



Выделение импульсов основного тона по речевому сигналу

Собакин А.Н.

ГОУ ВПО «Московский государственный лингвистический университет»
Россия, 119034, Москва, ул. Остоженка, 38.
Тел. 8 (495) 637-56-97. E-mail: ansobakin@yandex.ru

Речевые колебания преобразуются в импульсную последовательность (импульсное преобразование речи ИПС), синхронную с периодом колебаний голосового источника. По полученной последовательности определяются интервалы, соответствующие сомкнутым голосовым связкам. На этих интервалах определяются несмещённые характеристики речевого тракта и методом обратной фильтрации речевых колебаний вычисляются импульсы основного тона.

Фонационные характеристики речи составляют основу описания источника звуковых колебаний, возникающих в речевом тракте. Выделяют четыре основных типа источника речевого сигнала: голосовое (тональное) возбуждение, шумовое (турбулентное) возбуждение, смешанное возбуждение и взрыв (импульсное возбуждение).

Различие между звуками, определяемое типом источника, есть признак способа образования. Основные трудности определения типа источника и его параметров по речевому сигналу заключаются в следующем:

1. Некорректность задачи определения типа источника и его параметров из-за того, что речевой сигнал есть свёртка сигнала источника и передаточной функции речевого тракта.
2. Большой диапазон и статистическая природа параметров источника.
3. Наличие нелинейных эффектов взаимодействия между источником и речеобразующим трактом.
4. Возможность генерации звуков при одновременном участии нескольких типов возбуждения.

Шумовой источник возбуждения речевого тракта достаточно полно имитируется случайным сигналом в виде «белого шума», каждое значение которого имеет равномерное распределение по амплитуде и не зависит от предыдущих значений.

Напротив, *импульсное возбуждение* изучено слабо: нет надёжных моделей и его характеристик.

При исследовании *смешанного источника* возбуждения голосовой источник значительно превосходит энергетически шумовой источник. Поэтому исследо-

вание этого типа возбуждения можно осуществлять алгоритмами и методами, применимыми при изучении голосового источника.

Голосовой источник может быть представлен в виде пульсирующего воздушного потока. Форма импульсов воздушного потока имеет приблизительно треугольный вид или усечённую снизу синусоиду. Скорость воздушного потока примерно пропорциональна площади раскрытия голосовой щели. Площадь раскрытия голосовой щели монотонно увеличивается вплоть до максимального значения (10...20 мм²) на фазе открывания голосовых связок (рис. 1).

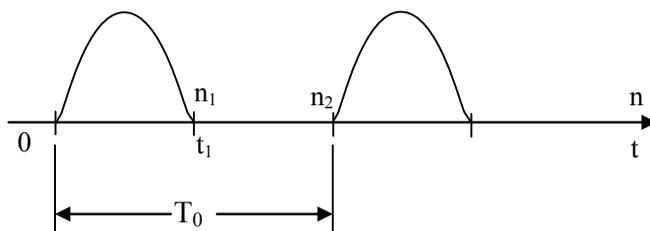


Рис. 1. Модельные импульсы основного тона

Наибольшей информационной ёмкостью из всех указанных источников обладает голосовой источник и ему в дальнейшем уделено основное внимание.

Можно выделить два основных параметра голосового источника:

1. Длительность периода колебаний голосовых связок T_0 , которая получила название *периода основного тона*. Величина, обратная периоду основного тона, $F_0 = 1/T_0$, называется частотой основного тона. Период основного тона состоит из двух интервалов: $T_0 = T_c + T_p$, где $T_c = T_0 - t_1$ — длительность интервала сомкнутых, и $T_p = t_1$ — длительность интервала разомкнутых голосовых связок.

2. Форма импульса воздушного потока глотки.

Диапазон изменения частоты основного тона очень широк и составляет примерно 60...400 Гц в разговорной речи. Частота основного тона может достигать 2300 Гц при пении. Интервал смыкания голосовых связок занимает 30...70% времени от периода основного тона и с ростом частоты отношение T_c/T_p остаётся примерно постоянным.

В научной литературе в модельных представлениях форму импульсов основного тона принято описывать положением максимума импульса, величиной (амплитудой) максимума, наклоном начального и конечного участков импульса. Однако следует отметить, что выделение импульсов основного тона по речевым колебаниям представляет собой весьма сложную задачу

В рамках линейной модели речеобразования задача определения возбуждающей функции сводится к задаче синтеза линейной системы, обратной воздействию речевого тракта. Сложность расчёта параметров обратной линейной системы усугубляется недостаточностью информации о сигнале возбуждения речевого тракта: известен и доступен для измерения только речевой сигнал (реакция прямой линейной системы). Многочисленные прогнозирующие методы расчёта обратной линейной системы по речевому сигналу не дают устойчивых и надёжных результатов. Одна из причин неудовлетворительности получаемых результатов заключается в том, что параметры обратного фильтра являются смещёнными оценками относительно истинных значений. Это приводит к искажению импульсов основного тона и усложнению формы звуковых колебаний. Характеристики голосового источника маскируются сложными колебательными процессами.

Одно из преобразований речевого сигнала [1] позволяет перейти к импульсной последовательности, синхронной с периодом $(0, T_0)$ колебаний голосовых связок.

Сложная форма речевых колебаний существенно упрощается предлагаемым нелинейным преобразованием. Получающаяся в результате функция имеет ярко выраженный импульсный характер. При этом импульсы соответствуют моментам размыкания голосовых связок, а моменты их смыкания соответствуют малым амплитудам, практически равным нулевым значениям.

Предлагаемое нелинейное преобразование речевого сигнала обладает следующими свойствами:

- на интервалах смыкания (t_0, T_0) голосовых связок речевой сигнал полностью компенсируется и становится близким к нулю;
- на интервалах размыкания ($0, t_0$) речевые колебания преобразуются в положительный импульс;
- форма получающегося импульса отражает форму импульса основного тона;
- параметры речевого тракта при этом не оцениваются и в вычислительном процессе не используются.

Все расчёты в предлагаемом методе основываются только на речевом сигнале и его основных свойствах:

- на интервалах смыкания голосовых связок функция возбуждения близка к нулю, и линейная система речевого тракта находится в режиме свободных колебаний;
- на интервалах размыкания, напротив, имеем вынужденные колебания линейной системы.

Предлагаемое нелинейное преобразование является компенсатором свободных колебаний линейной системы конечного порядка и позволяет производить микросегментацию речи на интервалы смыкания и размыкания голосовых связок.

Метод основан на уравнениях авторегрессии, моделирующих процесс речеобразования:

$$x_n + a_1 x_{n-1} + \dots + a_p x_{n-p} = h_n. \quad (1)$$

На интервалах смыкания возбуждающая функция h_n равна нулю, и неоднородное уравнение (1) преобразуется в однородное векторное следующего вида:

$$\bar{x}_0 + a_1 \cdot \bar{x}_1 + \dots + a_p \cdot \bar{x}_p = \bar{0}, \quad (2)$$

где $\bar{x}_k = \{x_k, x_{k+1}, \dots, x_{k+N}\}$, $k = 0, 1, \dots, p$ — векторы длины $(N+1)$, образованные из отсчётов $\{x_j\}$ речевого сигнала на интервале смыкания (t_1, T_0) голосовых связок.

Уравнение (2) соответствует режиму свободных колебаний в речевом тракте, и указывает на линейную зависимость системы векторов, образующих левую часть равенства (2) на интервале (t_1, T_0).

На интервалах размыкания голосовых связок уравнение линейной зависимости не выполняется, и однородное векторное уравнение преобразуется в неоднородное.

Появляется возможность сегментации периода основного тона на интервалы сомкнутых и разомкнутых голосовых связок по признаку линейной зависимости или линейной независимости системы векторов.

В теории линейных векторных пространств [2] это разграничение может осуществляться вычислением определителя Грамма автокорреляционной матрицы $R_{p+1}(n)$:

$$M(n) = |R_{p+1}(n)|, \quad (3)$$

где $p+1$ — порядок определителя; n — текущее дискретное время.

Тот факт, что указанный определитель является мерой зависимости или независимости системы векторов, легко обосновывается соответствующим преобразованием равенства (2) в однородную систему линейных уравнений, имеющую нетривиальное решение.

Для получения указанной системы векторное равенство (2) скалярно умножить на векторы $\bar{x}_k = \{x_k, x_{k+1}, \dots, x_{k+N}\}, k = 0, 1, \dots, p$ соответственно. Как известно [2] однородная система уравнений имеет ненулевое решение только в том случае, когда определитель при неизвестных будет равен нулю. На интервалах, соответствующих разомкнутым голосовым связкам, система линейных уравнений будет неоднородной, имеющей единственное решение. В этом случае главный определитель системы будет отличен от нуля. Определителем при неизвестных в рассмотренных случаях будет определитель Грамма (3), и, следовательно, этот определитель может служить мерой линейной зависимости и независимости системы векторов.

Значение определителя $M(n)$ вычисляется последовательно с шагом дискретизации речевого сигнала, при этом на каждом шаге n происходит обновление координат векторов $\bar{x}_k = \{x_k, x_{k+1}, \dots, x_{k+N}\}, k = 0, 1, \dots, p$.

Определитель автокорреляционной матрицы позволяет определять интервалы смыкания голосовых связок, на которых функцию возбуждения речевого тракта можно считать известной и равной нулю. На этом интервале формируется система линейных уравнений на основе уравнения (1) следующего вида:

$$R_{p \times p}(n) \bar{a} = \bar{b}, \quad (4)$$

где ...?... автокорреляционная матрица порядка $p \times p$, а координаты b_k вектора ...?... равны отрицательным значениям скалярных произведений $-(\bar{x}_0, \bar{x}_k)$ векторов \bar{x}_0 и \bar{x}_k , $k = 1, 2, \dots, p$ соответственно.

Система уравнений (4) получается последовательным умножением равенства (1) на векторы \bar{x}_k , $k = 1, 2, \dots, p$ и переносом первых членов в правую часть.

Решение $\bar{a} = \{a_1, a_2, \dots, a_p\}$ системы (4) определяет коэффициенты линейного предсказания, которые являются несмещёнными оценками резонансных характеристик речевого тракта. Они вместе с единичным первым коэффициентом составляют параметры обратного фильтра. Подставляя полученные значения в уравнение (1), вычислим функцию возбуждения h_n речевого тракта, образованную работой голосовых связок.

Полученная импульсная последовательность позволяет исследовать характеристики импульсов основного тона по речевому сигналу и определять несмещённые параметры речевого тракта.

Для повышения надёжности предложенного метода можно производить усреднение оценок параметров обратного фильтра, полученных на нескольких смежных интервалах сомкнутых голосовых связок, и выделять импульсы основного тона по речевому сигналу на базе усредненных оценок.

Эксперименты по определению интервалов T_c , соответствующих сомкнутым голосовым связкам, проводились для стационарных гласных «а», «о», «у», «и», «э», «ы» диктора I (муж., частота основного тона порядка $F_0 = 150$ Гц).

В процессе проведенного эксперимента результаты представлялись в графическом окне, каждое из которых состояло из двух частей: верхней и нижней. В нижней части вы-

водилась осциллограмма (т.е. значения $x(n)$ отсчетов речевого сигнала), а в верхней части — график изменения значений функции $M(n)$ порядка $p=9$ (рис. 2).

Как видно из приводимых графиков, интервалы смыкания голосовых связок (отмечены подчёркиванием на оси абсцисс) определялись достаточно стабильно для всех шести гласных.

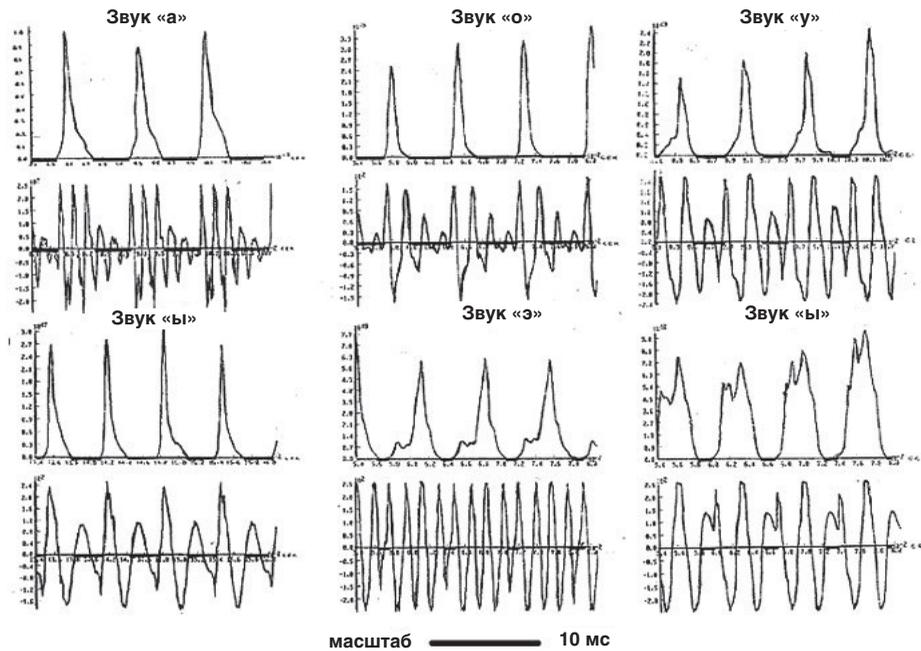


Рис. 2. Определение интервалов смыкания для гласных звуков речи

Далее в соответствии с приведённым алгоритмом на этих интервалах вычислялись несмещённые оценки обратного фильтра и методом обратной фильтрации (1) определялась функция h_n возбуждения речевого тракта.

Приведём эксперименты для стационарных гласных «и», «э» для диктора I (муж., $F_0 = 130$ Гц) и диктора II (жен., $F_0 = 300$ Гц). Полученные импульсы основного тона изображены на рисунке 3.

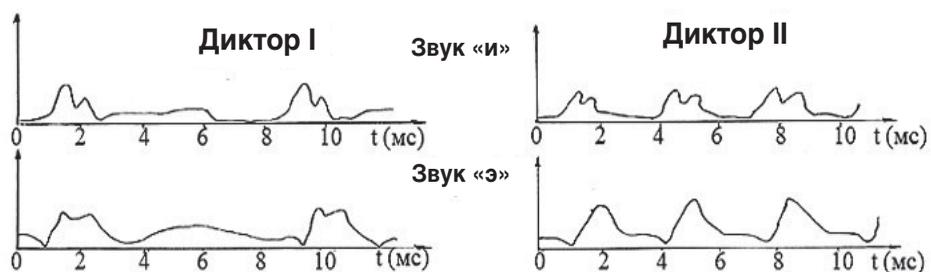


Рис. 3. Импульсы основного тона для гласных «и», «э» дикторов I и II

Собакин А.Н.

Выделение импульсов основного тона по речевому сигналу

Как видно из приведённых графиков форма импульсов основного тона существенно сложнее формы модельного импульса и зависит от произносимого звука речи и диктора.

Метод позволяет, таким образом, исследовать импульсы основного тона и их форму по речевым колебаниям, и ответить на ряд вопросов речеобразования:

- какова зависимость колебаний голосового источника от произносимого звука речи;
- какова доля индивидуальных характеристик диктора в форме импульсов основного тона?

ЛИТЕРАТУРА

1. Собакин А.Н. Артикуляционные параметры речи и математические методы их исследования // Монография. Вестник МГЛУ. Серия «Лингвистика». Вып. 517. М., 2005. 220 с.
2. Гантмахер Ф.Р. Теория матриц. М.: «Наука», 1967. 575 с.