

РОЛЬ ФАЗОВЫХ СООТНОШЕНИЙ В ВОСПРИЯТИИ РЕЧИ

М. А. Сапожков

Из ряда работ [1, 2] известно, что при синфазном возбуждении параллельно включенных контуров формантного синтезатора получаемые звонкие звуки речи имеют резкое дребезжащее тональное звучание. Если же сигналы возбуждения подавать на контуры в расфазированном виде [1, 3], то синтезируемые звонкие звуки становятся более близкими по звучанию к их натуральным прототипам. Если к тому же заставить флукутировать период сигнала возбуждения [1, 4], то исчезает резкая тональность звучания и синтезируемые звуки речи становятся мало отличающимися от натуральных звуков. Поскольку, согласно закону Гельмгольца [5] *, слух реагирует только на частоты и амплитуды составляющих сложного колебания и не воспринимает фазу колебания, то на основании сказанного выше формулировалось предположение о несоответствии этого закона процессу восприятия ряда сигналов, в том числе и речевого сигнала. В связи с этим нами был проведен ряд экспериментов с синтезированными тональными и речевыми сигналами, чтобы определить применимость закона Гельмгольца к восприятию речи.

При экспериментах качество звучания оценивали эксперты-аудиторы на слух. При этом они прослушивали сигналы через динамические телефоны с равномерной частотной характеристикой в широком (для речи) диапазоне частот (до 7000 гц).

Исследуемый тональный сигнал синтезировался с помощью специального гармонического генератора на 12 гармоник. Частота основного тона сигнала изменялась в пределах 100–300 гц, амплитуды составляющих подбирались в желаемых соотношениях, сдвиги фаз (относительное расположение составляющих во временном процессе) также подбирались по желанию. Уровни сигналов контролировались с помощью вольтметра на входе телефона. Для определения субъективных комбинационных составляющих использовался метод зондирующего (пробного) тона, вводимого в телефоны.

1-й эксперимент. К телефонам попеременно подводились синтезированные сигналы, состоящие из шести гармоник нечетного порядка:

$$U_1 = \sum_{k=0}^5 \frac{1}{2k+1} \sin(2k+1)\omega t$$

и

$$U_2 = \sum_{k=0}^5 \frac{(-1)^k}{2k+1} \sin(2k+1)\omega t.$$

На фиг. 1 дана форма колебаний обоих сигналов.

При низких (на слух) уровнях громкости оба колебания оценивались как идентичные. При повышении уровня (ориентировочно до 50–60 дб) стала ощущаться на слух небольшая разница в звучании обоих сигналов. Измерения с зондирующим тоном показали, что во втором сигнале появился субъективный разностный тон с частотой 2ω . У первого сигнала такой тон появился только при повышении его уровня еще на 6 дб. По мере дальнейшего повышения уровня сигналов разница в их восприятии, хотя и оставалась заметной, но постепенно уменьшалась. Оба сигнала становились дребезжащими.

2-й эксперимент. К телефонам подводился сигнал, состоящий из четырех составляющих:

$$U = \sum_{k=0}^3 \cos(3k+2)\omega t.$$

Форма этого колебания дана на фиг. 2.

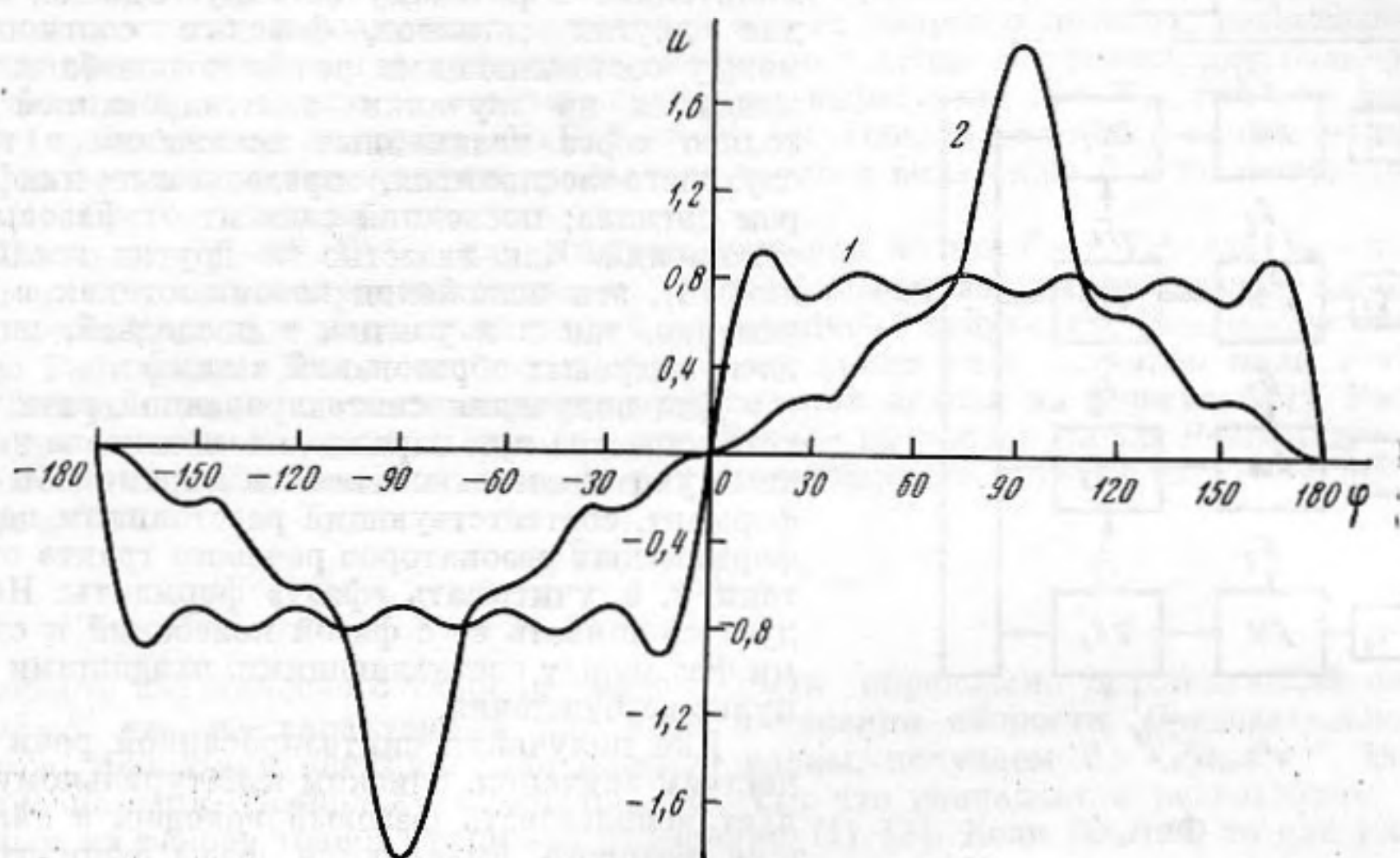
При повышении уровня этого сигнала до такого значения, при котором зондирующий тон показывал появление разностного тона с частотой 3ω , проводилось изменение фаз между составляющими. Замечено при этом, что субъективный разностный тон исчезал, когда «пикфактор» сигнала уменьшался (пикфактор определялся как отношение пикового значения к среднему квадратичному). При этом на слух также обнаруживалась разница в звучании: при появлении разностной составляющей прослушивались хрипы.

3-й эксперимент. К контурам формантного синтезатора подводились острые импульсы возбуждения, показанные на фиг. 3. Синтезируемую речь ** эксперты оцени-

* Среди ряда физиологов он носит название закона Ома.

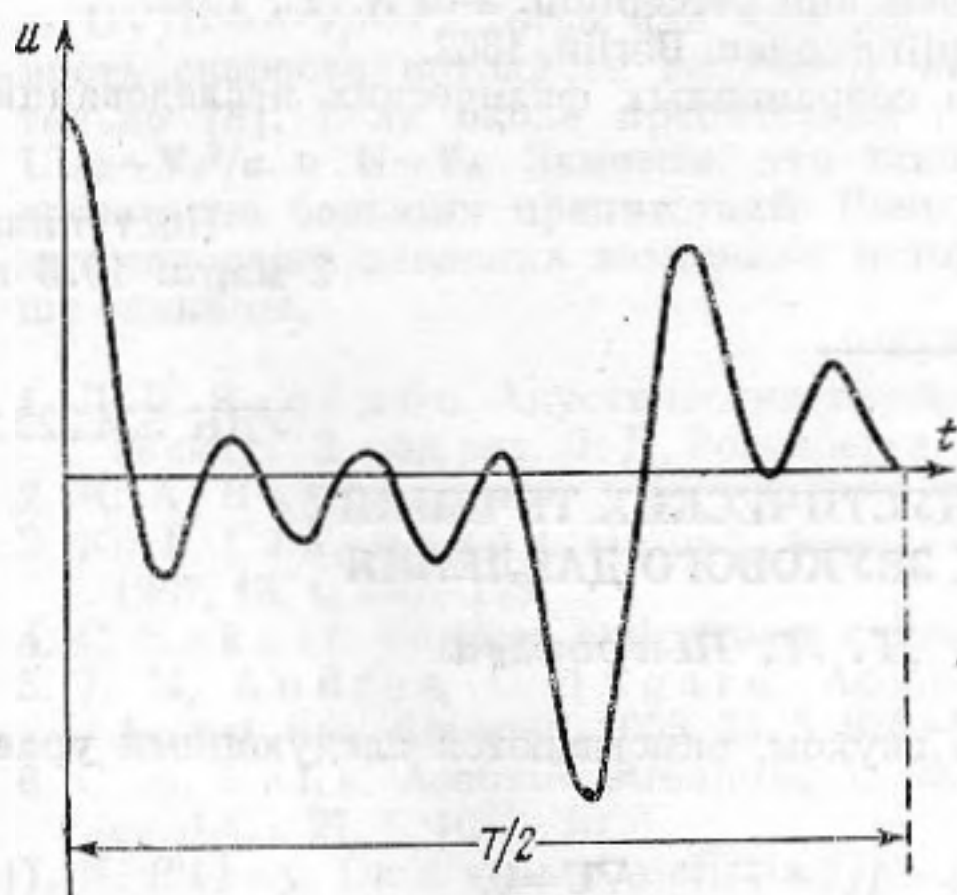
** Под речью подразумевается передача специальных артикуляционных слогов.

вали как «вокодерную». Затем в формантные каналы синтезатора, состоящие из амплитудных модуляторов АМ и формантных контуров с переменной настройкой ФК, были включены временные задержки τ (см. фиг. 4). Задержки подбирались разной величины для каждого канала, но максимальная из них не превышала 0,5 мсек. Это

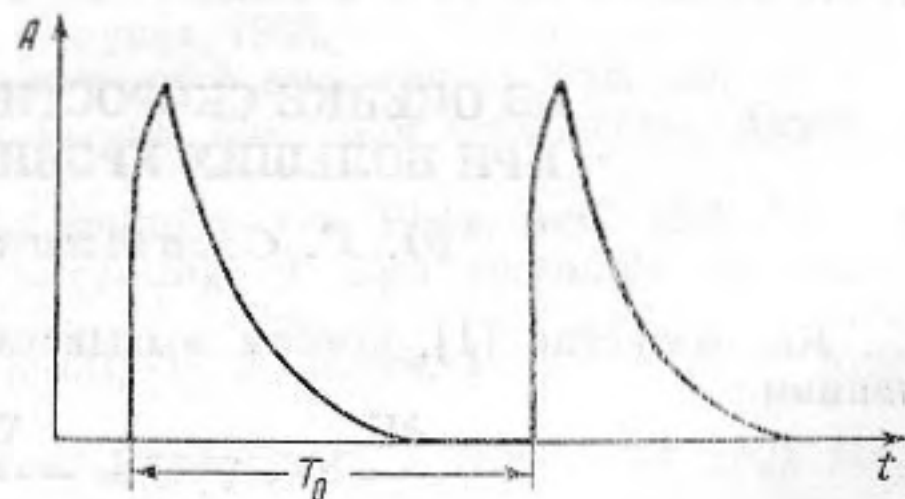


Фиг. 1

соответствует полной длине речевого тракта: от гортани до отверстия рта. Уже небольшие величины задержек (0,1 мсек) приводили к значительному улучшению звучания синтезированной речи. При этом пикфактор, определенный по осциллограммам, уменьшался в несколько раз. Различные комбинации величин задержек и перестановка последовательности порядка возбуждения контуров, хотя и приводили к изменению в звучании речевого сигнала, но эти изменения были гораздо меньше, чем первичное изменение. Кстати, в этих случаях и пикфактор сигнала изменялся несуще-



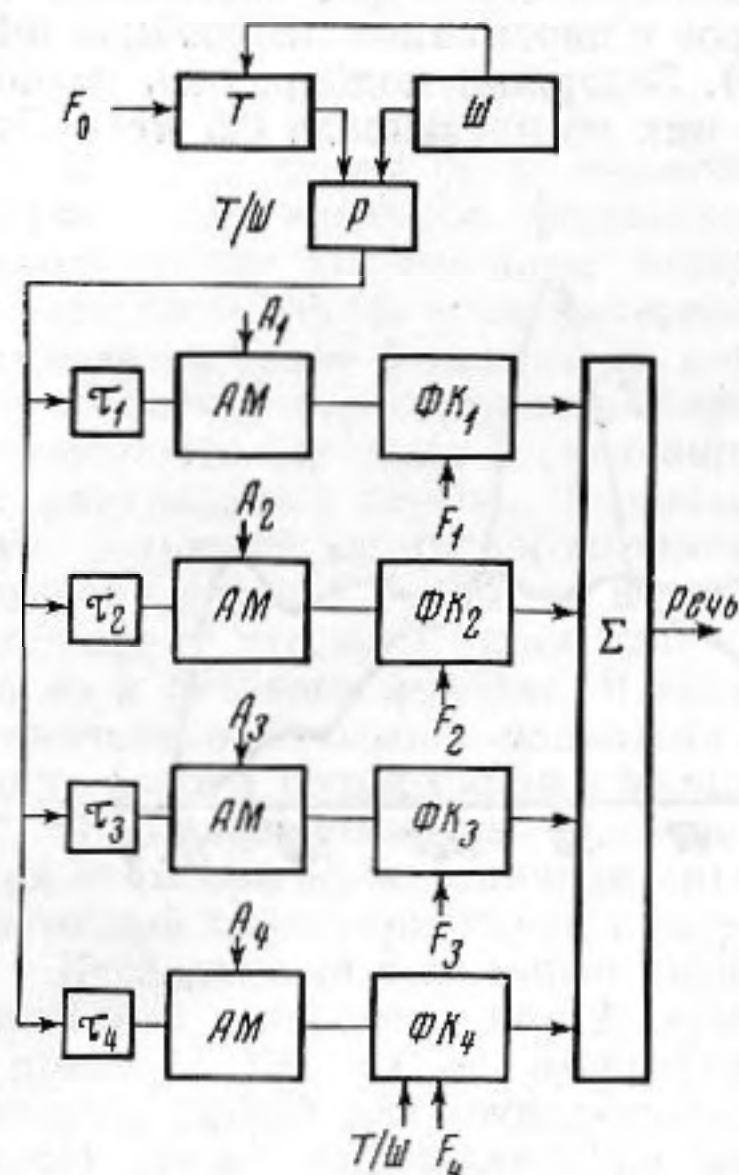
Фиг. 2



Фиг. 3

ственно. Исключением из этого правила была такая комбинация временных задержек, при которой формантные контура возбуждались через такие же интервалы времени, как и соответствующие им резонаторы речевого тракта. В этом случае получалось заметное повышение качества звучания синтезированной речи, хотя субъективные нелинейные искажения изменялись несущественно, так как пикфактор сигнала оставался примерно на прежнем уровне.

4-й эксперимент. К контурам формантного синтезатора подводились импульсы возбуждения типа гортанных с флуктуацией периода импульсов. Синтезированная



Фиг. 4

синтезаторе они вводятся соответственно методу Фланагана [4, 1]. Рекомендуется использовать генератор основного тона голоса T (см. фиг. 4) с флуктуирующим периодом, получающимся с помощью шумового генератора Ш. При таких флуктуациях существенно уменьшается пикфактор сигнала и, следовательно, нелинейные искажения.

ЛИТЕРАТУРА

1. М. А. Сапожков. О некоторых путях улучшения качества синтезируемой речи. Акуст. ж., 1971, 17, 4, 605–609.
2. Дж. Фланаган. Анализ, синтез и восприятие речи. М., «Связь», 1968.
3. А. А. Пирогов. Синтетическая телефония. М., Связьиздат, 1963.
4. J. L. Flanagan. Speech analysis, synthesis and perception. 2-d, N. Y., 1972.
5. Н. в. Helmholtz. Die Lehre von Tonempfindungen, Berlin, 1862.
6. С. Н. Ржевкин. Слух и речь в свете современных физических исследований. М., ОНТИ НКПТ СССР, 1936.

Московский электротехнический институт связи

Поступила
2 марта 1973 г.

УДК 534.222.2

ОБ ОЦЕНКЕ СКОРОСТИ АКУСТИЧЕСКИХ ТЕЧЕНИЙ ПРИ БОЛЬШИХ УРОВНЯХ ЗВУКОВОГО ДАВЛЕНИЯ

Ю. Г. Статников, Н. Л. Широкова

Как известно [1], потоки, вызываемые звуком, описываются следующими уравнениями:

$$\frac{\partial U}{\partial t} + (U \nabla) U = - \frac{\nabla p}{\rho_0} + \nu \Delta U + F, \quad \nabla U = 0, \quad (1)$$

где U — скорость потока, p и ρ_0 — постоянные составляющие давления и плотности среды, ν — кинематическая вязкость среды, F — сила, вызывающая поток:

$$F = - (\mathbf{V}_A \nabla) \mathbf{V}_A + \frac{1}{\rho_0^2} [\rho_A \nabla p_A - \eta \rho_A \Delta \mathbf{V}_A], \quad (2)$$

где \mathbf{V}_A , ρ_A , p_A — переменные составляющие скорости, плотности и давления, которые могут быть определены на основе уравнений акустики:

$$\frac{\partial \mathbf{V}_A}{\partial t} + (\mathbf{V}_A \nabla) \mathbf{V}_A = - \frac{\nabla p_A}{\rho_0} + \nu \Delta \mathbf{V}_A, \quad \frac{\partial \rho_A}{\partial t} + \rho_0 \nabla \mathbf{V}_A = 0.$$