УДК 534.87; 534.4

О ФАЗОВЫХ ОСОБЕННОСТЯХ НЕКОТОРЫХ ЗВУКОВ РЕЧИ

Митянок Вячеслав Владимирович, к.ф.–м.н, доцент,
Полесский государственный университет
MitsianokVatslav, PhD, Polessian State University, mitsianok@mail.ru

Аннотация: гласные звуки речи человека раскладываются на моды с дрейфующими амплитудами. Производится фазовый анализ полученных мод. Показано, что существуют фазовые комбинации, уникальные для каждого респондента.

Ключевые слова: распознавание речи, синтез речи, фазовый анализ звуков.

Задачи автоматического распознавания речи человека и автоматической верификации и распознавания личности по голосу до сих пор не имеют удовлетворительного решения. Во многом это связано с тем, что все еще неизвестны математические характеристики звуков человеческой речи. Кривые звукового давления во многих случаях представляют собой крайне запутанную картину. Представляет интерес вопрос о том, *что* именно делает звук «А» звуком «А», звук «О» звуком «О» и т.д. Какие именно математические характеристики звуков здесь существенны, какие привнесены несовершенством аппарата речеобразования человека, какие позволяют отличать одного диктора от другого, а какие вообще ни за что не несут ответственности, и попали в состав звуков случайно.

Как известно, метод преобразований Фурье, используемый для нахождения спектра звуков, обладает рядом недостатков. В частности, в спектре присутствуют фальшивые линии, линии спектра даже в случае идеального гармонического сигнала, но рассмотренного на ограниченном интервале времени, размыты. (В квантовой механике это обстоятельство является математической подоплекой хорошо известного соотношения неопределенностей). Спектр сигнала существенно зависит от его длительности. Если в исследу-

емом сигнале присутствуют малоинтенсивные моды, то они могут оказаться скрытыми под фальшивыми линиями. Поэтому в [1,2] была поставлена задача нахождения спектра звуковых сигналов методом аппроксимации, с учетом того, что звуковые сигналы, соответствующие отдельным звукам речи человека представляют собой сумму мод, параметры которых (амплитуды, частоты, фазы) могут слегка меняться в процессе звучания, дрейфовать, дрожать, то есть зависеть от времени.

Пусть формула звукового сигнала имеет вид

$$y(t_i) = B_{i,0}(t_i) + \sum_{k=1}^{n}\left[A_{i,k}(t_i)\sin(\omega_k i) + B_{i,k}(t_i)\cos(\omega_k i)\right], \quad (1)$$

где $A_{i,k}$ и $B_{i,k}$ $i = 1..n$ $k = 1..l$ – дрейфующие, т.е. медленно изменяющиеся (по сравнению с несущими частотами ω_k) амплитуды сигнала, $B_{i,0}$ дрейфующий нуль, t_i– время, индекс i нумерует оцифровываемые моменты времени, l –количество мод, составляющих звук, n – количество оцифрованных точек. Составим функционал

$$S = \sum_{i=1}^{n}\left[y(t_i) - y_1(t_i)\right]^2 + \alpha\sum_{i=1}^{n-1}\left[b_{i,0} - b_{i+1,0}\right]^2 + \alpha\sum_{k=1}^{l}\sum_{i=1}^{n-1}\left[a_{i,k} - a_{i+1,k}\right]^2 + \alpha\sum_{k=1}^{l}\sum_{i=1}^{n-1}\left[b_{i,k} - b_{i+1,k}\right]^2 \quad (2)$$

где $y(t_i)$ – аппроксимируемая функция, заданная выражением (1), а

$$y(t_i) = b_{i,0}(t_i) + \sum_{k=1}^{n}\left[a_{i,k}(t_i)\sin(\omega_k i) + b_{i,k}(t_i)\cos(\omega_k i)\right] \quad (3)$$

– аппроксимирующая. В (1) – (2) для простоты примем $t_i = i$, хотя это и не обязательно. Параметр α в (1) позволяет сглаживать изменения амплитуд волн при переходе от точки к точке по оси времени. Чем больше значение α, тем более гладкими являются амплитуды волн.

Вычисляя частные производные (2) по дрейфующим амплитудам и по дрейфующему началу отсчета и приравнивая результаты нулю, получим систему линейных алгебраических уравнений относительно параметров аппроксимирующей функции. Решив эту систему, найдем эти параметры и тем самым произведем разложение аппроксимируемой функции на набор волн с медленно меняющимися амплитудами. Просуммировав их по формуле (3), найдем аппроксимирующую функцию. Если ее вычесть из исходного звука (1) и подвергнуть разность преобразованиям Фурье, то часто выясняется, что существуют еще какие–то несущие частоты, которые не были замечены при первом разложении в интеграл Фурье по причине малой интенсивности несомых ими мод. В частности, этим способом в [3] было установлено, что в спектре звуков «З», «ЗЬ», «Ж», «ЖЬ» присутствуют полуцелые (по отношению к базовой) несущие частоты.

Каждую из мод, входящую в (3) можно переписать в физически более информативном виде:

$$a_{i,k}\sin(\omega_k t_i) + b_{i,k}\cos(\omega_k t_i) = c_{i,k}\sin(\omega_k t_i + \varphi_{i,k}). \quad (4)$$

Тогда аппроксимирующая функция выглядит так:

$$y_1(t_i) = b_{i,0} + \sum_{k=0}^{l} c_{i,k}\sin(\omega_k t_i + \varphi_{i,k}). \quad (5)$$

Здесь $c_{k,i}$ – дрейфующая общая амплитуда моды, $\varphi_{k,i}$ – дрейфующая фаза моды.

В настоящем исследовании изучались звуки «А», «О», «У», «Э», «Ы», «И». Эти звуки были отобраны для изучения по той причине, что их можно произносить достаточно долго и от этого они не теряют свою индивидуальность в отличие от звуков «Я», «Е» и других, которые при длительном звучании преобразуются соответственно в звуки «А», «Э» и т.д.

Для изучения вышеуказанных звуков, методом преобразований Фурье определялась в нулевом приближении система несущих частот, нижняя из которых назначалась базовой. Эта система несущих частот дополнялась теми частотами, которые остались незамеченными методом Фурье при первом его использовании. Для этого использовался вышеописанный способ. Затем эта система частот дополнялась полуцелыми частотами, составляющими 0.5, 1.5, 2.5, 3.5 от базовой. В результате получалась сеть несущих частот, которая и использовалась для окончательного разложения звуков на моды. При анализе фаз различных мод разложенных звуков выявилось следующее: Для всех звуков и всех респондентов фазы не являются независимыми. Так, если ввести в рассмотрение нормированные фазы целых мод

$$\varphi_{i,k}^{norm} = \frac{\varphi_{i,k}}{k}, \quad i = 1..m, \quad k = 1..l \tag{6}$$

то отчетливо видно, что фазы хотя и хаотичны, но синхронны (рис.1,2). Естественно считать, что модой, задающей поведение фаз других мод, является базовая мода.

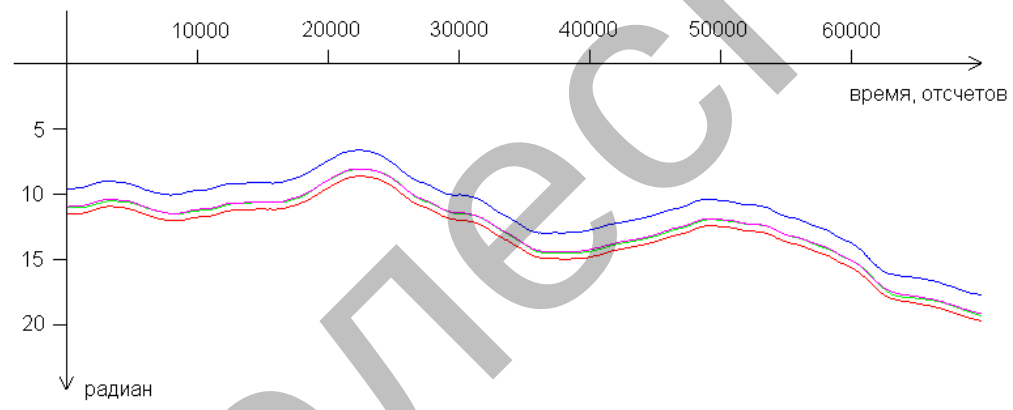

Рисунок 1 – Фазы нижних мод, несомых целыми частотами, деленные на номер частоты (нормированные фазы). Звук «О», респондент Митянок. Частота дискретизации 44100 Гц. Базовая мода – красный цвет, вторая мода – фиолетовый, третья – зеленый, четвертая – синий.

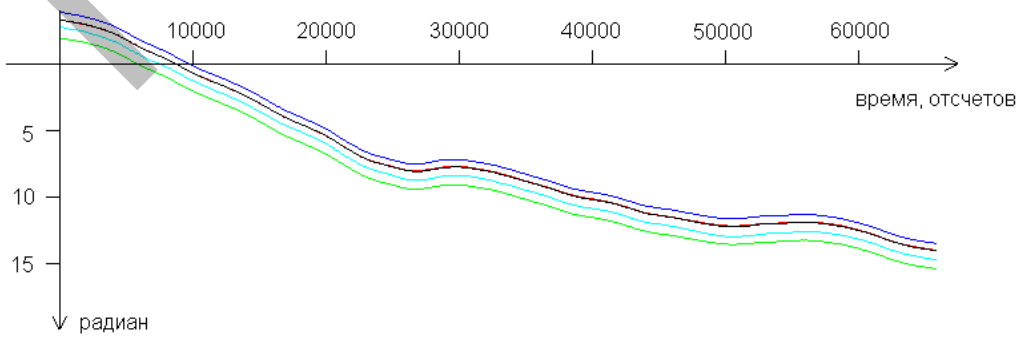

Рисунок 2 – Фазы мод, несомых целыми частотами, деленные на номер частоты (нормированные фазы). Звук «Э», респондент Янковский. Частота дискретизации 44100 Гц. Базовая мода – красный цвет, вторая мода –

фиолетовый, третья – зеленый, четвертая – синий , пятая – бирюзовый.

Из рисунков 1 и 2 также видно, что расстояния между нормированными фазами других мод не зависят от времени. Но они, в свою очередь, зависят от того, *какой* звук произносится, и *кто* его произносит. Это можно использовать для создания систем идентификации и верификации человека по голосу. Для этого подойдет критерий

$$K_i = \sum_{j=1}^{n} (\varphi_{1,j} - \frac{\varphi_{i,j}}{i}). \qquad (7)$$

На рис. 3 представлены гистограммы критерия (7)

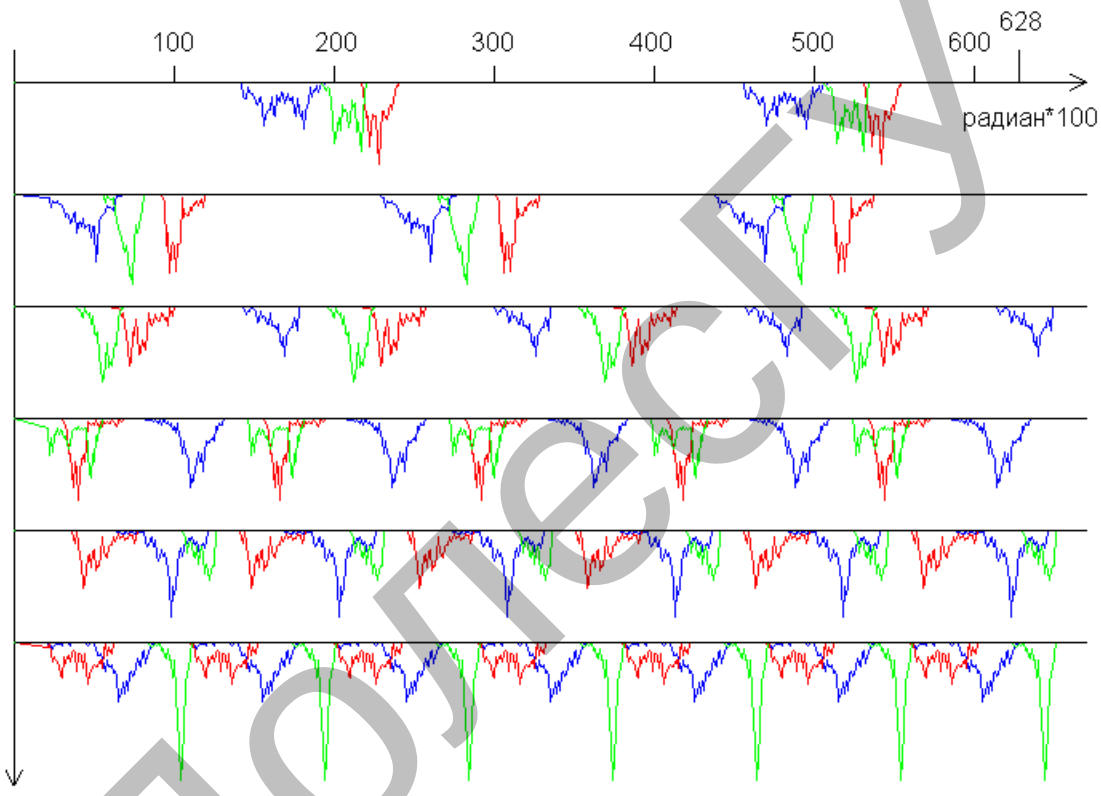

Рисунок 3 – Гистограммы критериев (7) звука А в исполнении респондентов Коновалова – красные линии, Романова – зеленые линии, Коваленко – синие линии

Для злонамеренного звукоподражателя, каким бы талантливым он ни был, станет сюрпризом, что его личность можно идентифицировать по тем характеристикам звуков, которые ухом не различаются.

Список использованных источников:

1. Митянок, В.В. Определение числовых характеристик высокочастотных звуков речи на основе аппроксимации гармоническими функциями. Известия НАН Беларуси, сер. ф.–м.н.–2009.– , №2– с.111–118.

2. Митянок, В.В. / Техническая акустика. – Электрон. журн.– 2008.–15.– Режим доступа: http://www.ejta.org, свободный

3. Митянок, В.В. О физической структуре звуков З, ЗЬ, Ж, ЖЬ [Электронный ресурс] /В.В. Митянок.// Техническая акустика. – Электрон. журн.– 2014.–9.– Режим доступа: http://www.ejta.org, свободный