

МОДЕЛИРОВАНИЕ СПЕКТРА ЗВУКОВ ЧЕЛОВЕЧЕСКОЙ РЕЧИ, ПРОИЗНЕСЕННЫХ В СОСТАВЕ СЛОГОВ, МЕТОДОМ АППРОКСИМАЦИИ

В. В. Митянок (Пинск, БЕЛАРУСЬ)

Задачу распознавания компьютером человеческой речи обычно решают, используя метод преобразований Фурье, который позволяет найти спектр *периодического* сигнала. Однако реальные кривые звукового давления, даже долгих гласных, не являются строго периодическими уже по одной той причине, что параметры звука (например частота, амплитуда) подвержены дрожаниям. Кроме того, метод преобразований Фурье имеет и собственные недостатки - результаты разложения в интеграл Фурье заметно зависят от длительности отрезка функции, отобранной для изучения, чувствительны к неточности измерения значений изучаемых функций, посторонним шумам и т. д. В противовес методу преобразований Фурье, в [1] предложен метод аппроксимации, в котором эти недостатки либо вообще не имеют место, либо заметно ослаблены.

Пусть требуется аппроксимировать (почти) периодическую функцию $y_i = y(x_i)$, заданную своими значениями в избранных точках x_i с помощью суммы (почти) гармонических функций. Для упрощения положим $x_i = i$ и рассмотрим невязку

$$S = \sum_{i=1}^n (y_i - b_{0i} - \sum_{k=1}^{l_1} a_{ki} \sin(\omega_k i) - \sum_{k=1}^{l_1} b_{ki} \cos(\omega_k i))^2 + \alpha \sum_{i=1}^{n-1} (b_{0i} - b_{0i+1})^2 + \\ + \alpha \sum_{k=1}^{l_1} \sum_{i=1}^{n-1} (b_{ki} - b_{ki+1})^2 + \alpha \sum_{k=1}^n \sum_{i=1}^{n-1} (a_{ki} - a_{ki+1})^2,$$

где l_1 – количество гармоник (мод), ω_k – их частоты (считаются известными), n – количество используемых точек изучаемой функции, a_{ki} , b_{ki} – соответственно дрейфующие амплитуды синус – и косинус волн, b_{0i} – дрейфующий нуль. Индекс i нумерует значения аргумента. Невязка S состоит из слагаемых – как содержащих так и не содержащих параметр α . Слагаемое, не содержащее α , описывает различие между аппроксимируемой

и аппроксимирующими функциями. Остальные описывают дрейф амплитуд. Неотрицательное число α показывает относительную важность слагаемых. Стандартный путь дифференцирования невязки по дрейфующим параметрам и последующее приравнение производных нулю приводит к составлению системы линейных алгебраических уравнений, методы решения которых хорошо известны.

Данный метод был применен к анализу слов, состоящих из повторяющихся несколько раз одинаковых слогов (мультислогов), в которых были использованы согласные Л,М,Н и гласные А,О,У,Э,Ы во всех возможных их комбинациях – всего 15 вариантов. Оказалось, что для разрезания кривой звуковой на фрагменты, соответствующие каждому из звуков слога, выгоднее всего использовать не те моды, которые имеют максимальный уровень амплитуды, а, наоборот – те, которые для одного из звуков имеют амплитуду, практически равную нулю, а для другого звука – заметно отличающуюся от нуля, но вовсе не обязательно – максимальную. Так, для мультислога ЛАЛАЛАЛА удачнее всего подходит моды N 6 и 7, для мультислога МАМАМАМА – подходят моды 3, 4, 5, 6, 7. Для мультислога МЫМЫМЫМЫ нужно использовать моды N 7 и 11 и т. д. Разрезание мультислога на отдельные звуки позволяет создавать динамические базы данных по амплитудам разных мод.

Литература

1. Митянок В.В. *Метод аппроксимации для нахождения числовых характеристик некоторых низкочастотных звуков человеческой речи* //Электронный журнал «Техническая акустика», <<http://www.ejta.org>>, 2008, 15.