

ВЫДЕЛЕНИЕ ОСНОВНОГО ТОНА РЕЧИ МЕТОДОМ SWIPE ИЗ СИГНАЛА, ОГРАНИЧЕННОГО ПОЛОСОЙ ТЕЛЕФОННОГО КАНАЛА

Вашурин Д.Ю.,

Гитлин В.Б.

Рассмотрена работа алгоритмов SWIPE и SWIPE' при выделении основного тона речевого сигнала, ограниченного полосой телефонного канала. Показано, что для реальных сигналов, ограниченных полосой частот телефонного канала, указанные алгоритмы склонны к переходу на гармоники основного тона. Предложен алгоритм, позволяющий повысить надёжность выделения основного тона алгоритмами SWIPE и SWIPE' для сигналов указанного типа.

• частота основного тона • скалярное произведение • пилообразный сигнал • интенсивность вокализации

SWIPE and SWIPE' pitch extraction algorithms were considered for speech signal, which is limited by telephone bandwidth. Prove that these algorithms tend to move to the harmonic pitch for real signals that bandlimited during regular telephone speech transmission. A pitch extraction algorithm was proposed for signals of this type. This algorithm improves the reliability of the pitch extraction by SWIPE and SWIPE' algorithms.

• pitch • scalar product • sawtooth signal • vocalization intensity

Введение

Вокализированные звуки речи образуются с участием голосового источника, одним из параметров которого является частота основного тона (ОТ). Частота ОТ является переносчиком большого количества информации, содержащейся в речевом сигнале. В нём представлена информация о интонационной структуре произнесения [1]...[5], о индивидуальности голоса диктора [6], [7] и его эмоциональном состоянии [8], [9], о возрастных [10] и патологических изменениях голосового аппарата. В мгновенных значениях частоты ОТ, в моментах включения и отключения голосового источника содержится информация о



фонемном составе и ритме произнесения [2], [11]. В тональных языках (китайский, вьетнамский и т.п.) основной тон играет роль смысловозначительного признака [12]. В частоте основного тона отражается информация о способе произнесения (нормальная, форсированная, несвязная речь) [13], информация об окружающей обстановке [14]. Характерным признаком живой речи, улучшающим её восприятие, служат микровариации основного тона [15]. Существенны значения 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 [16].

Выбор алгоритма выделения ОТ зависит от задач, которые поставлены перед выделителем ОТ. В системах связи, применяющих алгоритмы сжатия речи [17]...[21], выделение частоты ОТ выполняют в реальном масштабе времени. Допускаются погрешности в оценке частоты ОТ, которые несущественны при восприятии синтезированной в вокодер речи. В других применениях, например, в задачах распознавания человека по голосу [22], [23], возможно существенное увеличение времени получения траектории ОТ для конкретного произнесения. Одновременно, требования к точности и надёжности измерений параметров траектории основного тона могут ужесточиться. Если речь анализируется синхронно с основным тоном, либо выполняется анализ мгновенных изменений частоты ОТ, то необходимы оценка моментов возбуждения речевого тракта со стороны голосовых связок [16] и измерение длительности каждого периода ОТ [24]. В системах параметрического синтеза речи, в вокодерных системах, в системах измерения высоты звучания музыкальных инструментов этого не требуется [17], [25], [26]. При анализе интонационных характеристик речи поведение контура ОТ информативно в течение всего произнесения, в том числе на интервалах отсутствия вокализации и в паузах [4]. Если же требуются измерения частоты ОТ на интервалах начала и окончания вокализации, то, кроме оценки частоты ОТ, необходимо адекватное принятие решения о наличии вокализации в речи (решение «ТОН/НЕ ТОН» или короче (Т/НТ и НТ/Т)) [16]...[18], [28].

Работы по созданию новых алгоритмов выделения основного тона продолжают достаточно интенсивно. Можно отметить появившиеся в последнее время ряд новых алгоритмов выделения ОТ [24]...[32]. Возникает необходимость анализа этих алгоритмов с точки зрения их технических возможностей и области применения.

Алгоритмы SWIPE и SWIPE'

В отличие от автокорреляционных методов [25] в алгоритмах SWIPE и SWIPE' [26] частота ОТ оценивается по максимуму корреляции речевого сигнала с эталонным образцом, имеющим спектр, близкий к спектру анализируемого звука. В алгоритмах SWIPE и SWIPE' параметры эталонного образца изменяются таким образом, чтобы достичь максимально возможного сходства с анализируемым речевым сигналом. Выполняется адаптация эталона к речи путём изменения его основной частоты. После достижения максимального сходства, частота основного тона эталонного образца принимается за частоту основного тона сигнала.

В качестве адаптируемого эталонного образца в алгоритмах SWIPE и SWIPE' выбран пилообразный сигнал, что и определило название метода (Sawtooth Inspired Pitch Estimator) [26]. Спектр эталонного пилообразного образца авторами работы [26] модифицирован следующим образом. Дискретные гармоники периодического пилообразного сигнала с амплитудой, уменьшающейся обратно пропорционально частоте, заменены двуполярными лепестками косинуса. Сравнение анализируемого сигнала с адаптируемым эталонным образцом выполняют путём вычисления скалярного произведения между спектром сигнала и ядром интегрального преобразования в виде лепестков косинуса. Вычисление скалярного произведения производят в преобразованной шкале частот, близкой к шкале барков. С целью повышения меры сходства между гармониками анализируемого сигнала и косинусоидальными лепестками адаптируемого эталона для каждого нового значения частоты ОТ эталона выполняют пересчёт спектра анализируемого сигнала с новым весовым окном. В качестве весового окна выбрано окно Ханна, имеющее ширину, определяемую значением частоты ОТ эталона. Значение вычисленного скалярного произведения принимают за меру интенсивности предполагаемой частоты ОТ в анализируемом звуке. Окончательно за частоту ОТ анализируемого звука принимают то значение частоты адаптируемого эталона, которой соответствует максимум интенсивности. Одновременно, значение вычисленной интенсивности служит мерой вокализованности кадра анализа. Если максимум вычисленной интенсивности на кадре меньше заданного порога, обозначенного как «STHR», то анализируемый кадр сигнала считают невокализованным, в противном случае – вокализованным.

В алгоритме SWIPE' в отличие алгоритма SWIPE в расчёте скалярного произведения принимают участие только гармоники с простыми (не кратными) номерами, что уменьшает вероятность принятия решения о переходе выделенной частоты ОТ на субгармоники. Методы SWIPE и SWIPE' относятся к интегральным [16]. Они определяют значение частоты ОТ на кадре анализа длиной в четыре периода ОТ, не способны оценивать моменты возбуждения речевого тракта. Траектория частоты ОТ, получаемая этими методами, является сглаженной. Оба метода требуют значительного количества вычислений, предварительного выбора пределов измерения частоты ОТ и установки порога STHR, необходимого для принятия решения Т/НТ.

Сравнительные испытания алгоритмов GS2, SWIPE и SWIPE'

Алгоритм выделения ОТ на основе метода GS2 [28], [33] в отличие от методов SWIPE и SWIPE' относится к локальным [16]. Он оценивает положение моментов возбуждения речевого тракта, не требует предварительной подстройки, может работать в реальном масштабе времени, если не учитывать принятие решения Т/НТ на втором этапе алгоритма. Выделитель частоты ОТ на основе метода GS2 [28], [33] основан на использовании особенностей временной функции речевого сигнала. В отличие от методов SWIPE и SWIPE' метод GS2 не определяет высоту звучания музыкальных инструментов [34].

Результаты совместных испытаний алгоритмов SWIPE, SWIPE' и GS2 представлены в табл. 1. В экспериментах участвовали 10 дикторов мужчин и 7 дикторов женщин. Испытания проводили на чистом сигнале, на сигнале с добавленным шумом при соотношении Сигнал / Шум, равном 5 дБ, на сигнале, ограниченном полосой телефонного канала 300...3400 Гц, и на предельно ограниченном (клиппированном) сигнале.

Анализировались ошибки Т/НТ (ошибочно принятое решение «ТОН», эталонный сигнал не вокализован), ошибки НТ/Т (ошибочно принятое решение «НЕ ТОН», эталонный сигнал вокализован) и большие ошибки. К большим ошибкам относились ошибки оценки частоты ОТ, отличавшиеся от эталонных более чем на 15 %.



Эталонные контуры частоты ОТ получены путём визуальной обработки осциллограмм речи. Анализ малых ошибок не выполняли, поскольку алгоритмы SWIPE и SWIPE' не предназначены для оценки тонкой структуры контура ОТ.

Таблица 1

Результаты испытаний методов GS2, SWIPE и SWIPE'

Метод	Сигнал	Ошибки (в %)			
		T/HT	HT/T	T/HT ср	Большие ошибки
GS2	Без шума	2,17	2,42	2,29	1,90
	С/Ш 5 дБ	2,01	17,20	9,65	15,24
	Телефон	2,21	14,74	8,45	6,19
SWIPE	Без шума	2,64	2,20	2,42	1,42
	С/Ш 5 дБ	0,59	15,50	8,05	6,90
	Телефон	2,77	18,41	10,59	32,92
SWIPE'	Без шума	3,00	2,19	2,59	1,31
	С/Ш 5 дБ	1,88	12,50	7,19	6,66
	Телефон	6,13	10,36	8,30	47,54

Оценку качества работы алгоритмов выполняли с помощью программы «Compare_work» [35], написанной на языке Borland C++. Программа позволяет визуально сравнивать три контура частоты ОТ: два исследуемых и один эталонный, подсчитывать количество ошибок T/HT, HT/T, малых и больших ошибок. Пример интерфейса работы программы показан на рис.1. Фраза «Не видали мы такого невода», диктор мужчина. На нижнем графике показан эталонный контур частоты ОТ, на среднем – контур ОТ, построенный методом SWIPE', на верхнем – методом SWIPE. В нижней части рисунка помещены результаты анализа исследуемых контуров частоты ОТ.

Анализ ошибок алгоритма SWIPE при обработке сигнала, ограниченного полосой телефонного канала

Результаты, представленные в табл. 1, показывают, что алгоритмы SWIPE, SWIPE' и GS2 обеспечивают сопоставимые результаты на чистом сигнале. На сигнале с добавленным шумом при соотношении С/Ш = 5 дБ алгоритмы SWIPE и SWIPE' были предпочтительнее. Но на сигнале, ограниченном полосой телефонного канала, указанные два алгоритма показали существенное увеличение количества больших ошибок по срав-

нению с алгоритмом GS2, причём большее – для алгоритма SWIPE'. Худшее качество работы алгоритма SWIPE' по сравнению с алгоритмом SWIPE можно объяснить тем, что в полосе телефонного канала количество гармоник основного тона для алгоритма SWIPE' будет меньше по сравнению с алгоритмом SWIPE. Большое количество ошибок на сигнале, ограниченном полосой телефонного канала, потребовало анализа неудовлетворительного качества работы алгоритмов SWIPE и SWIPE' на этом типе сигнала.

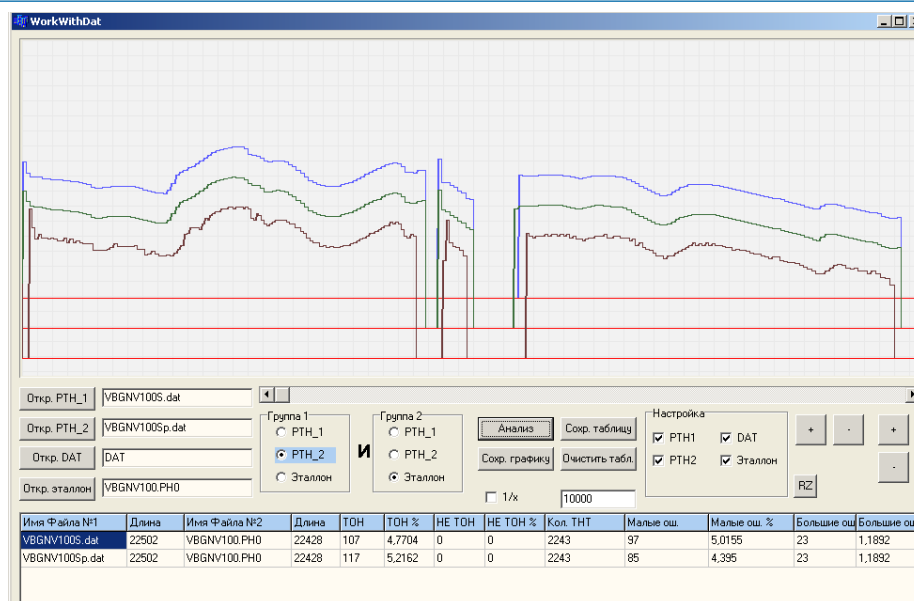


Рис. 1. Интерфейс программы «Compare_work»

Спектр речевого сигнала имеет спад примерно 6 дБ/октаву в сторону высоких частот [16]. Ядро интегрального преобразования алгоритмов SWIPE и SWIPE' соответствует среднему спектру речевого сигнала. В работе [26] показано, что понижение амплитуд лепестков ядра, пропорциональное частоте спектра сигнала, способствует защите алгоритмов SWIPE и SWIPE' от переходов на субгармоники. Однако для конкретных произнесений огибающая спектра звука может отличаться от указанной средней зависимости от частоты. Известно, что огибающая спектра сигнала сильно зависит от формантной структуры и от типа источника звука. Например, фонема /и/ имеет сильный спад огибающей спектра в полосе 500...1500 Гц и подъём в области 1700...2200 Гц [1], [16], [17]. Пример среднего спектра слога /не/ (фраза «Не видели мы такого невода», диктор VBG), иллюстрирующий отмеченные выше особенности спектра речи, показан на рис. 2. Показаны средние спектры чистого сигнала (без аддитивного шума) и без ограничения спектра – толстая линия) и сигнала, ограниченного полосой телефонного канала (тонкая линия). По оси абсцисс на рис. 2 отложена частота в Гц, по оси ординат – логарифм амплитуды. Спектр получен с помощью программы SIS.6x [22].

В случае ограничения полосы сигнала полосой телефонного канала (300...3400 Гц) низкочастотные гармоники спектра, включая частоту основного тона, оказываются подавленными. Отсутствие низкочастотных гармоник наряду с малыми значениями огибающей спектра в диапазоне 500...1500 Гц увеличивает долю высокочастотных компонент спектра в расчёте интенсивности вокализации. Повышение вклада высокочастотных компонент приводит к возрастанию вероятности перехода на гар-



моники ОТ. Необходимо также учесть влияние аппаратуры передачи и обработки речи.

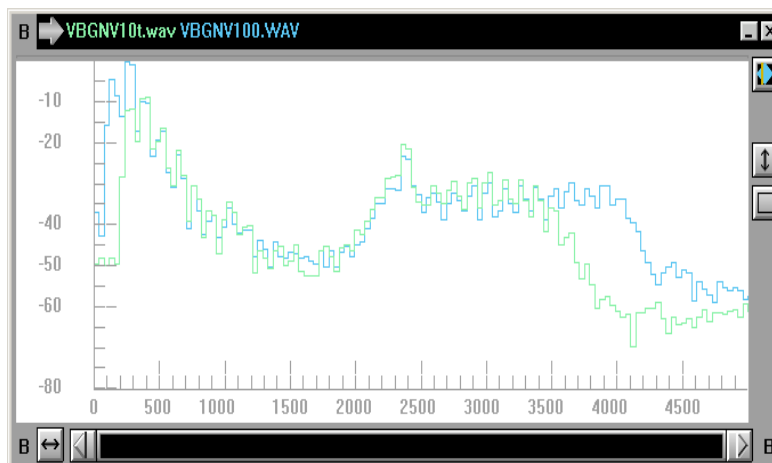


Рис. 2. Спектр слога /не/. Диктор VBG

В алгоритмах SWIPE и SWIPE' ещё одной причиной сбоя на гармоники ОТ может служить изменчивость гармонической структуры речевого сигнала. Гармоническая структура спектра речи сохраняется лишь до 2...2,5 кГц [5], [16], причём с повышением частоты форма спектральных составляющих гармоник размывается, выраженность максимума гармоники в спектре становится менее чёткой [1], [16], [17]. «Размытость» положения высокочастотных гармоник является характерным свойством речи. Она обусловлена квазипериодичностью речевого сигнала. Например, в работе И.С. Азарова, М.И. Вашкевича и А.А. Петровского [24] показано, что динамическое выравнивание периодов ОТ позволяет существенно повысить чёткость представления гармоник ОТ во всём диапазоне их существования.

На рис. 3 показаны осциллограмма и контур ОТ фразы «Не видели мы такого невода». Контур получен алгоритмом SWIPE для сигнала без шума, диктор VBG. Один из параметров алгоритма SWIPE, задаваемый при запуске алгоритма, – верхнее допустимое значение частоты ОТ, выбран равным $F_{0max} = 500$ Гц. За исключением оценки интервалов наличия и отсутствия вокализации, этот контур достаточно близко соответствует эталонному контуру ОТ, что можно отметить из сравнения с рис. 1. Частота ОТ на графике рис. 3 изменяется в диапазоне от 80 до 160 Гц.

На рис. 4 показаны контуры ОТ, полученные методами SWIPE' и SWIPE с помощью пакета MATLAB для сигнала, ограниченного полосой телефонного канала. Фраза и параметры алгоритма те же, что и на рис. 3. Видны многочисленные сбои на гармониках ОТ, значение частот которых приближается к F_{0max} .

На рис. 5 приведены сечения значений интенсивности ОТ как функции номера кандидата в частоты ОТ. Графики построены для моментов времени $t_1 = 200$ мс (верхняя панель) и $t_2 = 210$ мс (нижняя панель) от начала произнесения. На верхней панели максимум интенсивности соответствует истинному значению частоты ОТ. На нижней панели – сбойной ситуации.

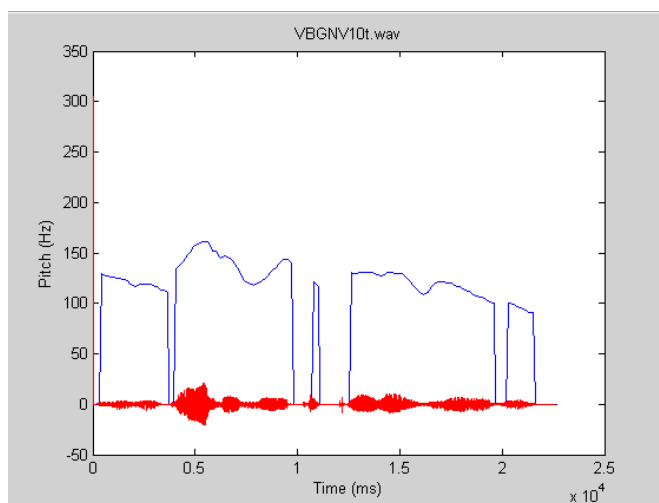


Рис. 3. Контур основного тона, полученный для чистого сигнала алгоритмом SWIPE

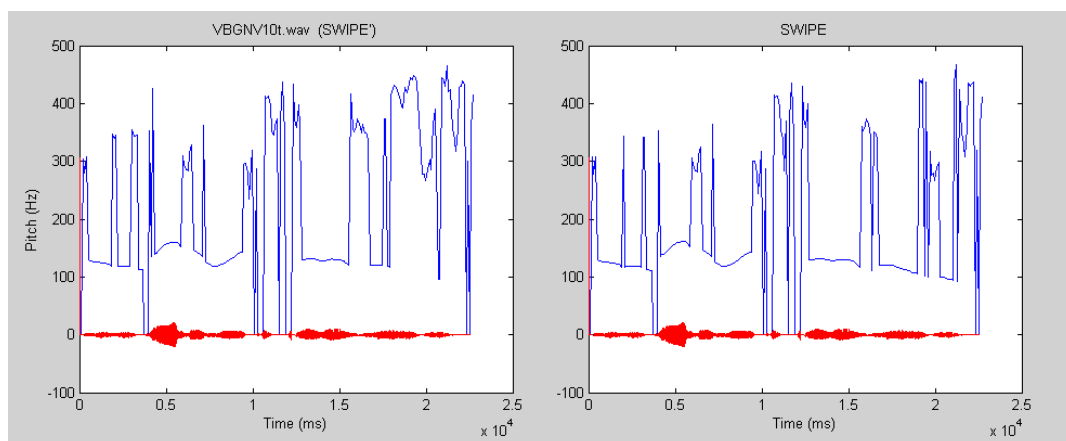


Рис. 4. Контурсы основного тона, полученные для сигнала, ограниченного полосой телефонного канала, методами SWIPE' и SWIPE

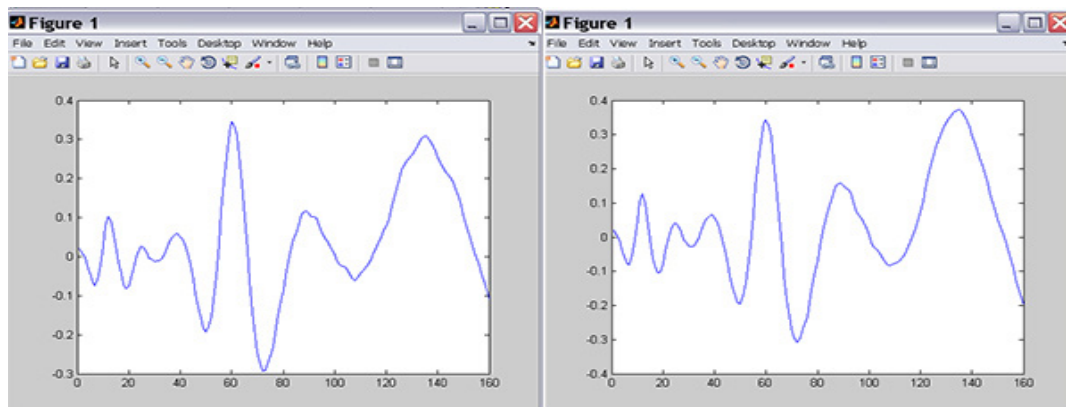


Рис. 5. Интенсивность кандидатов в ОТ



Из анализа рис. 5 можно сделать следующие выводы. Во-первых, высокочастотный максимум имеет более плавный характер и менее выражен по сравнению с низкочастотным максимумом, который соответствует истинному значению ОТ. Указанное свойство является следствием размывания спектра гармоник сигнала с повышением частоты. Сбои, вызванные переходом на более высокий по частоте максимум (рис. 4), можно объяснить указанным выше подчёркиванием высокочастотной части спектра в сигнале, ограниченном полосой телефонного канала. Во-вторых, информация об истинном значении частоты ОТ с большой вероятностью сохраняется и в графике интенсивности, соответствующем сбойной ситуации (рис. 5, нижняя панель). Следовательно, эту информацию можно тем или иным способом извлечь.

Предлагаемое изменение, вносимое в алгоритм SWIPE путём слежения за максимумом интенсивности

Учитывая эти результаты, в алгоритмы SWIPE и SWIPE' было решено сначала внести следующее изменение. Был добавлен поиск второго локального максимума на кривой интенсивности ОТ, соответствующей i -му моменту анализа t_i . В качестве второго дополнительного максимума выбирали максимум, расположенный непосредственно ниже основного максимума. Из найденных на $(i+1)$ -м моменте анализа двух максимумов решение принимали в пользу того, который был ближе к максимуму, принятому за основной в i -й момент времени.

На рис. 6 представлен пример работы модифицированных алгоритмов SWIPE и SWIPE' при выделении частоты ОТ сигнала, ограниченного полосой телефонного канала. Сравнивая эти результаты с результатами, показанными на рис. 4, видим, что количество больших ошибок существенно уменьшилось. Ошибки в основном возникали на невокализованных интервалах, неверно отнесённых алгоритмом к вокализованным.

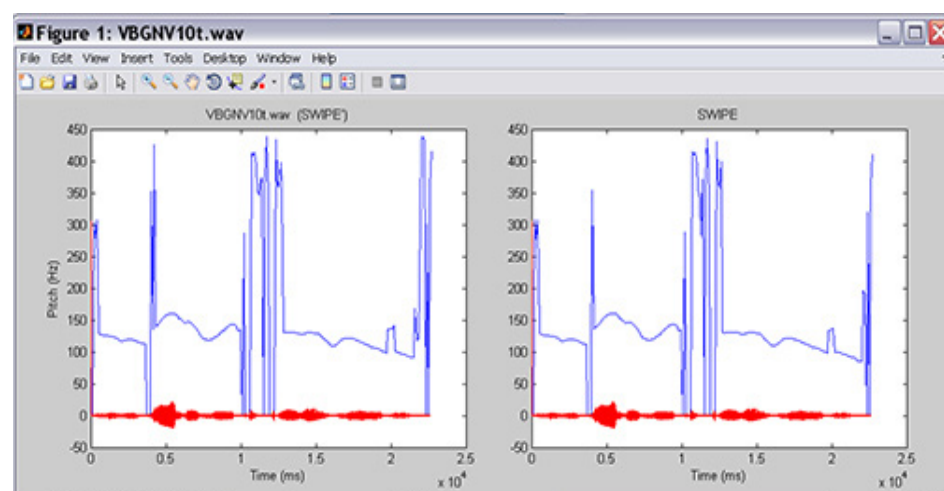


Рис. 6. Выделение ОТ модифицированными алгоритмами SWIPE и SWIPE'

В повторных испытаниях участвовали 20 дикторов мужчин и 17 дикторов женщин. Результаты этих испытаний исходным и модифицированным алгоритмом SWIPE для сигнала, ограниченного полосой телефонного канала, сведены в табл. 2. Количество больших ошибок исходного алгоритма SWIPE сопоставимо с результатами табл. 1 для указанного типа сигнала. В то же время слежение за максимумом позволило уменьшить количество больших ошибок примерно в два раза. Количество больших ошибок на женских голосах было меньше по сравнению с количеством больших ошибок на мужских голосах как для исходного, так и для модифицированного алгоритма. Данное соотношение количества больших ошибок, скорее всего, связано с меньшим влиянием ограничения снизу спектра женских голосов полосой телефонного канала.

Таблица 2

Результаты испытаний алгоритма SWIPE со слежением за максимумом (в %)

Дикторы	Ошибки Т/НТ	Ошибки НТ/Т	Средняя ошибка Т/НТ	Большие ошибки
Алгоритм SWIPE без слежения за максимумом				
Мужчины	14,16	2,75	8,46	48,96
Женщины	14,10	2,11	8,11	26,95
Общая группа	14,14	2,44	8,29	38,25
Алгоритм SWIPE со слежением за максимумом				
Мужчины	17,23	1,18	9,21	20,24
Женщины	13,20	3,40	8,40	16,88
Общая группа	15,33	2,24	8,79	18,65

Количество ошибок Т/НТ и НТ/Т существенно не изменилось. Для иллюстрации причин появления этих ошибок на рис. 7 приведен пример контура частоты ОТ, полученный при произнесении слова «Четыре», диктор VMV – женщина: нижний график – эталон, верхний график – контур частоты ОТ, выделенный модифицированным алгоритмом SWIPE со слежением за максимумом.

Так же, как и на рис. 6, значительное количество ошибок Т/НТ и НТ/Т на рис. 7 появлялось на переходах между вокализованными и невокализованными интервалами произнесения. Данное поведение алгоритма связано с интегральным характером алгоритма SWIPE, когда решение о наличии границы вокализованный/невокализованный звук принимается с точностью до ширины кадра анализа. Смещение границы в пределах кадра анализа может иметь место как в начале, так и в конце вокализации. По этой причине сравнение между алгоритмами по качеству принятия решения Т/НТ предпочтительнее проводить по среднему значению ошибки.

Из табл. 2 видно, что применение слежения за максимумом интенсивности привело к небольшому увеличению ошибки Т/НТ_{ср}.

Рассматривая рис. 7, можно сделать ещё два важных наблюдения. *Первое.* Можно отметить сглаживающие свойства алгоритма SWIPE при построении контура частоты ОТ. *Второе.* Большинство больших отклонений измеренного контура от эталонного можно отметить на интервалах повышенной изменчивости эталонного контура частоты ОТ, которые соответствуют переходам между вокализованными и невокализованными звуками. Ошибки принятия решения Т/НТ и НТ/Т оказываются сильно коррелированными с большими ошибками.



Рис. 7. Контур частоты ОТ, слово «Четыре», диктор VMV

Ограничение верхнего предела, определяющего допустимые значения частоты основного тона

Возвращаясь к рис. 5, можно предложить ещё один способ уменьшения количества ошибок, возникающих при выделении частоты ОТ алгоритмами SWIPE и SWIPE'. Рядом ошибок можно избежать, если выбрать меньшие значения верхнего предела $F_{0\max}$, определяющего границы диапазона поиска частот ОТ. На рис. 8 представлен контур ОТ, выделенный алгоритмом SWIPE для сигнала, ограниченного полосой телефонного канала, когда верхнее предельное значение частоты $F_{0\max} = 200$ Гц. Сравнивая между собой рис. 3, 4, 6 и 8, видим, что при таком выборе частоты $F_{0\max}$ большие ошибки выделения ОТ для сигнала, ограниченного полосой телефонного канала, в данном случае отсутствуют.

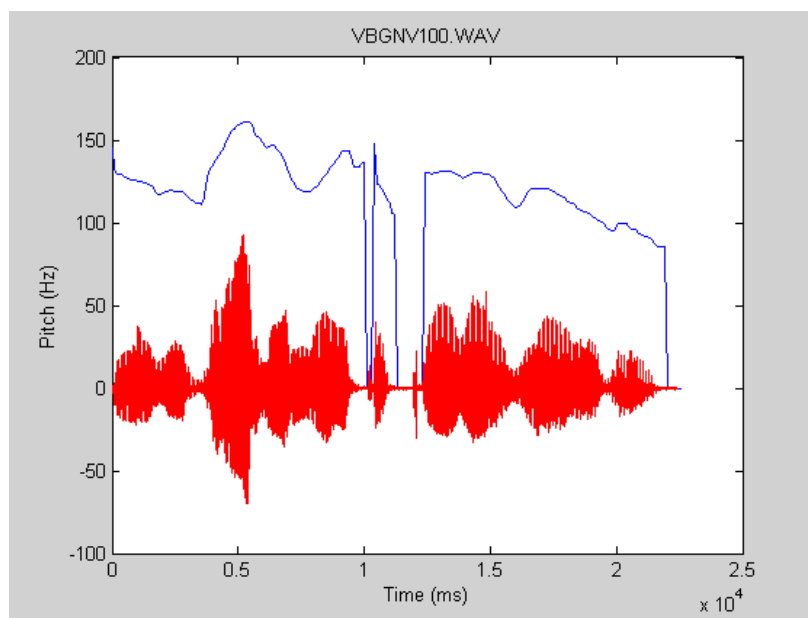


Рис. 8. Контур ОТ, полученный для сигнала, ограниченного полосой телефонного канала, методом SWIPE с верхней предельной частотой ОТ 200 Гц

Сравнивая рис. 8 и рис. 4, можно отметить, что количество ошибок Т/НТ и НТ/Т также снизилось. Причину этого явления можно увидеть из графика интенсивности ОТ, соответствующего моменту времени $t_i = 60$ мс от начала произнесения «Не видели мы такого невода», ограниченного полосой телефонного канала (рис. 9). Максимум интенсивности, превышающий порог принятия решения о наличии вокализации (в данном эксперименте $STHR = 0,16$), расположен в области частот, превышающих истинные значения периода ОТ данного произнесения. Подобное расположение максимума интенсивности приводит как к ошибочному отнесению невокализованного интервала к вокализованному, так и к ошибочному определению частоты ОТ. Снижение предельного допустимого значения F_{0max} , ограничивающего диапазон допустимых частот ОТ сверху, устраняет подобные ошибки.

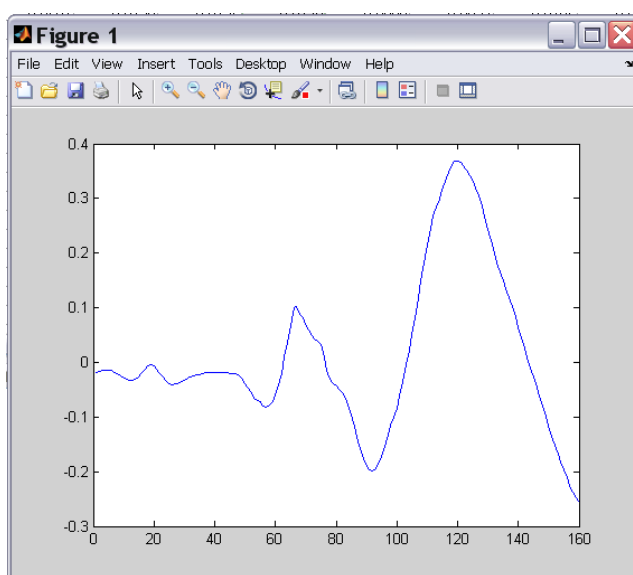


Рис. 9. Интенсивность ОТ для временного отсчёта на невокализованном интервале

При построении алгоритма автоматической оценки верхнего предела частоты ОТ [36] мы исходили из следующих предпосылок. Значения частот контура ОТ, расположенные ниже средней частоты ОТ, вычисленной по всем измеренным значениям частоты ОТ данного произнесения, скорее всего, принадлежат к истинным значениям частоты ОТ. Наличие больших скачков контура ОТ для относительно плавных интонационных изменений позволяет с определённой вероятностью предположить, что произошёл переход на гармонику ОТ. Переход на гармоники ОТ может происходить и относительно плавно. Возможны переход на субгармоники ОТ и ошибочное определение заниженных значений частоты ОТ. Такие решения случаются относительно редко.

Алгоритм поиска верхней границы значений частоты ОТ представлен на рис. 10.

После построения контура ОТ методами SWIPE или SWIPE' определялось среднее значение частоты ОТ: «СРЕДН.ОТ». Затем формировали массив отсчётов ОТ, больших нуля и меньших СРЕДН.ОТ, и вычисляли среднее значение разности между отсчетами этого массива. Исходя из значения этой разности, находили допустимую величину скачка, которая в любом варианте находилась в диапазоне 14...36 Гц. Выбор значений скачка в указанном диапазоне позволил устранить ошибочные решения относительно величины скачка в случае существенных плавных изменений траектории ОТ.

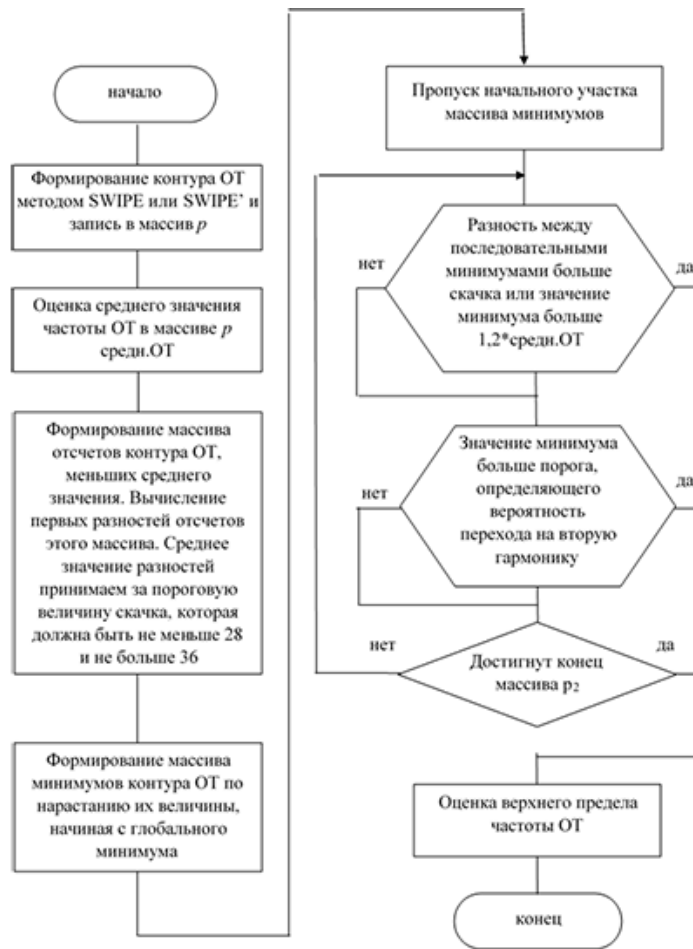


Рис. 10. Структура алгоритма оценки верхнего предела частоты ОТ

На следующем шаге алгоритма выполняли ранжирование значений выделенных минимумов и ранжированные отсчёты помещали в отдельный массив. Первые отсчёты этого массива в количестве, равном «kol_min_ispolz», отбрасывали. Общее количество минимумов ранжированного массива ограничивали значением, равным «kol_min_vsego». Эта операция позволила снизить вероятность перехода на гармоники и субгармоники. Оставшиеся отсчёты ранжированного массива проверяли на превышение допустимого значения скачка и на вероятность перехода на вторую гармонику. Если эти события имели место, или если количество тестируемых минимумов превышало kol_min_vsego, то переходили на оценку допустимого значения верхнего предела частоты ОТ F_{0max} .

Оценку допустимого значения верхнего предела алгоритм выполняет следующим образом. Из отобранных на предшествующих этапах минимумов отбрасываются начальные отсчёты в количестве, равном шестой части от оставшихся минимумов. Для пяти шестых этого массива вычисляется среднее значение, обозначенное как «min_sred». Верхняя граница диапазона определяется как

$$F_{0max} = min_sred + min_sred * procent / 100,$$

если $\min_sred < \text{СРЕДН.ОТ}$ и

$$F_{0max} = \text{СРЕДН.ОТ} + \min_sred * 0,2$$

в противном случае. Параметр «procent» по умолчанию выбран равным 30. Но в случае повышенной интонационной изменчивости контура основного тона его увеличивали в два–три раза.

На рис. 11 показано окно графического интерфейса, разработанного для организации экспериментов по выделению основного тона методами SWIPE, SWIPE' и алгоритмами с коррекцией верхнего предельного значения частоты ОТ [37]. Параметры алгоритмов SWIPE, SWIPE' и корректирующего алгоритма можно выбрать либо по умолчанию, либо вручную.

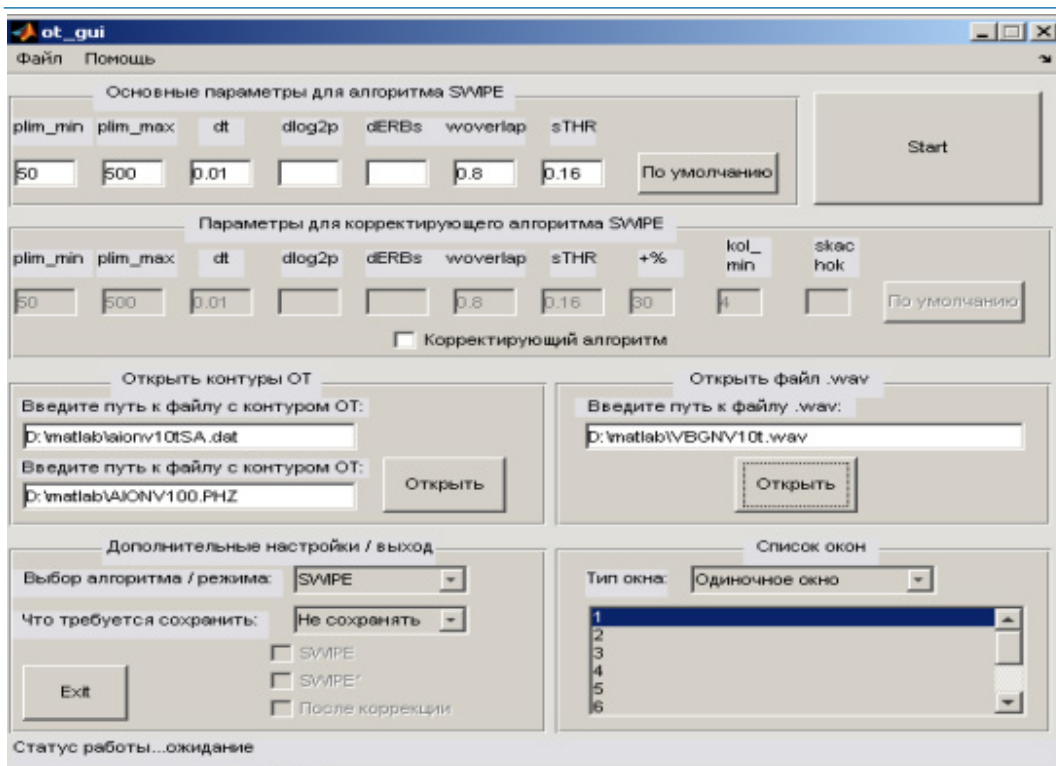


Рис. 11. Окно командного интерфейса

Корректирующий алгоритм может быть либо включён, либо отключён по команде оператора. Выбор файла данных, для которых необходимо выделить контур ОТ, выполняются на панели «Открыть файл данных» с помощью кнопки «Открыть». Панель «Дополнительные настройки / выход» позволяет выбрать алгоритм выделения ОТ: SWIPE, SWIPE' или оба алгоритма совместно. С помощью кнопок на этой же панели можно сохранить в памяти компьютера в формате «ROW» выделенные контуры ОТ. Запуск алгоритмов на выполнение осуществляется кнопкой «Start».

Результаты выделения частоты ОТ могут быть представлены либо в одиночном окне (рис. 3), либо для целей сравнения – в сдвоенном окне (рис. 4). Выбор одиночного или сдвоенного представления выполняется кнопкой «Тип окна». Одновременно могут сохраняться результаты работы до 10 экспериментов, названия которых отражены в списке окон. Записанные в памяти компьютера в формате «ROW» контуры ОТ позволяют анализировать панель «Контуры ОТ».



Результаты испытаний

Результаты испытаний алгоритмов SWIPE и SWIPE' с применением автоматической установки верхнего предельного значения диапазона частот ОТ представлены в табл. 3. Автоматическая установка F_{0max} после первого прохода («Корректирующий алгоритм») применялась как непосредственно к алгоритму SWIPE, так и к алгоритму SWIPE со слежением за максимумом.

Оба варианта алгоритма SWIPE с применением корректирующего алгоритма показали существенное снижение больших ошибок по сравнению с данными исходных вариантов алгоритма SWPE (табл. 2). Различие в количестве больших ошибок между исходным алгоритмом SWIPE и алгоритмом SWIPE со слежением за максимумом практически нивелировалось, хотя алгоритм SWIPE со слежением за максимумом и с автоматической установкой F_{0max} показал несколько лучшие результаты. Алгоритм SWIPE' (табл. 4) также улучшил свои показатели, но был хуже алгоритма SWIPE по количеству больших ошибок [38].

Таблица 3

Результаты испытаний алгоритмов SWIPE с автоматической установкой F_{0max} (в %)

Дикторы	Ошибки Т/НТ	Ошибки НТ/Т	Средняя ош. Т/НТ	Большие ошибки
Алгоритм SWIPE без слежения за максимумом				
Мужчины	2.56	10.83	6.7	1.73
Женщины	7.86	4.95	6.41	11.6
Общая группа	5.14	7.97	6.56	6.5
Алгоритм SWIPE со слежением за максимумом				
Мужчины	3.18	9.89	6.54	1.97
Женщины	8.16	5.97	7.06	9.41
Общая группа	5.94	8.04	6.99	5.39

Таблица 4

Результаты испытаний алгоритма SWIPE' с установкой F_{0max} (в %)

Дикторы	Ошибки Т/НТ	Ошибки НТ/Т	Средняя ош. Т/НТ	Большие ошибки
Общая группа	4.32	8.99	6,65	13,70

Все алгоритмы показали различное количество ошибок на мужских и женских голосах. Везде количество больших ошибок на женских голосах было больше по сравнению с мужскими голосами.

На женские голоса ограничение сигнала полосой телефонного канала влияет двояким образом. С одной стороны, минимальные значения частоты ОТ, которые для женских голосов лежат в диапазоне 200...350 Гц [16], [17], менее подвержены влиянию ограничения спектра сигнала снизу частотой 300 Гц. Предложенные выше алгоритмы по снижению количества больших ошибок, связанные с компенсацией влияния ограничения спектра сигнала снизу, меньше сказались на качестве обработки женских голосов.

С другой стороны, телефонный канал ограничивает сверху диапазон частот спектра частотой 3400 Гц. Число гармоник ОТ женских голосов, при высоких значениях частоты ОТ, будет также ограниченным. Это приводит к снижению надежности выделения ОТ алгоритмом SWIPE на женских голосах.

Как следует из сравнения данных табл. 2 и табл. 3, ограничение диапазона частот ОТ пониженным значением F_{0max} практически не сказалось на количестве ошибок Т/НТ и НТ/Т. Более того, использование алгоритма со слежением за максимумом совместно с автоматическим выбором значения F_{0max} даже несколько увеличило количество этих ошибок при снижении количества больших ошибок (табл. 3). Пример контура частоты ОТ для женского голоса, полученного алгоритмом SWIPE со слежением за максимумом и с автоматическим выбором частоты F_{0max} показан на рис. 12. Контур частоты ОТ выделялся для произнесения, состоявшего из десяти изолированных цифр от «один» до «десять». Можно отметить случаи «затягивания» ошибочных значений, принятых алгоритмом слежения за максимумами.

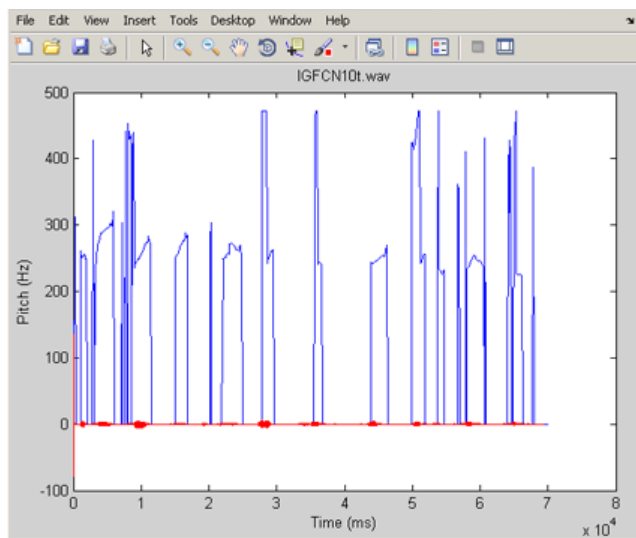


Рис. 12. Пример работы алгоритма SWIPE с автоматической установкой верхней предельной частоты основного тона F_{0max} . Женский голос

Заключение

Алгоритм SWIPE и его разновидность SWIPE' относятся к интегральным методам выделения ОТ. Они обеспечивают высокую надёжность выделения ОТ чистого и зашумленного речевого сигнала вплоть до соотношений сигнал/шум, равных 5 дБ. Имеющие место большие ошибки и ошибки Т/НТ и НТ/Т в большинстве случаев



легко диагностируются по виду контура частоты ОТ. Большая вычислительная сложность и соответствующее ей большое время вычислений не являются существенными в тех применениях, которые не требуют вычислений в реальном масштабе времени. Однако алгоритмы SWIPE и SWIPE' чувствительны к искажениям спектра сигнала, обусловленным, например, влиянием аппаратуры, предназначенной для обработки и передачи сигнала. Увеличение количества ошибок, вызванных искажениями спектра, в ряде случаев может быть скомпенсировано подбором параметров алгоритма.

Литература

1. *Кодзасов С.В., Кривнова О.Ф.* Общая фонетика: Учебник. — М.: Рос. гос. ун-т, 2001. — 592 с.
2. *Чистович Л.А., Венцов А.В., Гранстрем М.П. и др.* Физиология речи. Восприятие речи человеком. — Л.: Наука, 1976. — 388 с.
3. *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.
4. *Лобанов Б.М.* Исследование и разработка методов автоматического синтеза речи по фонемному тексту: автореф. дис. ... доктора. техн. наук. — Рига, 1984. — 50 с.
5. *Светозаров Н.Д.* Интонационная система русского языка. — Л.: ЛГУ, 1982. — 176 с.
6. *Рамишвили Г.С.* Речевой сигнал и индивидуальность голоса. — Тбилиси: Изд-во «Мецниереба», 1976. — С. 184.
7. *Кузнецов П.Г., Гитлин В.Г.* Идентификация голосов по средней частоте основного тона // Применение вычислительной техники в машиностроении. — Ижевск, 1977. — С. 68-74.
8. *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.
9. *Галунов В.И.* О возможности определения эмоционального состояния говорящего по речи // Речевые технологии. — 2008. — №1. — С. 60-66.
10. *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. — Vol. 78. — № 1 (Part 2). — P. 40-48.
11. *Златоустова Л.В.* Фонетические единицы русской речи. — М.: МГУ, 1981. — 108 с.
12. *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.
13. *Kolesnikov B.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.

14. *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.
15. *Ryalls J.H. and Lieberman P.* Fundamental frequency and vowel perception // J. Acoust. Soc. Am. — 1982. — Vol. 72. — № 5. — P. 1631–1634.
16. *Гитлин В.Б.* Основной тон речевого сигнала / Деп. в ВИНТИ, 1998. — № 1206-B98. — 739 с.
17. *Сапожков М.А.* Речевой сигнал в кибернетике и связи. — М.: Государственное издательство литературы по вопросам связи и радио, 1963. — 450с.
18. *Сапожков М.А., Михайлов В.Г.* Вокодерная связь. — М.: Радио и связь, 1983. — 248 с.
19. *Добровольская Н.Ф., Абилова М.А.* Стандарт сотовой связи GSM. Электронный ресурс. Режим доступа: <http://www.radioscanner.ru>.
20. *Михайлов В.Г.* Из истории исследований преобразования речи (часть 1) // Речевые технологии. — 2008. — № 1. — С.93–113.
21. *Михайлов В.Г.* Из истории исследований преобразования речи (часть 2) // Речевые технологии. — 2008. — №2. — С.81–96.
22. Система редактирования, анализа и шумочистки речевых сигналов SIS 6.x. Руководство пользователя / Центр речевых технологий. Санкт-Петербург.
23. *Лячканов С.Е.* Криминалистический учет лиц по фонограммам их речи // Речевые технологии. — 2008. — № 4. — С.111–118.
24. *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.
25. *Баронин С.П.* Автокорреляционный метод выделения основного тона речи // Речевые технологии. — 2008. — №2. — С.3–12.
26. *Samacho A., Harris J.G.* A sawtooth waveform inspired pitch estimator for speech and music // Journal of the Acoustical Society of America. — 2008, vol. 124. — P. 1638–1652.
27. *Rabiner L., Juang B-H.* Fundamental of Speech Recognitions. Prentice-Hall International Inc. — 1993. — 541 p.
28. *Гитлин В.Б., Лузин Д.А.* Совместный алгоритм выделения основного тона речи методами GS и автокорреляционной функции речи // Речевые технологии. — 2008. — № 3. — С. 39–42.
29. *Hong J.O. and Wolf P.J.* Model-based estimation of instantaneous pitch in noisy speech. Proceedings of INTERSPEECH. — 2009.
30. *Resch B., Nillson M., Ekman A. and Klejin W.B.* Estimation of the instantaneous Pitch of Speech // IEEE Trans on Audio, Speech and Lang. Process. — 2007.— Vol. 15. — № 3. — P. 813–822.
31. *Kobayashi T., Arifanto D., Masuko T.* Fundamental frequency estimation based on instantaneous frequency amplitude spectrum // Proc of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — 2002. — P. 329–332.
32. *Kawaara H., Masuda-Katsuse I. and de Cheveigné F.* Restructuring speech representation using a pitch-adaptive time-frequency smoothing of a repetitive structure in sounds // Speech Communication. — 1999. — № 27. — P. 187–207.
33. *Гитлин В.Б., Лузин Д.А.* Экспериментальная оценка точности выделения частоты основного тона обобщенным методом GS // Сборник трудов XX сессии Российского акустического общества. — М.: ГЕОС, 2008. — Т. 3. — С. 54–58.
34. *Гитлин В.Б., Девярых А.И.* Выбор метода оценки высоты музыкальных звуков // Сборник трудов III научно-технической конференции «Приборостроение в XXI. Интеграция науки, образования и производства» (Ижевск, 14-15 апреля 2006 г.).— Изд-во ИжГТУ, 2007. — С. 393–397.



35. *Ведерников А.А.* Разработка подсистемы оценки точности выделения основного тона и исследования на ее основе системы выделения основного тона и системы принятия решения «Тон/НЕ ТОН». Пояснительная записка к дипломной работе по специальности «Инженер-системотехник». — Ижевск: ИжГТУ. — 2007. — 154 с.
36. *Вашурин Д.Ю., Гитлин В.Б.* Выделение основного тона методом SWIPE из сигнала, прошедшего телефонный канал // Интеллектуальные системы в производстве. — 2012. — №2(22). — С. 123–126.
37. *Вашурин Д.Ю., Гитлин В.Б., Лузин Д.А.* Графический интерфейс алгоритма SWIPE // Сборник трудов региональной научной очно-заочной конференции «Информационные технологии в науке, промышленности и образовании» (Ижевск, 18 мая 2013 г.). / Науч. ред. В.А. Куликов. — Ижевск: Изд-во ИжГТУ, 2013. — С. 126–129.
38. *Вашурин Д.Ю., Гитлин В. Б., Лузин Д.А.* Сравнение алгоритмов выделения основного тона GS2 и SWIPE' // Сборник трудов региональной научной очно-заочной конференции «Информационные технологии в науке, промышленности и образовании» / Науч. ред. В.А. Куликов. — Ижевск: Изд-во ИжГТУ, 2012. — С. 30–34.

Сведения об авторах:

Гитлин Валерий Борисович,

доктор технических наук, профессор кафедры «Вычислительная техника» факультета «Информатика и вычислительная техника» ФГБОУ ВПО «Ижевский государственный технический университет имени М.Т. Калашникова». Область научных интересов: цифровая обработка сигналов, распознавание речи, информационно-измерительные и управляющие системы.

Вашурин Дмитрий Юрьевич,

магистрант ФГБОУ ВПО «Ижевский государственный технический университет имени М.Т. Калашникова». Область научных интересов: цифровая обработка сигналов, распознавание речи.