

The algorithm for constructing the scalar and vector potentials based on decomposition of the vector field of the dynamical system on the potential and solenoidal components by forming homotopy operator for the differential form corresponding to the vector field of dynamic system is considered in the paper. The invariants, constructed on these potentials, are used in pattern recognition of vector fields.

Keywords: vector field of dynamical system, potential of vector field, invariant of vector field, decomposition of vector field, homotopy operator.

Получено 28.04.2014

УДК 621.391

В. Б. Гитлин, доктор технических наук, профессор, Ижевский государственный технический университет имени М. Т. Калашникова

ПОВЫШЕНИЕ НАДЕЖНОСТИ ВЫДЕЛЕНИЯ ОСНОВНОГО ТОНА МЕТОДОМ SWIPE ИЗ СИГНАЛА, ПРОШЕДШЕГО ТЕЛЕФОННЫЙ КАНАЛ

Рассмотрена работа алгоритмов SWIPE и SWIPE' при выделении основного тона речевого сигнала, ограниченного полосой телефонного канала. Предложен метод, позволяющий повысить надежность выделения основного тона алгоритмами SWIPE и SWIPE'.

Ключевые слова: основной тон, гармоники, скалярное произведение, пилообразный сигнал.

Алгоритм выделения основного тона речи SWIPE (Sawtooth Inspired Pitch Estimator) [1] продолжает вызывать интерес исследователей, о чем свидетельствуют, в частности, ссылки в литературе на данный алгоритм [2]. Алгоритм SWIPE и его разновидность SWIPE' [1] обеспечивает надежное выделение частоты основного тона (ЧОТ) речи на чистом сигнале и на сигнале с низким соотношением сигнал/шум. Но на сигнале, ограниченном полосой телефонного канала, надежность выделения частоты ОТ резко падает [3].

Настоящую работу можно рассматривать как продолжение работы [3]. Рассматриваются причины появления больших ошибок выделения частоты ОТ при работе с сигналом, ограниченным полосой телефонного канала. Выполнен анализ качества работы алгоритмов SWIPE и SWIPE' отдельно для мужских и женских голосов. Предложен метод снижения количества больших ошибок методами SWIPE и SWIPE' при выделении ЧОТ сигнала, ограниченного полосой телефонного канала. Метод использует процедуру выбора частоты ОТ исходя из критерия ближайшего расстояния до предыдущего выбора этой частоты.

В отличие от автокорреляционных методов [4] в алгоритмах SWIPE и SWIPE' [1] частота ОТ оценивается по максимуму корреляции речевого сигнала с эталонным образцом, имеющим спектр, близкий к спектру анализируемого звука. В алгоритмах SWIPE и SWIPE' параметры адаптируемого эталонного образца изменяются таким образом, чтобы достичь максимально возможного сходства с анализируемым речевым сигналом. После достижения максимального сходства частота основного тона адаптируемого эталона принимается за частоту основного тона сигнала.

В качестве адаптируемого эталона в алгоритмах SWIPE и SWIPE' выбран пилообразный сигнал, что

и определило название метода (Sawtooth Inspired Pitch Estimator) [1]. Сравнение анализируемого сигнала с эталонным образцом выполняют при помощи интегрального преобразования путем вычисления скалярного произведения между спектром сигнала и ядром преобразования в виде лепестков косинуса [1]. Значение вычисленного скалярного произведения рассматривают как меру интенсивности кандидатов в частоты ОТ в анализируемом звуке. 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 от алгоритма SWIPE в расчете скалярного произведения принимают участие только гармоники с простыми (не кратными) номерами, что дополнительно уменьшает вероятность принятия решения о переходе выделенной частоты ОТ на субгармоники.

Методы SWIPE и SWIPE' относятся к интегральным [5]. Они определяют значение частоты ОТ на кадре анализа длиной в четыре периода ОТ. Оба метода требуют значительного количества вычислений, предварительного выбора пределов измерения частоты ОТ и установления порога STHR принятия решения ТОН/НЕ ТОН (Т/НТ). Ядро интегрального преобразования алгоритмов SWIPE и SWIPE' соответствует среднему спектру речевого сигнала примерно 6 дБ/октава в сторону высоких частот [1]. Однако для конкретных произнесений огибающая спектра, присущая конкретному звуку, может отличаться от указанной средней зависимости от частоты. Например, фонема /и/ имеет сильный спад огибающей

спектра в полосе 500...1500 Гц и подъем в области 1700...2200 Гц [6, 7]. Пример среднего спектра слога /не/ (фраза «Не видали мы такого невода», диктор VBG) показан на рис. 1. На этом рисунке: толстая линия – исходный спектр слога /не/, тонкая линия – спектр сигнала, ограниченного полосой телефонного канала. По оси абсцисс на рис. 1 отложена частота в Гц, по оси ординат – логарифм амплитуды. Спектр получен с помощью программы SIS.6x [8].

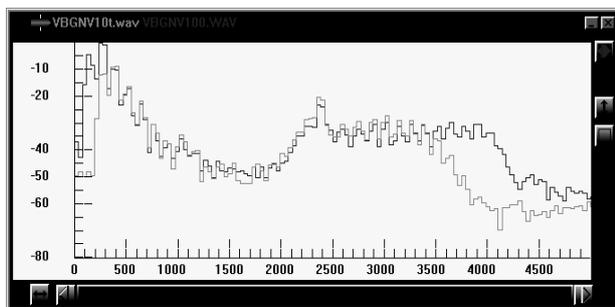


Рис. 1. Спектр слога /не/, диктор VBG

В случае ограничения полосы сигнала полосой телефонного канала (300...3400 Гц) низкочастотные гармоники спектра оказываются подавленными. Отсутствие низкочастотных гармоник наряду с малыми значениями огибающей спектра в диапазоне 500...1500 Гц увеличивает долю высокочастотных компонент спектра в расчете интенсивности вокализации. Повышение вклада высокочастотных компонент приводит к возрастанию вероятности перехода на гармоники ОТ. Необходимо также учесть влияние аппаратуры обработки речи.

В алгоритмах SWIPE и SWIPE' еще одной причиной сбоя на гармоники ОТ может служить изменчивость гармонической структуры речевого сигнала. Гармоническая структура спектра речи сохраняется лишь до 2...2,5 кГц, причем с повышением частоты форма спектральных составляющих гармоник размывается, выраженность максимума гармоники в спектре становится менее четкой [5–7]. «Размытость» положения высокочастотных гармоник является характерным свойством речи [2].

На рис. 2 приведен пример осциллограммы и контура частоты ОТ для фразы «Не видали мы такого невода». Контур получен алгоритмом SWIPE для сигнала без шума, диктор VBG. За исключением интервалов наличия и отсутствия вокализации этот контур достаточно близко соответствует эталонному контуру ОТ. Частота ОТ на рис. 2 изменяется в диапазоне от 80 до 160 Гц.

На рис. 3 показан контур ОТ, полученный методом SWIPE для сигнала, ограниченного полосой телефонного канала. Фраза и параметры алгоритма те же, что и на рис. 2. Видны многочисленные сбои на гармоники ОТ, значение частот которых приближается к предельной частоте $F_{0max} = 500$ Гц, соответствующей верхней границе диапазона поиска частот ОТ.

На рис. 4 проиллюстрированы сечения значений интенсивности ОТ как функция номера кандидата

в частоты ОТ для некоторого момента времени t_i . 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.

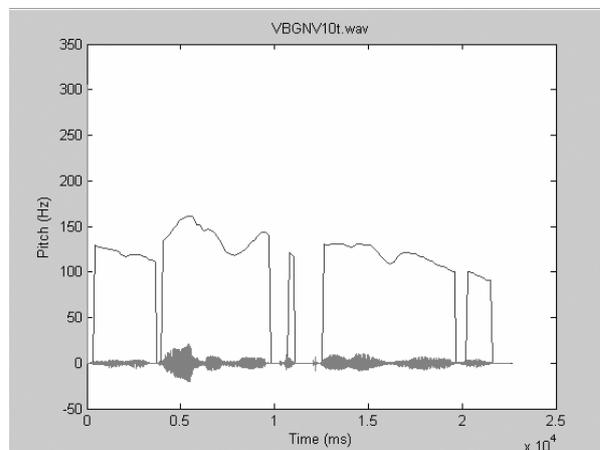


Рис. 2. Контур основного тона, полученный для чистого сигнала алгоритмом SWIPE

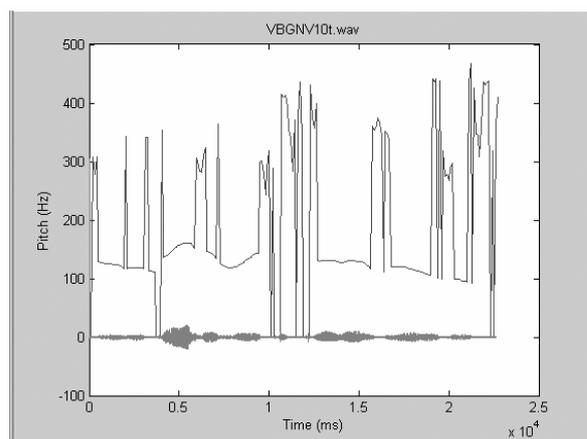


Рис. 3. Контур основного тона, полученный для сигнала, ограниченного полосой телефонного канала, методом SWIPE

Из анализа рис. 4 можно сделать следующие выводы. Во-первых, высокочастотный максимум имеет более плавный характер и менее выражен по сравнению с низкочастотным максимумом, определяемым истинным значением ОТ. Указанное свойство является следствием размывания спектра гармоник сигнала с повышением частоты [2]. Ошибочный переход на более высокий по частоте максимум вызван указанным выше подчеркиванием высокочастотной части спектра в сигнале, ограниченном полосой телефонного канала.

Во-вторых, информация об истинном значении частоты ОТ с большой вероятностью сохраняется в графике интенсивности, соответствующем сбойной ситуации (рис. 4, нижняя панель). Учитывая эти результаты, в алгоритмы SWIPE и SWIPE' было решено

внести следующее изменение. Был добавлен поиск второго локального максимума на кривой интенсивности ОТ, соответствующей t_i -му моменту анализа. В качестве второго дополнительного максимума выбирали максимум, расположенный непосредственно ниже основного максимума. Из найденных на t_{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.

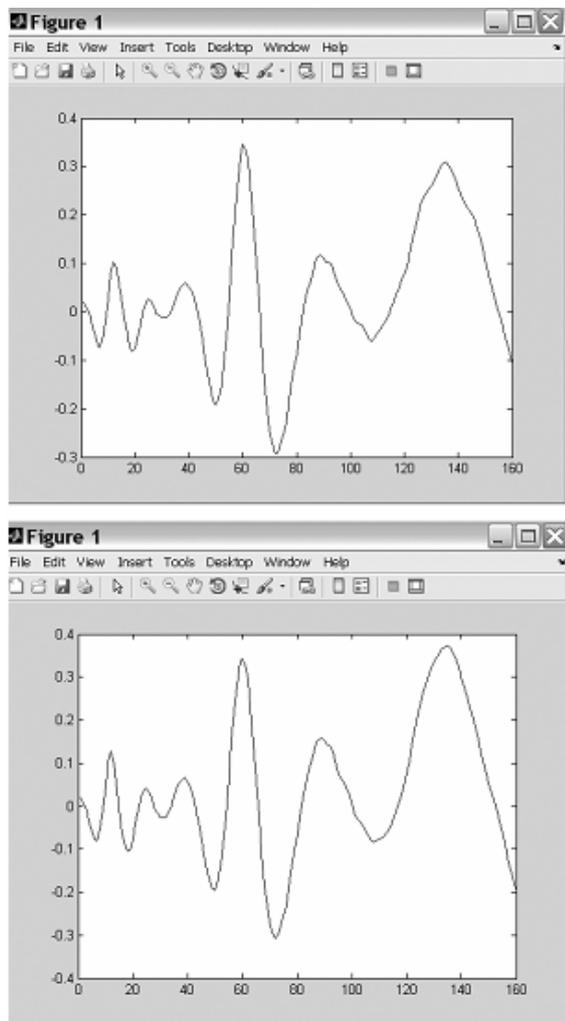


Рис. 4. Интенсивность кандидатов в ОТ

На рис. 5 представлен пример работы модифицированного алгоритма SWIPE при выделении частоты ОТ сигнала, ограниченного полосой телефонного канала. Сравнивая эти результаты с результатами, показанным на рис. 3, видим, что количество больших ошибок существенно уменьшилось. Ошибки в основном возникали на невокализованных интервалах, неверно отнесенных алгоритмом к вокализованным.

В испытаниях участвовало 20 дикторов мужчин и 17 дикторов женщин. Результаты выделения частоты ОТ исходным и модифицированным алгоритмом SWIPE для сигнала, ограниченного полосой телефонного канала, сведены в табл. 1. 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. Данное соотношение количества больших ошибок связано с большими значениями частот ОТ женских голосов и, соответственно, с меньшим влиянием ограничения спектра сигнала снизу частотой 300 Гц.

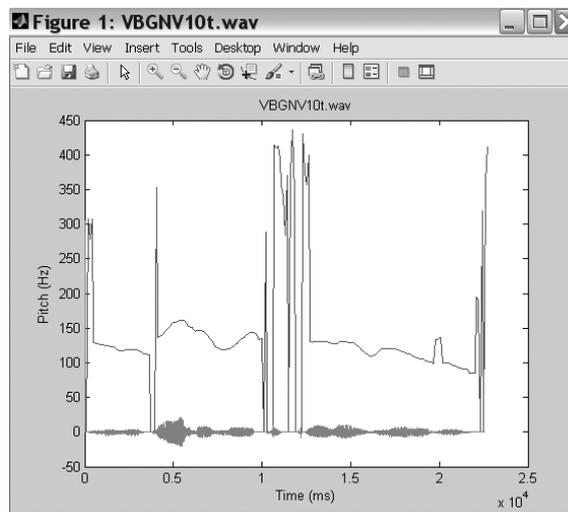


Рис. 5. Выделение ОТ модифицированным алгоритмом SWIPE

Применение алгоритма автоматического выбора частоты F_{0max} , описание которого дано в работе [3], позволило дополнительно снизить количество больших ошибок (табл. 2). 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 SWIPE (табл. 1). Различия 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. 3) также улучшил свои показатели, но был хуже алгоритма SWIPE по количеству больших ошибок [10].

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. Эти ошибки, в основном, имеют место на переходах от вокализованных интервалов к невокализованным, и наоборот. Решение Т/НТ принимается на кадре анализа, длительностью четыре периода ОТ. Длительности периодов ОТ женских голосов в среднем больше длительностей периодов ОТ, что, возможно, и служит причиной большего количества ошибок Т/НТ и НТ/Т на этих голосах.

Таблица 1. Результаты испытаний алгоритма SWIPE со слежением за максимумом, %

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

Таблица 2. Результаты испытаний алгоритмов SWIPE с автоматической установкой F_{0max} , %

Дикторы	Ошибки Т/НТ	Ошибки НТ/Т	Средняя ошибка 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 3. Результаты испытаний алгоритма SWIPE' с установкой F_{0max} , %

Дикторы	Ошибки Т/НТ	Ошибки НТ/Т	Средняя ошибка Т/НТ	Большие ошибки
Общая группа	4,32	8,99	6,65	13,70

Заключение

Алгоритм SWIPE и его разновидность SWIPE' относятся к интегральным методам выделения ОТ. Они обеспечивают высокую надежность выделения основного тона чистого и зашумленного речевого сигнала вплоть до соотношений сигнал/шум, равных 5 дБ. Большая вычислительная сложность и соответствующее ей большое время вычислений не являются существенными недостатками в тех применениях, которые не требуют вычислений в реальном масштабе времени. Решение Т/НТ принимается на кадре анализа длительностью в четыре периода ОТ, что приводит к большому количеству ошибок на переходах от невокализованных интервалов к вокализованным. Алгоритмы SWIPE и SWIPE' чувствительны к искажениям спектра аппаратурой обработки и передачи сигнала. Тем не менее информация об истинных значениях частоты ОТ сохраняется в вычисленных значениях интенсивности ОТ. Использование этой информации позволяет повысить надежность выделения ОТ.

Библиографические ссылки

1. Camacho 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.
2. Азаров И. С., Вашкевич М. И., Петровский А. А. Алгоритм оценки мгновенной частоты основного тона речевого сигнала // Цифровая обработка сигналов. – 2012. – № 4. – 49–57.
3. Ващури Д. Ю., Гитлин В. Б. Выделение основного тона методом SWIPE из сигнала, прошедшего телефонный канал // Интеллектуальные системы в производстве. – 2013. – № 2(22). – С. 123–126.
4. Баронин С. П. Автокорреляционный метод выделения основного тона речи // Речевые технологии. – 2008. – № 2. – С. 3–12.
5. Гитлин В. Б. Основной тон речевого сигнала / деп. в ВИНТИ, 1998. – № 1206-В98. – 739 с.
6. Кодзасов С. В., Кривнова О. Ф. Общая фонетика : учебник. – М. : Рос. гос. ун-т, 2001. – 592 с.
7. Сапожков М. А. Речевой сигнал в кибернетике и связи. – М. : Связьиздат, 1963.
8. Система редактирования, анализа и шумоочистки речевых сигналов SIS 6.x. Руководство пользователя. – Санкт-Петербург : Центр речевых технологий.
9. Ващури Д. Ю., Гитлин В. Б., Лузин Д. А. Сравнение алгоритмов выделения основного тона GS2 и SWIPE' / научн. ред. В. А. Куликов // Информационные технологии в промышленности и образовании : сб. науч. тр. регион. научн.-техн. очно-заочн. конф. – Ижевск : Изд-во ИжГТУ, 2012. – С. 30–34.

V. B. Gitlin, DSc in Engineering, Professor, Kalashnikov Izhevsk State Technical University

Improving Reliability of Pitch Extraction by SWIPE Algorithm for Telephone Bandwidth Limited Signal

Functioning of SWIPE and SWIPE' algorithms for telephone bandwidth limited signal is examined. The method is proposed to improve the reliability of pitch extraction by SWIPE and SWIPE' algorithms.

Keywords: pitch, harmonics, inner product, sawtooth signal.

Получено 06.05.2014