ФАЗОВЫЙ АНАЛИЗ НЕКОТОРЫХ ЗВУКОВ ЧЕЛОВЕЧЕСКОЙ РЕЧИ

В.В. Митянок
Полесский государственный университет
Пинск, Республика Беларусь
E-mail: mitsianok@mail.ru

Как хорошо известно, многие звуки речи человека являются почти периодическими функциями времени, то есть представляют собой сумму мод с различными параметрами – частотами, амплитудами, фазами. Так как параметры *реальных* звуков подвержены некоторым дрожаниям (возмущениям), то для адекватного математического описания звуковых кривых следует использовать выражение

$$y_i = b_{0i} + \sum_{k=1}^{l_1} c_{ki} \sin(k\omega_0 i + \varphi_{ki}), \quad k = 1..l_1,\ i = 1..n, \tag{1}$$

где y_i – оцифрованные значения кривой звукового давления, полученные через равные промежутки времени, c_{ki} – дрейфующие (медленно меняющиеся) амплитуды, φ_{ki} – дрейфующие фазы, ω_0 – величина базовой частоты сигнала, l_1 – количество мод, n – количество оцифрованных точек звуковой кривой. Индекс i нумерует последовательность моментов времени. Для вычисления дрейфующих параметров в [1, 2] предложены методы аппроксимации. Простой метод аппроксимации предполагает составление невязки, равной сумме (по всем точкам) квадратов отклонений аппроксимируемой функции от суммы нескольких гармонических функций с различными, но постоянными по изучаемому отрезку частотами и амплитудами. Частоты предполагаются известными, а амплитуды - нет. Как обычно поступают в методе наименьших квадратов, частные производные невязки по амплитудам приравниваются нулю и в результате получается система линейных алгебраических уравнений относительно неизвестных амплитуд. Решение данной системы позволяет найти эти амплитуды, что в итоге и дает решение задачи аппроксимации (при известных частотах). Если же частоты неизвестны, то можно поступить следующим образом. Выберем некоторые числа (наугад, либо исходя из каких-либо соображений) в качестве предполагаемого набора частот, проведем процедуру аппроксимации и, в конце, подсчитаем остаточную невязку. Затем выберем другой набор чисел в качестве возможных значений частот, вновь проведем процедуру аппроксимации и уже для нового набора частот подсчитаем остаточную невязку. После этого две полученные невязки сравним, и в качестве более правильного набора частот примем тот, который обеспечивает меньшую остаточную невязку. Перебор вариантов наборов частот можно проводить не наугад, а, например, методом скорейшего спуска. В результате переборов вариантов можно найти тот набор частот, который дает остаточную невязку меньшую, чем любой другой набор частот, и этот набор частот будет уже истинным. Частоты этого набора можно назвать несущими. Простой метод аппроксимации хорошо зарекомендовал себя на примере музыкальных звуков, однако он оказался недостаточно точен при анализе звуков человеческой речи. Одна из причин этого - проблемы выбора длительности отрезка звуковой кривой. Этот отрезок не может быть слишком коротким, так как итоговая точность вычислений будет недостаточно высокой, а с другой стороны, не может быть слишком длинным, так как параметры реальных речевых сигналов подвержены дрожаниям и дрейфу, и за длительное время претерпевают значительные изменения. Поэтому простой метод аппроксимации был усовершенствован. Усовершенствованный метод аппроксимации отличается от простого тем, что в невязку добавляются группы слагаемых, ответственных за величины прыжков значений параметров аппроксимирующих функций при последовательном переходе от точки к точке на звуковой кривой. В результате аппроксимации получается решение, допускающее изменение параметров аппроксимирующих функций при переходе от точки к точке, что, собственно и имеет место для реального голоса человека. В [3] этот метод применен к изучению поведения фаз различных мод гласных звуков. Оказалось, что хотя сами фазы мод являются равномерно распределенными на интервале случайными величинами, существуют такие их комбинации, которые являются компактными на диаграмме «базовая частота – фазовая комбинация», что может быть использовано для идентификации человека по его голосу и для распознавания речи. Среди специалистов, занимающихся проблемой автоматического распознавания речи, распространено мнение о том, что слуховой аппарат человека не воспринимает фазу звукового сигнала. Это, разумеется, не означает, что в произносимых звуках нет никаких закономерностей, связанных с фазами. В связи с этим в [3] выдвинуто предположение, что такие закономерности могут суще-

ствовать, и предложены комбинации (критерии)

$$Z = \sum_i \varphi_i - \sum_k \varphi_k, \text{ при условии } \sum i = \sum k. \quad (2)$$

Здесь i и k – номера мод звукового сигнала. В (2) произведено усреднение по всем точкам избранного отрезка звуковой кривой. Примерами критериев Z являются критерии $Z = 2\varphi_1 - \varphi_2$, $Z = \varphi_1 - 2\varphi_2 + \varphi_3$, $Z = \varphi_2 + \varphi_3 - \varphi_5$, $Z = \varphi_1 - \varphi_2 - \varphi_3 + \varphi_4$ и другие. Фазы, составляющие критерии, могут входить в них неоднократно. Выбор критериев в виде (2) обоснован тем, что они обладают двумя видами устойчивости. Во-первых, эти критерии не зависят от выбора начала отсчета времени, и, во-вторых, они не зависят от небольших погрешностей выбора базовой частоты ловящей сети. Изучение образцов гласных звуков, полученных от 11 респондентов показало, что в большинстве случаев наблюдаемые значения критериев (2) ложатся на диаграмме «базовый период – значение критерия» «кучно», часто их среднеквадратическое отклонение опускается ниже 0.3, а иногда достигает 0.15 (Для сравнения – случайная величина, равномерно распределенная на интервале $[0, 2\pi]$ имеет среднеквадратическое отклонение $\approx$1.81). Причем группировки точек, соответствующих одному и тому же звуку, но полученных от различных респондентов, находятся на диаграмме «базовый период – значение критерия» в разных местах. Таким образом, открывается перспектива разработки компьютерной программы, позволяющей производить идентификацию человека по голосу. Кроме того, критерии (2) для *различных звуков*, но полученных от *одного и того же человека* также занимают на этой диаграмме разные места. То есть можно также надеяться и на разработку программы распознавания речи на основе фазовых критериев.

Дрожание параметров мод звуковых кривых – это неизбежное явление, вызванное несовершенством речевого аппарата человека, и программы распознавания речи и идентификации человека должны это учитывать. Само понятие периода в этом случае уже не является строго определенным. Поэтому следует быть готовым к тому, что речевой сигнал придется анализировать на основе системы несущих частот, известных с некоторой погрешностью. К каким последствиям это может привести? Рассмотрим идеальный гармонический сигнал, состоящий из единственной несущей частоты ω, имеющий амплитуду A, фазу φ и пусть он анализируется методом аппроксимации на основе частоты, имеющей значение $\omega_1 \neq \omega$. Проведем очевидные преобразования

$$\begin{aligned} A\sin(\omega t + \varphi) &= A\sin(\omega_1 t + \varphi + \Delta\omega t) = \\ &= A\cos(\Delta\omega t)\sin(\omega_1 t + \varphi) + \\ &+ A\sin(\Delta\omega t)\cos(\omega_1 t + \varphi), \end{aligned} \quad (3)$$

где $\Delta\omega = \omega - \omega_1$. Как видно из (3), метод аппроксимации должен дать для дрейфующих амплитуд периодические функции с частотами, равными погрешности истинной частоты. Тестовые примеры подтверждают это предположение.

1. Митянок, В. В. О числовых характеристиках некоторых низкочастотных звуков человеческой речи / В. В. Митянок // Электронный журнал технической акустики www.ejta.org. – 2008. – № 15.
2. Митянок, В. В. Определение числовых характеристик высокочастотных звуков речи на основе аппроксимации гармоническими функциями / В. В. Митянок // Известия НАН Беларуси, Сер.ф.-м.н. – 2009. – № 2. – С. 111.
3. Митянок, В. В. Применение фазового анализа звуков речи для распознавания человека по его голосу / В. В.Митянок, Н.В.Коновалова // Электронный журнал технической акустики www.ejta.org. – 2013. – № 4.