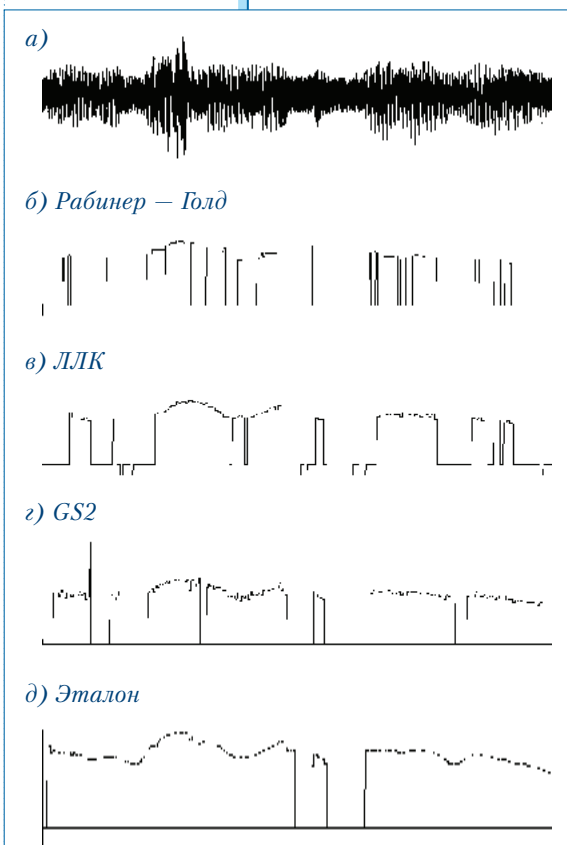


Рис. 7. Диктор VBG, фраза «Не видали мы такого невода», сигнал при С/Ш = 5 дБ

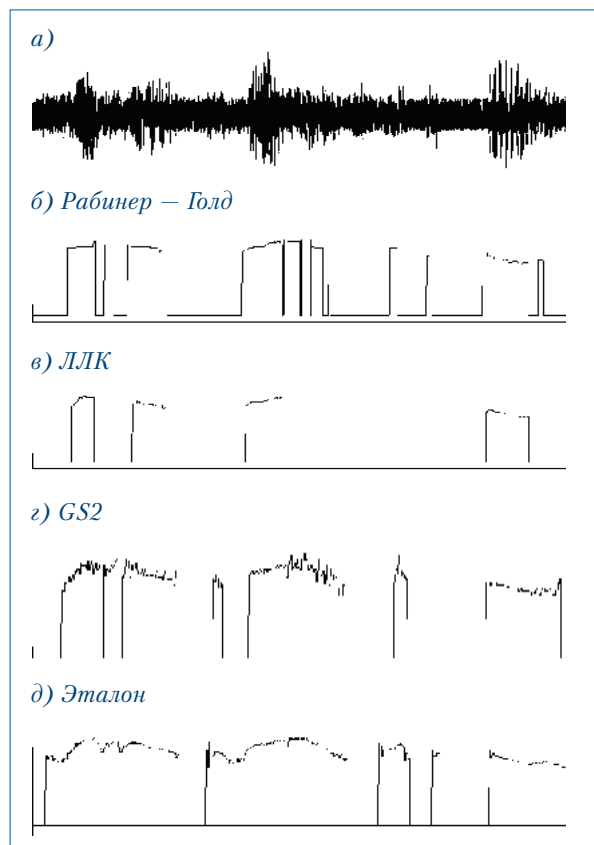


Рис. 8. Диктор GNA, фраза «Жирные сазаны ушли под палубу», сигнал при С/Ш = 5 дБ

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам G5 и автокорреляционной функции спектра

Литература

1. Потапова Р.К. Перспективы развития прикладного речеведения // Речевые технологии. № 1, 2008. С. 5–17.
2. Сорокин В.Н. Фундаментальные исследования речи и прикладные задачи речевых технологий // Речевые технологии. № 1, 2008. С. 18–48.
3. Суханов В.А., Жигулевцев Ю.Н. Исследования методов речевого диалогового управления в МВТУ — МГТУ им. Н.Э. Баумана // Речевые технологии. №1, 2008. С.78–82.
4. Хитров М.В. Распознавание русской речи: состояние и перспективы // Речевые технологии. № 1, 2008. С. 83–87.
5. Ламин К.В. // Речевые технологии. №1, 2008. С. 88–92.
6. Михайлов В.Г. Из истории исследований преобразования речи // Речевые технологии. №1, 2008. С. 93–113.
7. Кодзасов С.В., Кривнова О.Ф. Общая фонетика: Учебник. М.: Рос. гос. ун-т, 2001.
8. Чистович Л.А., Венцов А.В., Гранстрём М.П. и др. Физиология речи. Восприятие речи человеком. Л.: Наука, 1976.
9. Галунов В.И. Принципы переработки сложных речевых сообщений // Автоматическое распознавание слуховых образов: Тез. докл. и сообщ. 12 Всесоюзн. школы-семинара. Киев, 1982. С.349–351.
10. Лобанов Б.М. Исследование и разработка методов автоматического синтеза речи по фонемному тексту / Автореферат диссертации на соискание учёной степени доктора технических наук. Рига, 1984.
11. Светозарова Н.Д. Интонационная система русского языка. Л.: ЛГУ, 1982.
12. Кузнецов П.Г., Гитлин В.Г. Идентификация голосов по средней частоте основного тона // Применение вычислительной техники в машиностроении. Ижевск, 1877. С. 68–74.
13. Галунов В.И., Пиктурна В.В., Янушавичус В.Й. Акустические корреляты эмоциональной речи // Акустика речи и слуха: Материалы докл. и сообщ. 5 Всесоюзн. совещ.-симпозиума. Одесса, 1989. С. 16–25.
14. Галунов В.И. О возможности определения эмоционального состояния говорящего по речи // Речевые технологии. №1, 2008. С. 60–66.
15. Linville S.E., and Fisher H.B. Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females // J. Acoust. Soc. Am. 1985. 78, № 1 (Part 2). P. 40–48.
16. Златоустова Л.В. Фонетические единицы русской речи. М.: МГУ, 1981.
17. Bond Z.S., and Moore T.J. Speech produced under adverse circumstances // Eleventh Int. Congr. Phonetic Sci.: Proc. XI ICPHS. Tallin, 1987. Vol. 2. P. 73–76.
18. Kolesnikov V.M., and Zakharov L.M. Acoustic and perception of speech in various modes of articulation // Eleventh Int. Congr. Phonetic Sci.: Proc. XI ICPHS. Tallin, 1987. Vol. 2. P. 207–210.
19. Высоцкий Г.Я., Нгуен Ань Туан, Трунин Донской В.Н. Исследование фонетики тонального языка и автоматическое распознавание тонированных слогов // Автоматическое распознавание слуховых образов: Тез. докл. и сообщ. 10 Всесоюзн. школы-семинара. Тбилиси: Мецниереба, 1978. С.112–114.
20. Ryalls J.H. and Lieberman P. Fundamental frequency and vowel perception // J. Acoust. Soc. Am. 1982. 72, № 5. P.1631–1634.
21. Галунов В.И. О возможности определения эмоционального состояния говорящего по речи // Речевые технологии. № 1, 2008. С. 60–66.
22. Кривнова О.Ф. Алгоритмы просодического оформления синтезированных высказываний // Автоматическое распознавание слуховых образов: Тез. докл. и сообщ. 15 Всесоюзн. семинара. Таллин, 1989. С. 25–31.
23. Касевич В.Б., Шабельникова Е.М., Рыбин В.В. Ударение и тон в языке и речевой деятельности. Л.: Изд-во ЛГУ, 1990.
24. Лобанов Б.М. О развитии речевых технологий в Белоруссии // Речевые технологии № 1, 2008. С. 48–59.



Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

25. Сапожков М.А. Речевой сигнал в кибернетике и связи. М.: Связьиздат, 1963.
26. Гитлин В.Б. Основной тон речевого сигнала / Деп. В ВИНТИ, 1998. № 1206-В98.
27. Гитлин В.Б. Применение понятия о генерируемом солитоне для выделения особенностей речевого сигнала // Распознавание образов и анализ изображений: новые информационные технологии. 4-я Всероссийская с международным участием конференция. РОАИ-98. 1998 г. Новосибирск, 1998. Часть 1. С. 64–68.
28. Архипов И.О., Гитлин В.Б. Метод выделения основного тона на основе понятия о генерируемом солитоне. // Распознавание образов и анализ изображений: новые информационные технологии. 4-я Всероссийская с международным участием конференция. РОАИ-98. 1998 г. Новосибирск, 1998. Часть 1. С. 23–27.
29. Архипов И.О., Гитлин В.Б. Оценка точности выделения основного тона методом GS. // Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. М.: «ГЕОС», 1999. С. 38–42.
30. Архипов И.О. Сегментация речи по признаку ТОН/НЕ_ТОН синхронно с основным тоном. // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск, изд. ИПМ УрО РАН 1998. С. 5–8. В.Б. Гитлин, Д.А. Лузин. Алгоритм поиска оптимального значения коэффициента асимметрии в выделителе основного тона по методу GS // Сб. тр. XVIII сессии Российского акустического общества. Том 3. М.: «ГЕОС», 2006. С.11–14.
31. Гитлин В.Б., Лузин Д.А. Интегральная оценка полярности речевого сигнала, поступающего на вход локального выделителя ОТ // Информационные системы в промышленности и образовании: Сб. тр. Молод. учёных. Вып.2. Ижевск, 2007. С. 46–50.
32. Гитлин В.Б., Лузин Д.А. Алгоритм поиска оптимального значения коэффициента асимметрии в выделителе основного тона по методу GS // Сб. тр. XVIII сессии Российского акустического общества. Том 3. М.: «ГЕОС», 2006. С.11–14.
33. Лузин Д.А. Алгоритм начальной оценки основного тона речи для выделителя основного тона речи по методу АКФС // Сб. тр. XVIII сессии Российского акустического общества. Том 3. М.: «ГЕОС», 2006.
34. Гитлин В.Б., Лузин Д.А. Стабильность отношений последовательных периодов основного тона речи как критерий переключения совместно работающих выделителей основного тона // Сб. тр. XVIII сессии Российского акустического общества. Том 3. М.: «ГЕОС», 2006. С.23–26.
35. Лузин Д.А. Экспериментальная оценка точности выделения основного тона речевого сигнала по автокорреляционной функции спектра // Сб. тр. XIX сессии Российского акустического общества. Том 3. М.: «ГЕОС», 2007. С. 19–22.
36. Архипов И.О., Гитлин В.Б. Оценка частоты среза ФНЧ, используемого для выделения основного тона. // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск, изд. ИПМ УрО РАН 1998. С. 12–16.
37. Методические рекомендации по практическому использованию программы SIS при работе с речевыми сигналами / Центр речевых технологий. Санкт-Петербург, 1997.
38. Архипов И.О., Гитлин В.Б. Оценка точности выделения основного тона методом GS. // Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. М.: ГЕОС, 1999. С. 38–42.
39. Зубов Г.Н., Хитров М.В. Состояние и перспективы голосовой биометрии (по итогам участия ЦРТ в Voice Biometrics Conference 2007). «Центр речевых технологий». Россия.

Гитлин В.Б., Лузин Д.А. Совместный алгоритм выделения основного тона речи по методам GS и автокорреляционной функции спектра

Гитлин Валерий Борисович —

доктор технических наук, профессор кафедры «Вычислительная техника» Ижевского государственного технического университета, Заслуженный деятель науки Удмуртской республики. Ведёт дисциплины «Схемотехника ЭВМ», «Цифровая обработка сигналов». Окончил Ижевский механический институт в 1962 г. Методами обработки речи занимается с 1967 года. Кандидатскую диссертацию защитил в Каунасском политехническом институте в 1974 г., диссертацию на степень доктора технических наук — в Ижевском государственном техническом университете в 2000 г. Опубликовано более 125 работ по речевой тематике и тематике, связанной с учебно-методической деятельностью на кафедре.

Лузин Дмитрий Александрович —

аспирант кафедры «Вычислительная техника» Ижевского государственного технического университета. Окончил университет в 2003 г. и защитил диплом на звание инженера-системотехника. Опубликовано более 20 работ, связанных с обработкой речи.