

Формантный анализ речевого сигнала в фазовой области ¹

А.С.Леонов*, В.Н.Сорокин**

* *Национальный исследовательский ядерный университет «МИФИ», Москва, Россия*

** *Институт проблем передачи информации, Российская академия наук, Москва, Россия*

Поступила в редколлегию 01.06.2021

Аннотация—Исследуются математические свойства некоторой комплексной аналитической функции частоты и времени, получаемой по абстрактному сигналу. Эти свойства связывают амплитудные и фазовые характеристики сигнала. Установлено, что если собственные частоты генератора сигнала не меняются во времени, то они могут быть определены по экстремумам или точкам перегиба производной фазы по частоте при условии, что смешанная производная фазы по частоте и времени отрицательна. Описываются численные эксперименты по определению с помощью теории, развитой в статье, формантных частот синтетических и реальных речевых сигналов.

КЛЮЧЕВЫЕ СЛОВА: формантные частоты, аналитический сигнал, фазы речевого сигнала.

1. ВВЕДЕНИЕ

Знание резонансных частот речевого тракта необходимо для решения обратной задачи определения формы речевого тракта, аутентификации диктора, автоматического распознавания речи, определения диалекта языка и акцента носителя иностранного языка. Эти частоты в реальном речевом сигнале отождествляются с пиками его амплитудно-частотной характеристики (спектра). Области этих пиков, называются формантами, а частоты пиков – формантными частотами. Существует два основных метода нахождения формантных частот: непосредственный анализ амплитуд спектра речевого сигнала в частотной области и определение полюсов передаточной функции речевого тракта методом линейного предсказания [1]. Исследования показали, что оба метода могут давать существенные погрешности. Поэтому необходима разработка новых методов формантного анализа речевого сигнала.

Одним из них может оказаться определение формантных частот с использованием фазовых характеристик сигнала ("в фазовой области"). Традиционно стараются найти такое описание речевого сигнала в фазовой области, которое воспроизводило бы амплитудно-частотный спектр и позволило бы оценить формантные частоты по его пикам. Известны единичные примеры такого подхода. В них для оценки формант используется производная фазы по частоте для кратковременного преобразования Фурье сигнала (так называемая групповая задержка) или аналогичная производная по времени (мгновенная частота) [2–5].

Строгое математическое исследование фазовой функции невозможно без существенных предположений о свойствах сигнала. Так, в [6] такое исследование выполняется в предположении, что сигнал генерируется линейной системой с дробно-рациональной передаточной функцией, на параметры которой накладываются ряд ограничений. В работе [7] исследования фазы проводятся для так называемого аналитического сигнала, т.е. комплексного сигнала, зависящего от действительного аргумента и равного пределу некоторой аналитической функции

¹ Первый автор использовал финансовую поддержку Программы повышения конкурентоспособности Национального исследовательского ядерного университета МИФИ (проект 02.а03.21.0005 от 27.08.2013).

при стремлении мнимой части ее аргумента к нулю. Во всяком случае, необходимо отличать реальный сигнал от его идеализированных моделей. Для реального речевого сигнала даже простейшие модельные предположения, не адекватные его физической природе, часто приводят к математическим результатам, не позволяющим использовать их в приложениях. Тем не менее, идеализированные модели речевого сигнала продолжают использоваться в речевых технологиях и в ряде случаев приводят к практически приемлемым результатам. Например, крайне идеализированная модель источника голосового возбуждения в виде последовательности знакопеременных δ -образных импульсов позволила в [8] сформулировать успешный алгоритм определения моментов начала и конца действия реального голосового источника, частоты основного тона и длительности интервала открытой голосовой щели.

В этой статье представлено исследование математических свойств некоторой комплексной аналитической функции, сформированной по заданному абстрактному сигналу, зависящей от времени и частоты и представленной с помощью ее амплитудной и фазовой функции. Считается, что максимумы амплитудной функции по частоте (при фиксированном времени) соответствуют "формантным частотам". Цель работы состоит в определении возможных условий оценки формантных частот по фазовой функции и в экспериментальной проверке возможности использования установленной связи между амплитудными и фазовыми характеристиками для определения формантных частот различных речевых сигналов.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

При анализе сигналов различной природы часто используется их преобразование Фурье. Если исследуются дискретизированные сигналы с относительно небольшой частотой отсчетов, то вместо преобразования Фурье обычно применяется кратковременное преобразование Фурье (КПФ, см., например, [6]), использующее скользящий сегмент сигнала с центром в момент времени t . В результате получается комплексная функция $S(\omega, t)$, зависящая от круговой частоты ω и t . Аналогичная функция может быть получена с помощью различных вейвлет-преобразований и другими способами. В случае речевого сигнала функция $S(\omega, t)$ содержит информацию о формантных частотах, спектре фрикативных звуков, частоте основного тона, типе смычки. Эту функцию можно представить в виде

$$S(\omega, t) = u(\omega, t) + iv(\omega, t) = A(\omega, t)e^{i\Psi(\omega, t)}.$$

В дальнейшем будем называть величину

$$A(\omega, t) = |S(\omega, t)|$$

амплитудой функции $S(\omega, t)$ (или амплитудно-частотной характеристикой, АЧХ), а $\Psi(\omega, t)$ – ее фазой (фазово-частотной характеристикой, ФЧХ). Определим "формантные частоты" абстрактного сигнала как точки максимумов АЧХ при фиксированном t . Задача состоит в определении условий, при которых можно найти (оценить) эти формантные частоты по ФЧХ.

Математически, это формулируется как следующая проблема: можно ли оценить положения максимумов $\omega_0 = \omega_0(t)$ функции $A(\omega, t)$ (амплитуды сигнала) при каждом фиксированном t , используя только функцию $\Psi(\omega, t)$ (фазу). Простые примеры показывают, что без серьезных ограничений на функцию $S(\omega, t)$ этого сделать нельзя. Поэтому мы в дальнейшем считаем, что $S(\omega, t)$ есть аналитическая функция комплексного переменного $\omega + it$ в некоторой области $G = (\omega_1, \omega_2) \times (t_1, t_2)$. Обозначим частные производные по ω штрихом, а по t точкой. Для аналитической функции $S(\omega, t)$ переменных ω, t в области G должны выполняться условия Коши–Римана: $u' = \dot{v}$, $\dot{u} = -v'$, и это является основным фактором, обеспечивающим нахождение максимумов амплитуды по фазе.

Производные фазы по частоте и времени можно записать как

$$\begin{aligned}\Psi'(\omega, t) &= \frac{v'u - u'v}{u^2 + v^2}, \quad \Psi''(\omega, t) = \frac{(v''u - u''v)(u^2 + v^2) - 2(v'u - u'v)(uu' + vv')}{(u^2 + v^2)^2}, \\ \dot{\Psi}'(\omega, t) &= \frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega, t) = -\frac{(\dot{u}^2 + \dot{v}^2 + u\ddot{u} + v\ddot{v})(u^2 + v^2) - 2(u\dot{u} + v\dot{v})^2}{(u^2 + v^2)^2}.\end{aligned}\quad (1)$$

Выражения (1) получаются дифференцированием из соотношения

$$i\Psi(\omega, t) = \ln(u(\omega, t) + iv(\omega, t)) - \frac{1}{2} \ln(u^2(\omega, t) + v^2(\omega, t))$$

с учетом условий Коши–Римана. Кроме формул (1), будет также использована

Лемма. Пусть функция $A(\omega, t)$ имеет при фиксированном t , $t \in (\alpha, \beta)$, максимум в точке ω_0 , $\omega_0 \in (\omega_1, \omega_2)$, причем $A(\omega_0, t) \neq 0$. Тогда справедлива формула

$$\Psi''(\omega_0, t) = \left[\frac{v''u - u''v}{u^2 + v^2} \right]_{\omega=\omega_0} = \left[\frac{\ddot{u}v - \ddot{v}u}{u^2 + v^2} \right]_{\omega=\omega_0}. \quad (2)$$

Доказательство. Точка ω_0 будет также точкой локального максимума функции

$$A^2(\omega, t) = |S(\omega, t)|^2 = u^2(\omega, t) + v^2(\omega, t),$$

и поэтому из необходимого условия максимума амплитуды: $A'(\omega_0, t) = 0$ – следует:

$$\frac{1}{2} (A^2(\omega, t))' \Big|_{\omega=\omega_0} = A(\omega_0, t)A'(\omega_0, t) = (u(\omega, t)u'(\omega, t) + v(\omega, t)v'(\omega, t)) \Big|_{\omega=\omega_0} = 0. \quad (3)$$

Последнее равенство можно записать, применяя условия Коши–Римана так:

$$u(\omega_0, t)u'(\omega_0, t) + v(\omega_0, t)v'(\omega_0, t) = u(\omega_0, t)\dot{v}(\omega_0, t) - v(\omega_0, t)\dot{u}(\omega_0, t) = 0. \quad (4)$$

Далее, используя (1) при $\omega = \omega_0$ получим:

$$\Psi''(\omega_0, t) = \left[\frac{(v''u - u''v)(u^2 + v^2) - 2(v'u - u'v)(uu' + vv')}{(u^2 + v^2)^2} \right]_{\omega=\omega_0} = \left[\frac{(v''u - u''v)}{u^2 + v^2} \right]_{\omega=\omega_0} \quad (5)$$

Здесь учтено, что в силу (4) $(uu' + vv')|_{\omega=\omega_0} = 0$. Из условий Коши–Римана следует гармоничность функций $u(\omega, t), v(\omega, t)$: $u''(\omega, t) + \ddot{u}(\omega, t) = 0$, $v''(\omega, t) + \ddot{v}(\omega, t) = 0$. Поэтому

$$(v''u - u''v)|_{\omega=\omega_0} = \ddot{u}(\omega_0, t)v(\omega_0, t) - \ddot{v}(\omega_0, t)u(\omega_0, t).$$

Подставляя это в (5), получим доказываемую формулу.

Центральным теоретическим результатом этой статьи является

Теорема 1. Предположим, что $S(\omega, t)$ есть аналитическая функция комплексного переменного $\omega + it$ в некоторой области $G = (\omega_1, \omega_2) \times (t_1, t_2)$, а ее амплитуда $A(\omega, t)$ имеет при каждом $t \in (\alpha, \beta)$ максимум в одной и той же точке $\omega_0 \in (\omega_1, \omega_2)$, причем $A(\omega_0, t) \neq 0$. Тогда точка $\omega = \omega_0$ есть стационарная точка функции $\Psi'(\omega, t)$ при каждом фиксированном $t \in (\alpha, \beta)$, т.е. выполнено равенство $\Psi''(\omega_0, t) = 0$.

Доказательство. Введем функции переменной t : $U(t) = u(\omega_0, t)$, $V(t) = v(\omega_0, t)$. Тогда равенство (4) можно записать как $U(t)\dot{V}(t) - V(t)\dot{U}(t) = 0$, $\forall t \in (\alpha, \beta)$, откуда $\frac{d}{dt} \left(\frac{V}{U} \right) = 0$. Поэтому $V(t) = C_0 U(t)$ для любого $t \in (\alpha, \beta)$. Учтем это равенство в формуле (2):

$$\begin{aligned} \Psi''(\omega_0, t) &= \frac{\ddot{u}(\omega_0, t)v(\omega_0, t) - \ddot{v}(\omega_0, t)u(\omega_0, t)}{u^2(\omega_0, t) + v^2(\omega_0, t)} = \frac{\ddot{U}(t)V(t) - \ddot{V}(t)U(t)}{U^2(t) + V^2(t)} = \\ &= \frac{\ddot{U}(t)C_0U(t) - C_0\ddot{U}(t)U(t)}{U^2(t) + C_0^2U^2(t)} = 0, \quad \forall t \in (\alpha, \beta). \end{aligned}$$

Таким образом, при любом $t \in (\alpha, \beta)$ выполнено необходимое условие экстремума функции $\Psi'(\omega, t)$ в точке $\omega = \omega_0$: $\Psi''(\omega_0, t) = 0$. Теорема доказана.

Теперь предположим, что функция $A(\omega, t)$, а поэтому и $A^2(\omega, t)$, при каждом $t \in (\alpha, \beta)$ имеет зависящую от t точку локального максимума $\omega = \omega_0(t)$ и последняя функция непрерывно дифференцируема. Выясним, выполнено ли при каждом $t \in (\alpha, \beta)$ необходимое условие экстремума $\Psi''(\omega_0(t), t) = 0$.

Теорема 2. $\Psi''(\omega_0(t), t) = 0$ только для тех $t \in (\alpha, \beta)$, для которых $\dot{\omega}_0(t) = 0$ и при этом

$$\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega_0(t), t) < 0$$

для любого $t \in (\alpha, \beta)$.

Доказательство. Запишем условия максимума функции $A^2(\omega, t)$ при каждом $t \in (\alpha, \beta)$, используя условия Коши–Римана и гармоничность функций u, v :

$$u(\omega_0(t), t)u'(\omega_0(t), t) + v(\omega_0(t), t)v'(\omega_0(t), t) = u(\omega_0(t), t)\dot{v}(\omega_0(t), t) - v(\omega_0(t), t)\dot{u}(\omega_0(t), t) = 0, \quad (6)$$

$$(A^2(\omega_0(t), t))'' = (u'^2 + uu'' + v'^2 + vv'')_{\omega=\omega_0(t)} = (\dot{v}^2 - u\ddot{u} + \dot{u}^2 - v\ddot{v})_{\omega=\omega_0(t)} < 0.$$

Здесь первое равенство – это необходимое условие экстремума (3), а второе неравенство есть достаточное условие максимума. Из него следует, что

$$(u\ddot{u} + v\ddot{v})_{\omega=\omega_0(t)} > (\dot{u}^2 + \dot{v}^2)_{\omega=\omega_0(t)}. \quad (7)$$

Тогда из формул (1) получим, используя неравенство (7), а затем равенство (6):

$$\begin{aligned} \frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega_0(t), t) &= - \frac{(\dot{u}^2 + \dot{v}^2 + u\ddot{u} + v\ddot{v})(u^2 + v^2) - 2(u\dot{u} + v\dot{v})^2}{(u^2 + v^2)^2} \Big|_{\omega=\omega_0(t)} < \\ &< - \frac{2(\dot{u}^2 + \dot{v}^2)(u^2 + v^2) - 2(u\dot{u} + v\dot{v})^2}{(u^2 + v^2)^2} \Big|_{\omega=\omega_0(t)} = - \frac{2(u\dot{v} - v\dot{u})^2}{(u^2 + v^2)^2} \Big|_{\omega=\omega_0(t)} = 0. \end{aligned}$$

Поэтому $\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega_0(t), t) < 0$ для любого $t \in (\alpha, \beta)$.

Теперь исследуем величину $\Psi''(\omega, t)$ при $\omega = \omega_0(t)$. Функция $\omega = \omega_0(t)$ удовлетворяет уравнению (6). Поэтому ее производная находится из равенства

$$\begin{aligned} \frac{d\omega_0(t)}{dt} &= - \frac{\frac{\partial(u(\omega, t)u'(\omega, t) + v(\omega, t)v'(\omega, t))}{\partial t}}{\frac{\partial(u(\omega, t)u'(\omega, t) + v(\omega, t)v'(\omega, t))}{\partial \omega}} \Big|_{\omega=\omega_0(t)} = - \frac{(u\dot{v} - v\dot{u})'_t}{u'^2 + uu'' + v'^2 + vv''} \Big|_{\omega=\omega_0(t)} = \\ &= - \frac{\dot{u}\dot{v} + u\ddot{v} - \dot{v}\dot{u} - v\ddot{u}}{\dot{v}^2 - u\ddot{u} + \dot{u}^2 - v\ddot{v}} \Big|_{\omega=\omega_0(t)} = - \frac{u\ddot{v} - v\ddot{u}}{\dot{v}^2 + \dot{u}^2 - (u\ddot{u} + v\ddot{v})} \Big|_{\omega=\omega_0(t)}. \end{aligned}$$

Таким образом,

$$(\ddot{u}v - \ddot{v}u)|_{\omega=\omega_0(t)} = \dot{\omega}_0(t) (\dot{v}^2 + \dot{u}^2 - (u\ddot{u} + v\ddot{v}))|_{\omega=\omega_0(t)}.$$

Отсюда и из (1) получим:

$$\Psi''(\omega_0(t), t) = \frac{\dot{\omega}_0(t) (\dot{v}^2 + \dot{u}^2 - (u\ddot{u} + v\ddot{v}))}{u^2(\omega, t) + v^2(\omega, t)} \Big|_{\omega=\omega_0(t)}.$$

В силу неравенства (7) эта величина может обращаться в нуль только для тех времен $t \in (\alpha, \beta)$, для которых $\dot{\omega}_0(t) = 0$. Теорема доказана.

Таким образом, в случае зависимости точки максимума от времени: $\omega = \omega_0(t)$ необходимое условие экстремума функции $A(\omega, t)$ по аргументу ω выполняется только для тех t , для которых $\dot{\omega}_0(t) = 0$. Например, это верно для случая $\omega_0(t) = \text{const}$ (как в Теореме 1).

Заметим, что неравенство $\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega_0, t) < 0$ остается верным и при условиях теоремы 1.

Предполагая, что реальный речевой сигнал позволяет найти аналитическую функцию $S(\omega, t)$, подчиненную условиям теорем 1 и 2, можно интерпретировать результаты этих теорем следующим образом. Если формантная частота ω_0 речевого сигнала не зависят от времени, то она находится среди стационарных точек (точек, подозрительных на экстремум) производной функции фазы $\Psi'(\omega, t)$, т.е. может быть найдена из уравнения $\Psi''(\omega_0, t) = 0$ для каждого допустимого t . Если же она зависит от времени, т.е. $\omega_0 = \omega_0(t)$, и такой трек форманты не имеет точек остановки (т.е. $\dot{\omega}_0(t) \neq 0$), то условие $\Psi''(\omega_0, t) = 0$ в этом случае не выполнено при всех допустимых t . Но при этом для точек экстремума и, в частности, для формант $\omega_0(t)$ выполнено неравенство $\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega_0(t), t) < 0$ при всех $t \in (\alpha, \beta)$. Поэтому в рамках сделанных предположений о $S(\omega, t)$ найти форманты по стационарным точкам фазы можно, вообще говоря, лишь для стационарных формант.

3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Проверим экспериментально, насколько теоретические результаты теорем 1 и 2 для аналитической функции $S(\omega, t)$, полученной по абстрактному сигналу, справедливы для речевых сигналов. На первом этапе такой проверки используем синтетические речевые сигналы, для которых и формантные частоты, и форма сигнала, и последовательность импульсов голосового возбуждения известна. Синтезируем несколько сигналов, соответствующих гласным /а, э, о, у, и, ы/ с фиксированными резонансными частотами ω_k , $k = 1, 2, \dots$, чтобы удовлетворить требованию $\dot{\omega}_k(t) = 0$ длительностью 0.5 с. До и после каждого гласного встраивается сигнал с нулевой амплитудой и с длительностью 0.5 с. Функция $S(\omega, t)$ вычисляется по таким сигналам $f(t)$ с помощью дискретизированного кратковременного преобразования Фурье на скользящем интервале в 40 мс в окне Гаусса $w(t)$:

$$S(\omega, t) = \int_{-\infty}^{+\infty} w(t - \tau) f(\tau) e^{-i\omega\tau} d\tau. \quad (8)$$

Фазы $\Psi(\omega, t)$ именно таких функций использована ниже. Поскольку ожидается, что в области резонансных частот гласного смешанная производная $\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega, t)$ отрицательна, рассмотрим ее поведение на частоте первого резонанса. На рис. 1а показана последовательность импульсов нормированной объемной скорости, определяющей голосовой источник, и смешанная производная фазы $\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega_1, t)$ на частоте первого резонанса $\omega_1 = 600$ Гц синтезированного стационарного гласного /а/ (рис. 1б). Видно, что смешанная производная стабильно отрицательна в конце периода основного тона, когда влияние источника на собственные колебания тракта

минимально и в начале следующего периода, когда действие источника на колебания в тракте только начинает сказываться. Отметим, что используемая для анализа формант функция (8) в данном случае не является аналитической функцией комплексного аргумента. Формально, применять теорию из раздела 2 здесь нельзя. Тем не менее, этот численный эксперимент показывает, что, по крайней мере качественно, результаты теорем 1 и 2 могут реализовываться для синтетических гласных. Подробнее об этом сказано в следующем разделе.

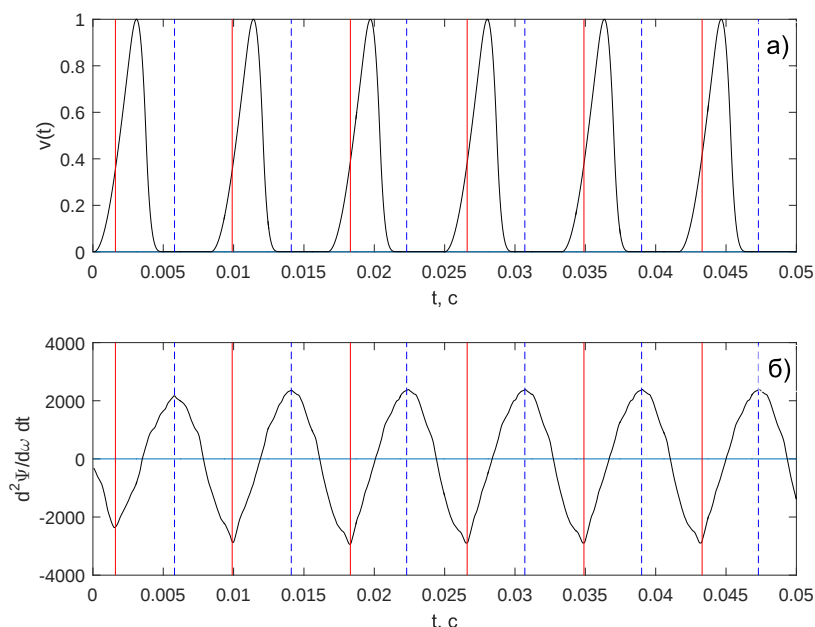


Рис. 1. Нормированная объемная скорость импульсов голосового источника (а) и смешанная производная фазы на частоте первого резонанса (б).

Само по себе, условие отрицательности смешанной производной не указывает на то, в какой момент времени, согласно теоремам 1 и 2, экстремум или перегиб функции $\Psi'(\omega, t)$ следует принимать за k -ю резонансную частоту. Поэтому на практике необходимо использовать дополнительные условия идентификации форманты. Например, можно выбирать в качестве оценки те стационарные точки этой функции по частоте (при фиксированном времени), где смешанная производная в этот же момент времени имеет локальный минимум по частоте. На рис. 2а показаны спектр сигнала (АЧХ) для гласного /а/, функция $\Psi'(\omega, t_n)$ (рис. 2б) и смешанная производная $\frac{\partial^2 \Psi}{\partial \omega \partial t}(\omega, t_n)$ (рис. 2в) для последовательности моментов времени t_n , в которых смешанная производная минимальна (см. рис. 1б). Резонансные частоты, соответствующие максимумам спектра, отмечены вертикальными линиями. Видно, что отрицательным пикам смешанной производной с хорошей точностью соответствуют положительные экстремумы производной фазы по частоте. При этом и отрицательные пики смешанной производной также хорошо аппроксимируют резонансные частоты гласного.

Оказалось, что такое соотношение производной фазы по частоте и смешанной производной фазы справедливо для всех синтезированных гласных. Вместе с тем иногда наблюдается и не предсказанное в разделе 2 явление, когда, хотя резонансным частотам и соответствуют (отрицательные) экстремумы производной по частоте, но смешанная производная в этих точках положительна.

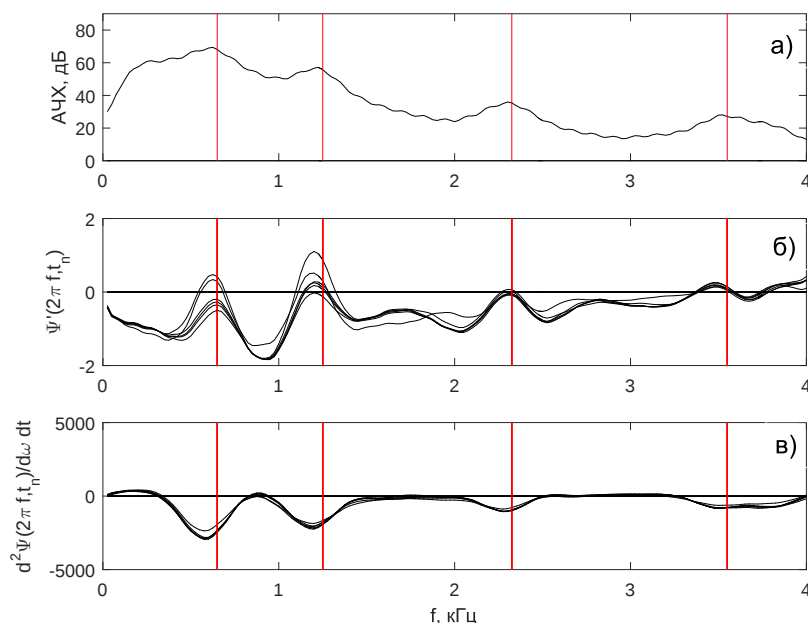


Рис. 2. Амплитудно-частотная характеристика (а); производная фазы по частоте $\Psi'(2\pi f, t_n)$ (б); смешанная производная $\frac{\partial^2 \Psi}{\partial \omega \partial t}(2\pi f, t_n)$ (в).

Синтезированные гласные удовлетворяют предположению о фиксированных резонансных частотах. Но в реальном речевом сигнале разные формантные частоты непрерывно изменяются с разной скоростью в силу изменения граничных условий со стороны голосовой щели и артикуляторных движений. При этом одна частота может практически оставаться неизменной, тогда как другая изменяется с большой скоростью. Это демонстрируется сонограммой слова /два/ на рис. 3. Поэтому, в соответствии с теоремой 2, экстремумы и точки перегиба производной фазы могут определять формантные частоты только на интервалах их относительного постоянства.

Сказанное можно проиллюстрировать: на рис. 4 и 5 стационарные точки хорошо приближают форманты. Однако, на рис. 4, в отличие от синтетического гласного, отрицательный пик смешанной производной на сегменте гласного этого слова лишь один-два раза совпадает с формантными частотами. Условие отрицательности смешанной производной при этом обычно выполняется. На рис. 5 показан случай, когда первой форманте соответствует точка перегиба производной фазы, второй и третьей форманте - минимумы производной фазы, а четвертой - ее максимум.

4. ОБСУЖДЕНИЕ

Связь амплитудно-частотных и фазово-частотных характеристик сигналов различного рода исследовалась и ранее. Например, в [6, с. 792] было установлено, что для некоторых линейных систем, описываемых дифференциальными уравнениями с постоянными коэффициентами, и обладающих передаточными функциями со специальными свойствами (минимальность фазы, специальное положение полюсов передаточных функций и т.д.), такая связь АЧХ и ФЧХ преобразования Фурье передаточных функций имеется. Однако, вопросы определения характеристик речевых сигналов (формант и др.) по найденным там связям "абсолютного значения и фазы" [6, с. 791] не исследовались. В книге [7] связь амплитудных и фазовых характеристик

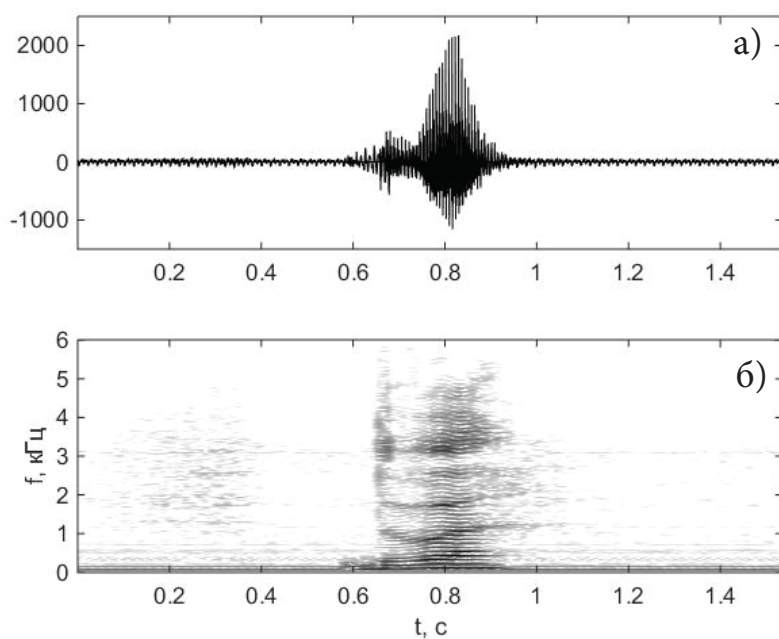


Рис. 3. Осциллограмма слова /два/ (а); его сонограмма (б).

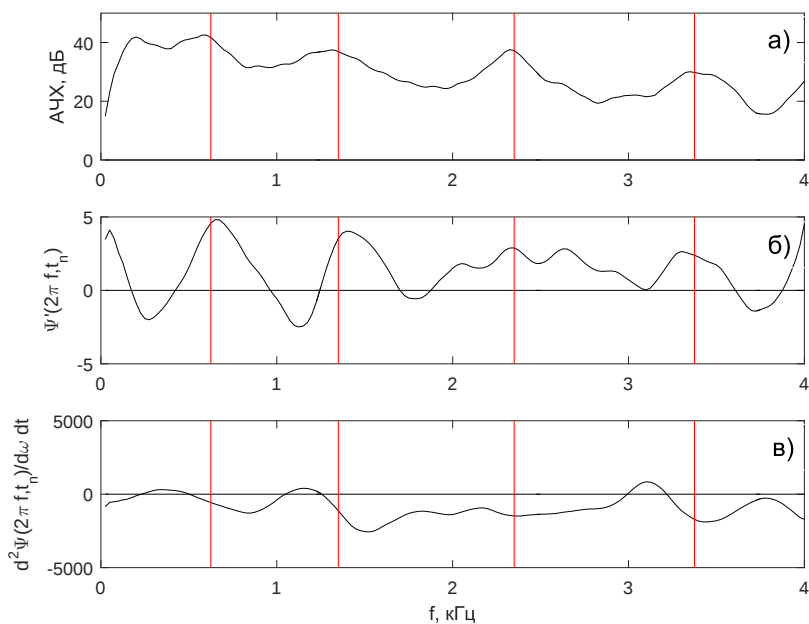


Рис. 4. Амплитудно-частотная характеристика звука /а/ (а); производная фазы по частоте (б); смешанная производная (в).

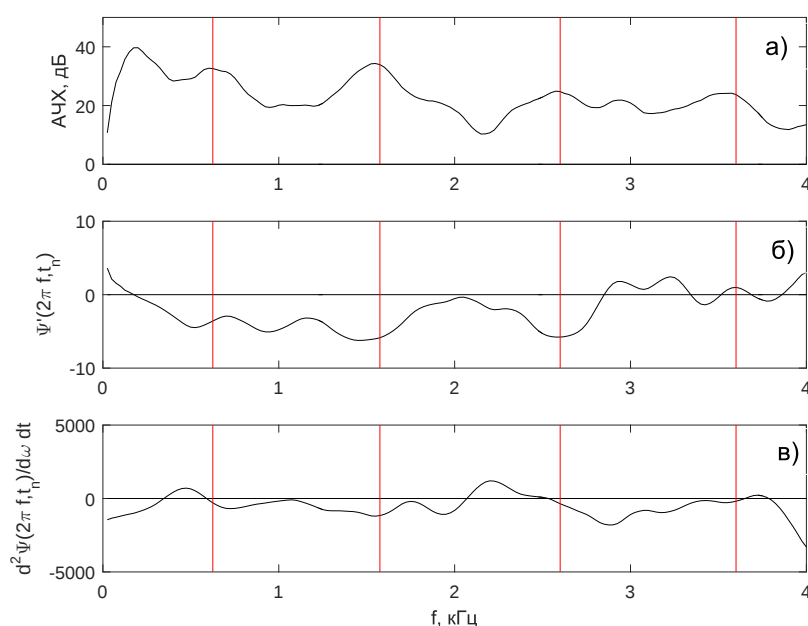


Рис. 5. Амплитудно-частотная характеристика звука /а/ (а); производная фазы по частоте (б); смешанная производная (в).

рассмотрена для так называемого аналитического сигнала. Отметим, что все эти результаты получены в частотной области и не имеют отношения к функциям типа $S(\omega, t)$, введенным выше.

Теория из раздела 2 основана на применении функции $S(\omega, t)$ – комплексной аналитической функции. В численных экспериментах роль этой функции играло кратковременное преобразование Фурье дискретного речевого сигнала. Как правило, эта комплексная функция аналитической не является. Именно поэтому важно проверить в численных экспериментах, насколько результаты теорем 1 и 2 применимы именно к речевым сигналам различной природы. Численные эксперименты показали, что эта теория может быть с определенными поправками применена для формантного анализа как синтетических, так и реальных речевых сигналов. Выполненные эксперименты позволяют уточнить и расширить условия оценки формантных частот по фазовой функции.

Например, оказалось, что фазовую характеристику функции $S(\omega, t)$ можно применять для оценки формант, если анализировать ее поведение на интервалах времени, близких к началу действия голосового источника. Отметим, что анализ формант другими методами часто проводится именно на таких интервалах (см. например [9]), и этот прием называется "синхронным анализом". Таким образом, методику синхронного анализа можно дополнить исследованием стационарных точек функции $\Psi'(\omega, t)$.

Далее, в проведенных численных экспериментах была уточнена связь нахождения формант по функции $\Psi'(\omega, t)$ с условием отрицательности смешанной производной. Оказалось, что часто наилучшие моменты для оценки формантных частот приходятся на положения отрицательных экстремумов смешанной производной во времени. Для реальных речевых сигналов теоретически найденные условия оценки формантных частот, как стационарных точек производной $\Psi'(\omega, t)$ (теорема 1), соблюдаются довольно редко. Это объясняется изменением резонансных частот речевого тракта во времени. Последнее вызвано, с одной стороны, изменением граничных условий на голосовой щели при колебаниях голосовых складок, а с другой

стороны – артикуляторными движениями. Кроме того, как и для синтетической речи, для реальных речевых сигналов, имеющих дискретный характер, используемая в анализе функции $S(\omega, t)$, как правило, аналитической не является. В этой связи, по результатам численных экспериментов, использование смешанной производной при формантном анализе реальной речи, представляется более перспективным, чем анализ стационарных точек функции $\Psi'(\omega, t)$.

На традиционной (амплитудной) сонограмме, представляющей динамический спектр мощности, формантам соответствуют пики амплитудного спектра. В результате выполненного в данном работе математического исследования и экспериментов с реальными сигналами установлено, что, в отличие от амплитудной сонограммы, на фазовой сонограмме формантам могут соответствовать не только пики, но и минимумы фазовой производной, и даже ее перегибы. Это требует пересмотра подхода к фазовому анализу, как самостоятельному процессу, и диктует необходимость совместного анализа амплитудного и фазового спектров.

5. ЗАКЛЮЧЕНИЕ

Найденные математические свойства комплексной аналитической функции $S(\omega, t)$ получаемой по сигналу, которые связывают амплитудные и фазовые характеристики сигнала, приводят к конструктивному алгоритму поиска формантных частот гласных речевых сигналов. Эти частоты могут быть определены как положения экстремумов или точек перегиба производной фазы по частоте на таких интервалах времени, где форманты не меняются и где смешанная частная производная меньше нуля. Численные эксперименты позволили уточнить и развить эти алгоритмы при формантном анализе синтетических и реальных речевых сигналов.

СПИСОК ЛИТЕРАТУРЫ

1. Маркел Дж.Д., Грей А.Х. Линейное предсказание речи. М.: Связь, 1980.
2. Yegnanarayana B., Murthy H. Significance of group delay functions in spectrum estimation. IEEE Trans. Signal Process, 1992, vol. 40, no 9, pp. 2281–2289.
3. Stark A.P., Paliwal K.K. Speech analysis using instantaneous frequency deviation. Proc. ISCA Interspeech. Brisbane, Australia: ISCA Archive, 2008, pp. 22–26.
4. Stark A.P., Paliwal K.K. Group-delay-deviation based spectral analysis of speech. Proc. ISCA Interspeech. Brighton, United Kingdom: ISCA, 2009, pp. 1083–1086.
5. Murthy H.A., Yegnanarayana B. Group delay functions and its applications in speech technology. Sadhana, 2011, vol. 36, Part 5, pp. 745–782.
6. Oppenheim A.V., Schaffer R.W., Buck J.R. Discrete-Time Signal Processing. New Jersey: Prentice Hall, 1999.
7. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989.
8. Сорокин В.Н., Леонов А.С. Фазовый анализ активности голосового источника. Акустический ж., 2021, том 67, № 2, стр. 185–202.
9. Леонов А.С., Макаров И.С., Сорокин В.Н. Частотные модуляции в речевом сигнале. Акустический ж., 2009, том 55, № 6, стр. 809–821.

Formant analysis of a speech signal in the phase domain

Leonov A.S., Sorokin V.N.

The mathematical properties of a certain complex analytical function of frequency and time, obtained from the signal, which relate the amplitude and phase characteristics of the signal, are investigated. It was

found that if the eigen-frequencies of the signal generator do not change in time, they can be determined from the extrema or inflection points of the phase derivative with respect to frequency, provided that the mixed derivative with respect to frequency and time is negative. Numerical experiments are described to determine the formant frequencies of synthetic and real speech signals using the developed theory.

KEYWORDS: formant frequencies, analytical signal, speech signal phase.