

КОНФЕРЕНЦИИ  
И СОВЕЩАНИЯ

УДК 534.78

НЕКОТОРЫЕ ПРОБЛЕМЫ АКУСТИЧЕСКОЙ ТЕОРИИ  
РЕЧЕОБРАЗОВАНИЯ

© 2002 г. В. И. Галунов

Санкт-Петербургский государственный университет  
ООО "Одитек"  
198904 С.-Петербург, Ст. Петергоф, ул. Суворовская 7/2-13  
E-mail: auditech@online.ru

Классическая акустическая теория речеобразования в своем основном виде остается неизменной со времен ее создания Гельмгольцем в 1870 г. В то же время имеет смысл рассмотреть целый комплекс вопросов, которые в настоящий момент вызывают сомнения.

Акустическая теория речеобразования в ее научном оформлении впервые была сформулирована Гельмгольцем в 1870 г. [1]. Основная идеология этой работы осталась без изменения до настоящего времени и в явном или неявном виде принимается большинством речевиков. При этом учитывается, конечно, целый ряд математических и методико-технологических усовершенствований, введенных начиная от забытой по естественным причинам работы Чибо и Каджияма [2] и кончая классическими трудами Фанта [3] Унгехойера [4] и последними работами [5, 6].

Вероятно следует указать две основные особенности модели Гельмгольца.

1. Процесс речеобразования состоит из двух независимых компонент: возбуждения как такового звука и формирования фонетического качества звука за счет возбуждения резонансных частот артикуляционного тракта (у Гельмгольца) или фильтрации (в современном рассмотрении).

2. Фонетическое качество звука определяется так называемыми формантами, которые определяются как резонансные частоты артикуляционного тракта (или же полюса передаточной функции артикуляционного фильтра) или как максимумы спектра речевого сигнала [7].

Указанные особенности в явном виде не подвергаются пересмотру, хотя, как всем очевидно, они не соответствуют действительности. Связано это с тем, что не видно, к каким научным последствиям такой пересмотр должен привести.

Коснемся прежде всего независимости источника звука и артикуляционного фильтра. В работе [8] было показано, что фонетическое качество гласных в значительной степени сформировано уже в гортани при исключении влияния артикуляционного фильтра. Более того, в классической литературе хорошо известно, что голосовой ис-

точник обладает своими полюсами и нулями, что в явном виде сказывается на качестве формируемого речевого сигнала. Это все заставляет принять предположение о том, что если существуют определяющие фонетическое качество форманты, то они являются максимумами в спектре, а не требующими операции деконволюции полюсами передаточной функции, как предполагает классическая теория.

Второй вопрос – являются ли процессы речеобразования и речевосприятия симметричными, как это предполагается в классической теории. То, что с помощью таких формант можно добиться определенного фонетического качества звучания, не вызывает сомнения. (Прекрасный пример – речь птиц). Но определяют ли именно эти максимумы спектра это качество? Первые сомнения были внесены еще в 30-х годах после создания полосного вокодера [9]. В начале 60-х годов была сформирована на основе большого экспериментального материала теория расчета разборчивости речи, принявшая за основу полосное представление речевого сигнала [10, 11], исключавшая из рассмотрения форманты (при этом следует отметить, что в русскоязычной версии этой теории за основным расчетным параметром сохранялось название "формантная разборчивость"). Все это заставило вести гипотезу (Варшавский Л.А. и Литвак И.М.) о том, что фонетическое качество звуков определяется определенным уровнем соотношений мощности в спектральных полосах, а форманты (т.е. максимумы в спектре) являются лишь доступным для речеобразующего аппарата способом достижения необходимых полосных соотношений.

Опираясь на изложенную выше двойную идеологию принципов формирования акустического облика речевых звуков на фонетическом уровне,



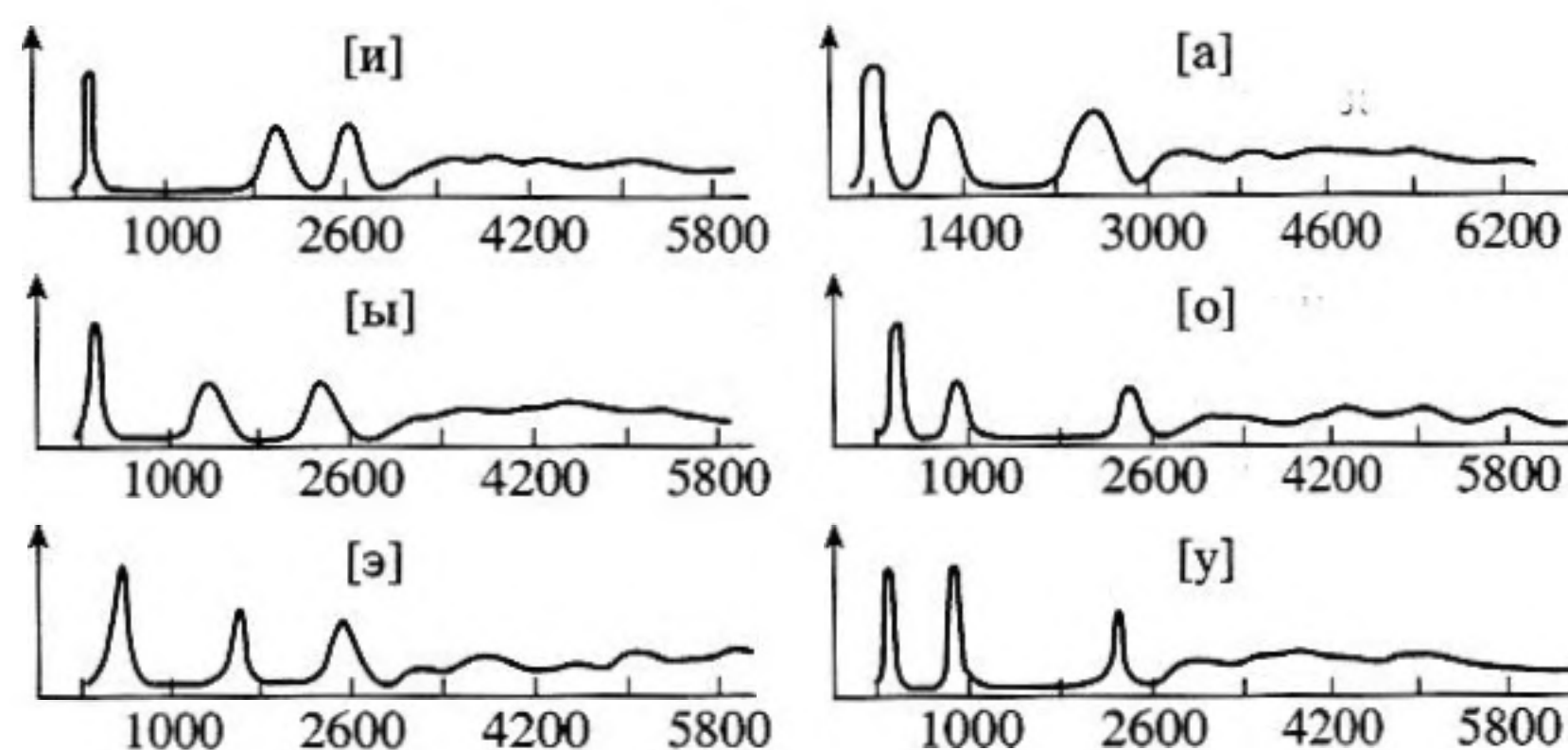


Рисунок.

естественно предположить, что существует несколько параллельно функционирующих систем звукообразующих признаков. Именно за счет существования нескольких систем признаков различного типа обеспечивается устойчивость речи как коммуникативной системы по отношению к воздействию в довольно широком диапазоне помех, шумов и искажений.

Следующий вопрос – сколько формант существует? Известно расположение формант для русских гласных, произносимых в изолированной позиции или в слогах. Каких либо заметных отклонений от результатов, полученных еще в классическом труде Г. Фанта [3] практически нет. Однако известно, что в реальном речевом сигнале картина представления формант заметно изменяется. Наблюдается появление лишних максимумов, расщепление максимумов на месте расположения форманты в идеальном варианте, исчезновение классических формантных максимумов. Указанные ненормативные отклонения могут проявляться в результате влияния по крайней мере 3-х факторов:

- индивидуальные особенности говорящего,
- контекстуальное окружение,
- ситуационные особенности (психо-физиологическое состояние диктора, манера произношения в данной аудитории и т.п.).

Если принять гипотезу о том, что артикуляторный аппарат обладает конечными акустическими резонансными и фильтровыми возможностями, но при этом человек при формировании речевых звуков стремится к определенной структуре различительных признаков, одним из которых является определенным образом организованная система формант, то несмотря на все перечисленные выше возможности искажения “правильной” формантной структуры на гистограммах вероятности распределения спектральных максимумов будут выделяться зоны упомянутых “правильных” формантных структур.

В работе [12] было показано, что для всех гласных в низкочастотной части спектра на распределении вероятностей появления спектральных максимумов наблюдается три четко выраженных максимума, соответствующих трем первым формантам (см. рис.). В промежутках и в высокочастотной области распределение вероятности появления максимума имеет практически равномерный характер и выделение формантных областей невозможно. Указанные формантные зоны при этом соответствуют принятым в научном обиходе положениям 3-х первых формант русских гласных. Единственно следует указать, что в слитной речи для мужских голосов часто появляется дополнительный спектральный максимум между 2-ой и 3-ей формантой. Для женских же голосов часто происходит исчезновение 2-ой форманты. (Это явление еще более характерно для детских голосов).

В связи с проблемой значимости формант для восприятия следует указать на еще одну проблему, которая в явном виде не обсуждается. Известно, что в реальной речи основную роль для формирования лингвистического образа сказанного играют переходы между звуками. Стационарные участки даже могут переставляться без ущерба для восприятия смысла. Непонятно, как это соотносится с явным значением стационарных участков, с их формантной структурой при восприятии изолированных гласных или слогов.

Рассмотрим теперь возможности полосного представления речевого сигнала. Очевидно, что из-за чисто физических ограничений человек в процессе речеобразования не может управлять большим числом спектральных составляющих. Это следует из физических ограничений возможностей артикуляционного тракта [13]. Если мы в данный момент измерили даже большое количество спектральных составляющих, то во времени они смогут изменяться только коррелированно. Независимые спектральные составляющие могут



быть обнаружены путем анализа матриц корреляции временных огибающих составляющих спектра. Естественно предположить, что получаемые таким образом независимые составляющие и являются именно теми существенными переменными, которые определяют качество сигнала. В работе [12] с помощью факторного анализа матриц корреляции спектральных составляющих речевого материала были получены следующие приблизительно независимые спектральные полосы: 80–400, 400–750, 750–1350, 1350–1750, 1750–2200, 2200–2900, 2900–5000 Гц. Выделенные составляющие хорошо совпадают с границами формантных диапазонов (интересно, что выделяется полоса 1350–1750 Гц, не соответствующая каким-либо формантам для русских гласных, но дающая часто ложную форманту для мужских голосов – см. предыдущий раздел).

Возвратимся к традиционной схеме. Основное предположение – информационная звуковая волна формируется в речевом тракте за счет возбуждения собственных колебаний звукового волновода, каковым считается речевой тракт. Из открытого конца волновода волны распространяются в упругой среде и в конечном счете возбуждают собственные колебания волновода, представляющего слуховой тракт. Эти собственные колебания анализируются мозгом и воспринимаются человеком как речь. С принципиальной точки зрения подобная схема, разумеется, может существовать, и у нее к настоящему времени немало успехов. Однако есть моменты, которые, с нашей точки зрения, плохо согласуются с реальностью. Во-первых, реальные спектры одних и тех же сигналов от разных индукторов часто иные, что коренным образом противоречит традиционной схеме. Тем более не ясно, что происходит в случае речи при сильном изменении артикуляционного тракта (например, с сигаретой во рту). Более того, при прохождении через телефонный канал речевого сигнала его спектр значительно искажается, но не происходит искажения смыслового содержания. Уже одного этого было бы достаточно, что бы подвергнуть сомнению адекватность традиционной модели речеобразования. Есть еще и другие сомнения, которые, однако, могут быть отнесены к числу эмоциональных. Зачем природе понадобилась такая сложная система? Действительно, для формирования определенного временного хода на выходе речевого тракта необходимо не просто возбудить конкретные моды волновода, но возбудить их с определенными фазовыми и интенсивностными соотношениями. При этом совершенно непонятна роль значительных пространственных деформаций речевого тракта в процессе речеобразования.

По крайней мере, эти вопросы снимаются в другой модели, которую будем называть модуля-

ционной моделью речеобразования (ММР). Здесь можно условно выделить два этапа. На первом этапе возникает звуковая волна, не содержащая в себе никакой информации и играющая роль несущей. На втором этапе несущая модулируется, и эта модуляция и содержит всю информацию о речи. В этой модели роль речевого тракта сводится к роли модулятора, и потому становятся естественными значительные пространственные деформации тракта. Снимается вопрос о строгих фазово-интенсивностных соотношениях между собственными колебаниями, так как возбуждаемая волна может носить (и носит, по нашему мнению) случайный характер. Кстати, это позволяет объяснить и несовпадение спектров одних и тех же речевых сигналов, а следовательно, и особенности прохождения через телефонные кабели. Одна из возможных схем модуляции нами предлагается в другом докладе [14].

#### СПИСОК ЛИТЕРАТУРЫ

1. *Helmholtz H.* Die Lehre von der Tonempfindungen als physiologische Graudlage fur die Theorie der Musik, Brounshweig, 1870.
2. *Chiba T., Kajima M.* The vowel, its nature and structure. Tokyo, 1941.
3. *Фант Г.* Акустическая теория речеобразования. М.: Наука, 1964.
4. *Ungeheuer G.* Elemente einer akustischen Theorie der Vokalartikulation. Berlin: Springer, 1962.
5. *Kent R.D. et al.* (Eds) Papers in Speech Communication: Speech production. Ac. Soc. of America, 1991.
6. *Сорокин В.Н.* Теория речеобразования. М.: Радио и связь, 1985.
7. *Фланаган Дж.Л.* Анализ, синтез и восприятие речи. Связь. М., 1968.
8. *Галунов В.И., Крылов Б.С., Станкевич С.А., Хантемиров Р.Г.* Исследование аэродинамических процессов в гортани. III съезд отоларингологов РСФСР, 1972.
9. *Dudley H.* The Vocoder. Bell Labs. Record 17. 122–126. 1939.
10. *Kryter K.D.* Methods for the calculation and use of the articulation index. JASA 34. 1689–1697 (1962).
11. *Покровский Н.Б.* Расчет и измерение разборчивости речи. Связьиздат, 1962.
12. *Галунов В.И., Гарбарук В.И.* Акустическая теория речеобразования и система фонетических признаков. Материалы международной конференции: 100 лет русской экспериментальной фонетики. СПб., 2001.
13. *Галунов В.И.* Исследование вариативности речевого поведения человека, докт. дисс., 1975.
14. *Галунов В.И., Уваров В.К.* Еще раз о механизме голосообразования. XI сессия Российского акустического общества. М., 2001. Т. 3. С. 17–20.



## Some Problems of the Acoustic Theory of Speech Production

V. I. Galunov

*ООО Одитек, St. Petersburg State University,  
ul. Suvorovskaya 7/2-13, Peterhof, St. Petersburg, 198904 Russia  
e-mail: auditech@online.ru*

**Abstract**—The classical acoustic theory of speech production remains basically intact since 1870 when it was first developed by Helmholtz. However, there are some questions that require special consideration, because the answers to them are still unclear.