Исследование голосового источника речи

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

Преобразование речи в импульсную последовательность, синхронную с колебаниями голосовых связок, позволяет исследовать форму полученных импульсов методами математической статистики. Для этого предлагается производить нормировку полученных импульсов по их центрам и осуществлять сложение нормированных импульсов. Эти процедуры позволяют получить статистически значимый «образ» полученной последовательности импульсов в виде нечёткого множества.

Слуховое восприятие речи в процессе эволюции достаточно хорошо согласовано по своим характеристикам со структурой речевого сигнала и, в частности, с восприятием источника звуковых колебаний. Слуховая система человека определяет тип источника (голосовой, шумовой, смешанный и импульсный) и осуществляет, в частности, регистрацию микроколебаний воздушного потока голосового источника. Это обстоятельство указывает на значительную информационную ёмкость голосового источника [1] и на важность изучения формы импульсов основного тона, а также на важность выделения из речевого сигнала характерных особенностей функционирования гортани в процессе речеобразования.

Проблема информационной ёмкости колебательного процесса голосовых связок с точки зрения речеведения может быть сформулирована следующим образом: зависит ли характер этих вибраций гортани, пересекающих практически постоянный (медленно меняющийся) воздушный поток от диктора и от произносимого звука речи? Утвердительный ответы на поставленный вопрос повлечёт за собой целый ряд новых задач в исследовании речи в области моделирования процессов образования речи, а также идентификации и верификации диктора по речи, медицинской диагностики и распознавании речи. Классическая линейная модель речепроизводства, сосредотачивающая информацию о произносимой фонеме лишь в резонансных параметрах артикуляционного аппарата, при этом должна быть скорректирована и видоизменена [2,3,4].

Один из возможных алгоритмов выделения импульсов основного тона описан в настоящем сборнике [5]. Он также применим к изучению характерных особенностей функционирования голосового источника и основан на изучении формы импульсов, полученных в процессе нелинейного преобразования речевых колебаний. Как отмечалось в статье [5], метод «речевые

колебания» преобразует в импульсную последовательность на основе вычисления определителя автокорреляционной матрицы (определителя Грамма).

Сравнительный анализ импульсов для различных стационарных гласных одного диктора приведён в работе [6]. Он указывает на существовании зависимости формы колебаний голосовых связок от произносимого звука речи. Полученные импульсы для одного и того же диктора имели значительные отличия при исследовании различных гласных звуков речи. В настоящей работе приводятся результаты сравнения одного и того же гласного звука в слове, произнесённого разными дикторами.

Предложенный метод исследования фонационных характеристик речи имеет компенсационный и дифференцирующий характер по отношению к речевым колебаниям [6], что накладывает свой отпечаток на поведение получающейся импульсной функции. Этот метод анализа по своей конструктивной сути убирает из речевого сигнала и из функции возбуждения все затухающие гармоники, соответствующие свободным колебаниям линейной системы конечного порядка. При этом с увеличением порядка матрицы возрастают компенсационные возможности метода, и возрастает количество убираемых гармонических компонент. Так, например, для звука речи, содержащего четыре форманты, для их компенсации достаточно рассмотреть определитель девятого порядка (удвоенная величина количества формант, плюс единица). Но поскольку предложенное преобразование действует на весь речевой сигнал в целом, то помимо компенсации формантных колебаний, порожденных артикуляционным аппаратом, аннулируются такие составляющие из функции возбуждения. Если теперь представить ситуацию, в которой отклик линейной системы (имитация речевого сигнала) образован функцией возбуждения в виде затухающих колебаний (плохое «подражание» функционированию голосовых связок»), то предложенный метод анализа (при соответствующем выборе порядка) производит почти полную компенсацию колебательного процесса.

В результате сохраняются только выбросы амплитуд в местах «склейки» двух «плавных» (дифференцируемых) функций, образующих сигнал возбуждения линейной системы. Места нарушения аналитичности функции возбуждения будут в этом случае подчёркнуты в виде кратковременных всплесков амплитуды, что с математической точки зрения напоминает операцию дифференцирования сигнала на входе линейной системы. Метод исследования фонации подчёркивает «разрывы» производных в аналоговом или выделяет разности в дискретном варианте представления функции возбуждения речевого тракта.

Теоретически высказанное предположение можно подтвердить. Наиболее просто это сделать для дискретного варианта представления речевого сигнала, когда порядок матрицы совпадает размерностью векторов, на которых образуется автокорреляционная матрица. Последнее условие соответствует локальному методу анализа резонансных свойств речевого тракта [6].

В этом случае матрица R может быть представлена в виде матрицы X, умноженной на сопряжённую матрицу X^1:

$$R = X X^1, \tag{1}$$

где элементы — x_{ij} матрицы X образуются по дискретным значениям x речевого сигнала следующим образом:

$x_{ij} = x_{i+j1}$ для $i,j = 1,2,\ldots, m$.

Определитель матрицы R будет равен произведению определителей матрицы X и транспонированной матрицы X^1, которые будут (ввиду их симметричности) равны [7]:

$$M(n) = |R| = |X|^2 \tag{2}$$

Легко показать, что определитель $|X|$ матрицы X зависит от разностей различного порядка, образованных по дискретным отсчетам x_n речевого сигнала:

$$|X| = \begin{vmatrix} x_1, x_2, \ldots, x_m \\ x_2, x_3, \ldots, x_{m+1} \\ - - - - - - - \\ x_m, x_{m+1}, \ldots, x_{2m} \end{vmatrix} = \begin{vmatrix} x_1, x_2, \ldots x_m \\ \Delta_1^1, \Delta_2^1, \ldots, \Delta_m^1 \\ \Delta_1^{m-1}, \Delta_2^{m-1}, \ldots \Delta_m^{m-1} \end{vmatrix}, \quad (3)$$

где разности $\Delta_p^q = \Delta_{p+1}^{q-1} - \Delta_p^{q-1}$ q —го порядка, выраженные рекуррентным образом через разности (q - 1) — го порядка ($p,q = 0,1,\ldots, m$-1), при этом $\Delta_p^o = x_{p+1}, p = 0,1,\ldots,m-1$.

Равенство (3) получается рекуррентным вычитанием *к*-ой строки исходного определителя из всех последующих строк $к+1, к+2,\ldots, m$. Эта процедура начинается с *к* равного единице, и продолжается до *к* равного (m-1). Как известно [7], подобное преобразование строк не изменяет значения исходного определителя, и он будет равен вновь образованному определителю, зависящему от разностей разного порядка. Функция $M(n)$, равная определителю матрицы R, вычисляется по указанным разностям и содержит подчёркнутые моменты нарушения гладкости функции возбуждения.

Нетрудно распространить этот результат на общий случай, при котором выбор порядка автокорреляционной матрицы и величины усреднения на связаны столь жёстким условием равенства друг другу. Матрицу X, вообще говоря, прямоугольную, можно предварительно представить в виде соответствующих разностных компонент так же, как это сделано выше. Далее осуществить умножение её на сопряжённую матрицу (1) и лишь после этого вычислять определитель полученной матрицы R. Равенство (2) в рассматриваемом обобщённом варианте, естественно, нарушается. Тем не менее, общий вывод остаётся верным.

Применимость предложенного метода к исследованию фонационной картины речи проверялась для значений параметров преобразования в достаточно широких пределах. Предварительно речевой сигнал пропускался через фильтр (с конечной импульсной характеристикой) с полосой пропускания 300–3400 Гц, имитирующий телефонный канал связи. Запись производилась через звуковую карту персонального компьютера в среде Windows 98. Частота дискретизации — 12 кГц, количество бит на отсчёт — 16.

Размер матрицы *R* варьировался от трёх до девяти, а величина усреднения (продолжительность окна анализа, определяемая размерностью вектора) — от 2 до 30 мс. Во всех случаях импульсный характер преобразования сохранялся, а сами импульсы имели достаточно ярко выраженный характер по отношению к интервалам смыкании голосовых связок. В этих пределах значений параметров преобразования метод сохранял работоспособность [6].

Реализация данного метода была выполнена в среде MATLAB 5.2.

Для каждого дискретного момента времени *j* из выбранного участка формируется одномерный массив (вектор), состоящий из значений речевого сигнала $\{x(j),\ldots, x(j+N+p-1)\}$, где N — размерность векторов, p — порядок автокорреляционной матрицы. Каждый из этих массивов обрабатывается следующим образом:

1) ассив преобразуется в элементы матрицы X;

2) на основе формулы (2) формируется матрица $R = \{r_{km}\}$ размером p x p;

3) вычисляется определитель этой матрицы;

4) значение определителя запоминается в одномерном массиве.

Эта процедура повторяется в цикле для каждого дискретного момента времени выбранного диапазона. Таким образом, образуется одномерный массив, содержащий значения функции $M_p(n)$, длина которого равна длине выбранного пользователем участка.

Примером получаемых результатов может служить осциллограмма ударного звука «а» в слове «бар***а***нка» (Рис. 1), произнесённого диктором I мужчиной.

На следующем этапе полученная импульсная последовательность в линейном масштабе сегментировалась на импульсы, при этом значения меньше порогового 0,01 полагались равными нулю. Далее определялись координаты центров импульсов и все импульсы нормировались таким образом, чтобы ординаты центров принимали общее значение равное 1. Произведённые преобразования формы импульсов были направлены на то, чтобы убрать разброс импульсов по амплитуде и получить суммированием усреднённую форму импульса для каждого диктора.

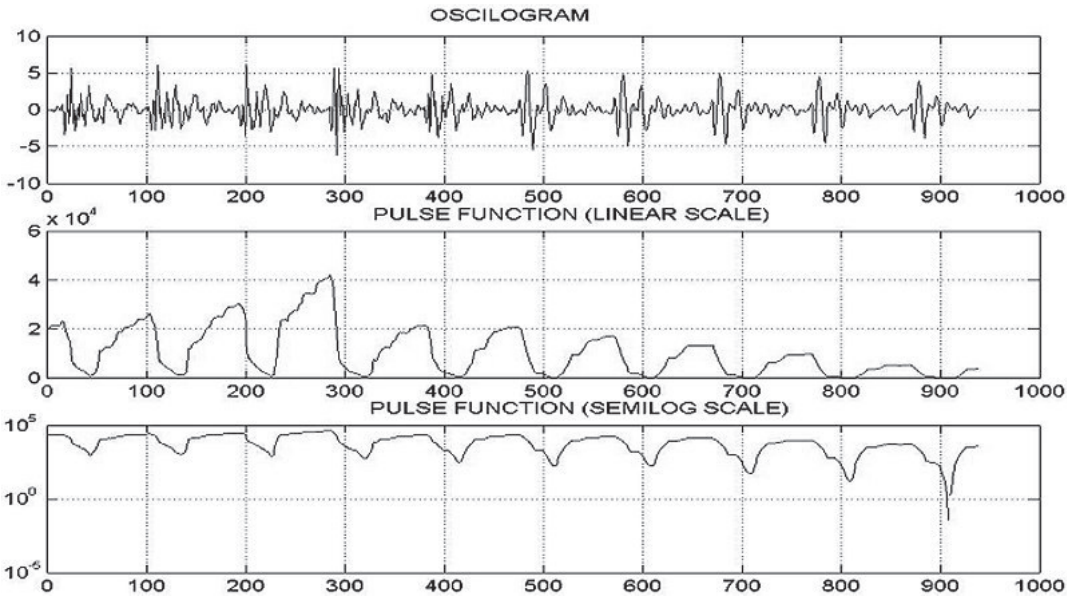

Рис.1. Осциллограмма звука «а» (верхний график), импульсная последовательность в линейном (средний график) и логарифмическом (нижний график) масштабе.

Прежде всего, усреднялись все импульсы по амплитуде. Затем каждый импульс отображался на общем рисунке. В результате получались усредненные импульсы и доверительные интервалы, соответствующие удвоенному среднеквадратическому отклонению. Полученные результаты изображены на рисунке 2.

Аналогичные результаты были получены для другого диктора II мужчины для того же слова «бар***а***нка» (рис. 3,4).

Анализ показывает явную зависимость формы полученных импульсов от диктора. Форма импульсов содержит некоторые компоненты, присущие, вероятно, индивидуальным особенностям фонационного аппарата диктора.

Для выявления этих характеристик достаточно провести корреляционный анализ полученных импульсов и получить «обобщённый» портрет аналога импульса с указанием соответствующих доверительных интервалов.

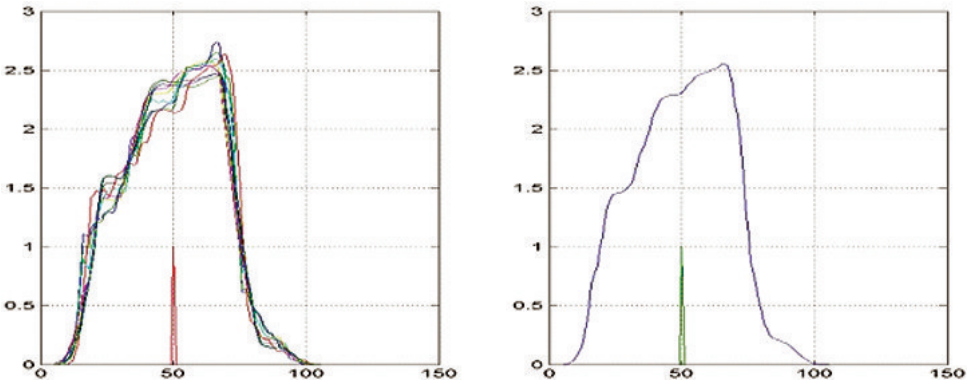

Рис. 2. Выделенные импульсы (левый график) и усреднённые значения (правый график)

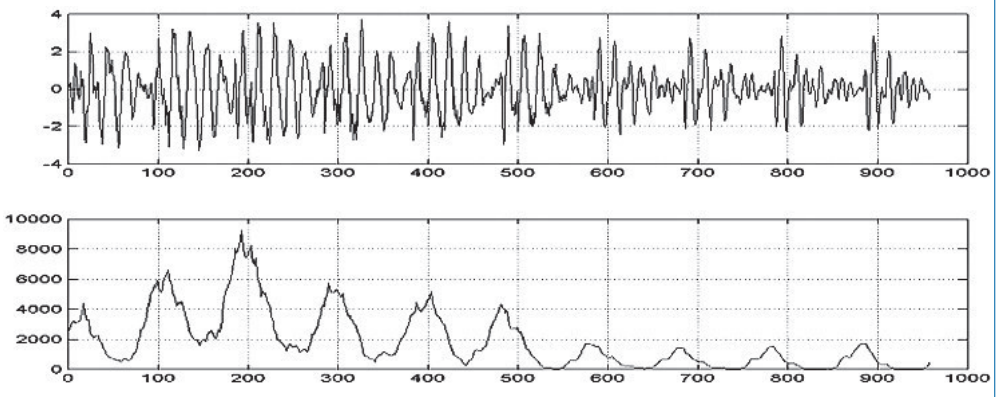

Рис. 3. Осциллограмма (верхний график) и импульсы (нижний график)

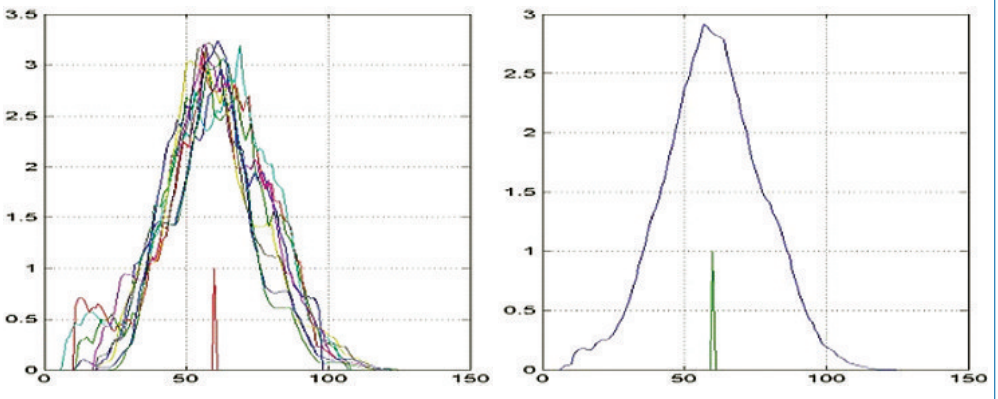

Рис. 4. Выделенные импульсы (левый график) и усреднённый импульс (правый график)

Полученные результаты можно использовать для сравнения обобщённых «портретов» импульсов между собой с целью верификации или идентификации разных дикторов.

ЛИТЕРАТУРА

1. *Ondrachkova J*. Glottographical research in sound Groups // Модели восприятия речи. Международный психологический конгресс. М.,1966. Л., 1966. P. 90–94.

2. *Галунов В.И., Тампель Н.Б*. Механизм работы голосового источника / Акустический журнал. М., 1981. Т. 27. Вып. 3. С. 321–334.

3. *Коваль С.Л., Лапина Л.В., Сапожкова И.Ф*. Синтез речи по правилам: Проблемы и перспективы // XV Всес. школа-семинар АРСО-XV: Тез. докл. и сообщ. Таллин, 1989. С. 25–31.

4. *Сорокин В.Н*. Теория речеобразования. М.: Радио и Связь, 1985. 312 с.

5. *Собакин А.Н*. Выделение импульсов основного тона по речевому сигналу. В наст. сб.

6. *Собакин А.Н*. Артикуляционные параметры речи и математические методы их исследования. Монография// Вестник МГЛУ. Вып. 517. М.: МГЛУ, 2006. 220 с.

7. *Гантмахер Ф.Р*. Теория матриц. М.: Наука, 1967. 567 с.