

УДК 81.342

**Кринов С.Н., Брянцева Т.Н., Хмелев А.К.**

*Московский государственный областной университет*

## ВЛИЯНИЕ ВОКАЛИЧЕСКОГО КОНТЕКСТА НА ИДЕНТИФИКАЦИЮ ГЛАСНОГО ЗВУКА

*Аннотация.* В статье описывается механизм контекстного влияния на восприятие гласного звука. Обсуждаются известные эксперименты по восприятию английских гласных синтезированной речи, демонстрирующих изменение идентификации тестового слова в зависимости от формантной структуры вводного предложения. Приводится новая интерпретация результатов обсуждаемого эксперимента. Демонстрируется эксперимент по изменению категориального восприятия гласного в контексте естественной речи.

*Ключевые слова:* восприятие гласных звуков, категория гласного звука, фонетический контекст, фонетическое пространство, формантные частоты гласных звуков.

**S. Krinov, T. Bryantseva, A. Khmelev**

*Moscow State Regional University*

## THE VOCALIC CONTEXT INFLUENCE ON THE VOWEL SOUND IDENTIFICATION

*Abstract.* The mechanism of contextual influence on perception of a vowel sound is described. Known experiments on perception of the English vowels of synthetic speech are discussed, especially those showing the change of the test word identification depending on the formant structure of an introductory sentence. New interpretation of the discussed experiments results is given. An experiment on changing the categorical perception of a vowel in the context of natural speech is shown.

*Key words:* perception of vowel sounds, category of a vowel, phonetic context, phonetic space, formant frequencies of vowel sounds.

Настоящая работа ограничивается рассмотрением одного из механизмов сложного процесса классификации гласных звуков, связанного с окружающим вокалическим контекстом.

Известно, что качественные различия между гласными звуками определяются взаимным расположением частот первых трёх формант, из которых для восприятия наибольшую значимость имеют первые две форманты

( $F_1$  и  $F_2$ ). При этом в сочетаниях с согласными формантная картина подвергается систематическим контекстным изменениям [1, с. 160–161]. Впервые распределение английских гласных звуков в пространстве  $F_1$ ,  $F_2$  было продемонстрировано, как только появился один из первых спектрографов [11]. Поскольку абсолютные значения формантных частот зависят от многих факторов, в том числе от размеров речевого тракта говорящего, то их следует рассматривать как ориентиры.

Результаты первых экспериментов по влиянию вокалического контекста на восприятие английского гласного в тестовом слове были опубликованы в 1957 г. Ладефогедом [10]. В этих экспериментах использовался параметрический синтезатор речи с возможностью управления формантными частотами. Для обсуждения полученных в этой работе результатов и их новой интерпретации приведём краткое описание эксперимента и результатов.

Были синтезированы шесть вариантов вводной фразы: Please say what this word is. Диапазоны изменения формантных частот в этих вариантах фразы приведены в таблице 1.

Таблица 1

Частотные характеристики формант шести вариантов вводной фразы

Номер варианта	Отличие от варианта 1	Частотный диапазон, Гц	
		F1	F2
1	...	275–500	600–2500
2	F1 смещение вниз	200–380	600–2500
3	F1 смещение вверх	380–660	600–2500
4	F2 смещение вниз	275–500	400–2100
5	F2 смещение вверх	275–500	800–2900
6	F1 смещение вниз F2 смещение вверх	200–380	800–2900

В дополнение к этим вводным фразам были синтезированы четыре односложных тестовых слова *b-(гласный)-t*. Были составлены 11 комбинаций «вводная фраза – тестовое слово».

В таблице 2 приведены значения формантных частот для четырёх синтезированных слов.

Таблица 2

Частотные характеристики формант четырёх тестовых слов

Тестовое слово	Частота, Гц	
	F1	F2
A	375	1700
B	450	1700
C	575	1700
D	600	1300

В эксперименте принимали участие 60 слушателей, которые предварительно идентифицировали каждое из тестовых слов в изолированном звучании. Результаты, демонстрирующие неоднозначную идентификацию этих слов, приведены в таблице 3.

Таблица 3

Результаты идентификации тестовых слов слушателями

Тестовое слово	Количество слушателей, идентифицировавших слово как:			
	bit	bet	bat	but
A	52	8		
B	14	46		
C		27	33	
D		1	14	45

Прежде чем обсудить результаты прослушивания комбинаций, приведём результаты собственных исследований над двумя из 11 комбинаций.

Американское акустическое общество в честь 80-летия Ладефогда провело специальное заседание в 2005 г., на котором в одном из докладов были воспроизведены два варианта комбинаций: *Вводная фраза 2 – Тестовое слово В* и *Вводная фраза 3 – Тестовое слово В* [8]. Доклад сопровождался презентацией со звучанием соответствующих комбинаций [8]. Звук был извлечён из презентации для настоящего исследования.

На приведённом ниже рисунке 1, представляющем собой копию окна анализа звукового редактора Praat, изображены осциллограмма (вверху) и спектрограмма (ниже) извлечённой из презентации комбинации: *Вводная фраза 2* (диапазон частоты  $F_1$  200–380 Гц) – *Тестовое слово В*. Разметка произнесения и значения формантных частот в герцах для соответствующих участков также были выполнены средствами Praat. На спектрограмме (по горизонтали – время, по вертикали – частота, степень почернения – ампли-

туда) чётко видны траектории трёх формантных частот. Видно, что на некоторых гласных звуках значения формантных частот изменяются, что было отмечено в нижней части рисунка 1.

Тестовое слово в конце произнесения в эксперименте Ладефогедда было идентифицировано 57 слушателями как *bet*, одним слушателем как *bit* и двумя как *bat*. Во второй комбинации *Вводная фраза 3 – Тестовое слово В* был изменен диапазон частоты  $F_1$  смещением вверх (380–600 Гц). В этом случае 58 слушателей идентифицировали тестовое слово как *bit* и двое как *bet*.

Эксперимент с прослушиванием этих двух произнесений в рамках нашей работы дал подобные результаты идентификации, только идентификация тестового слова для второго варианта как *bit* составила только 50 % из 24 слушателей – студентов первого курса. Это объясняется тем, что английский для слушателей не является родным языком. Кроме того, у студентов первого курса нет опыта аудирования, так как у них ещё не было практических занятий по фонетике. Следует заметить, что в публикации Ладефогедда [10] приведены рисунки, где каждый гласный в пространстве  $F_1, F_2$  отмечен точкой, в том числе и для голоса Ладефогедда. Однако известно, что в силу изменчивости голоса, даже для голоса одного диктора, в указанном пространстве каждому гласному соответствует область значений [6; 12]. Кроме того, существует так называемый «магнитный эффект восприятия». В слуховом пространстве гласных звуков для слушателя существуют соот-

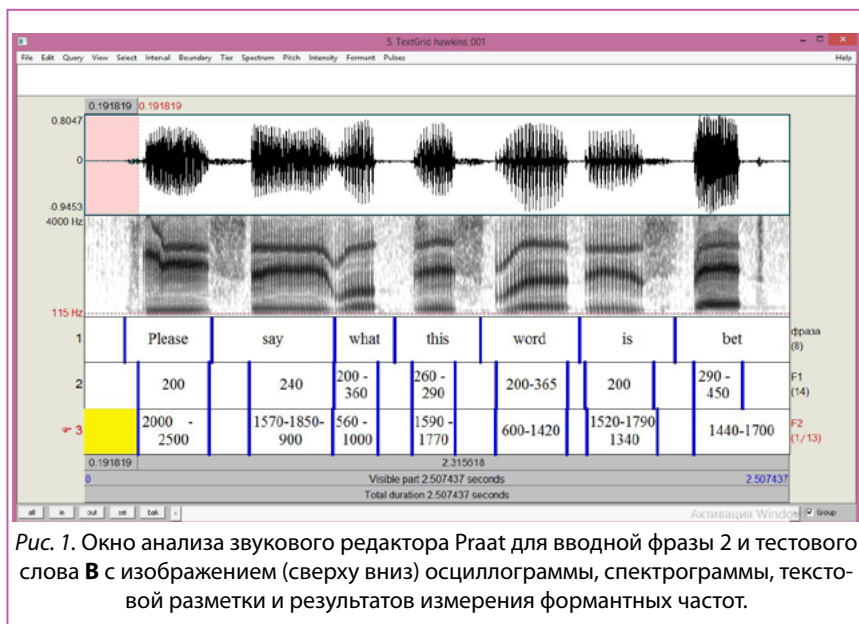


Рис. 1. Окно анализа звукового редактора Praat для вводной фразы 2 и тестового слова В с изображением (сверху вниз) осциллограммы, спектрограммы, текстовой разметки и результатов измерения формантных частот.

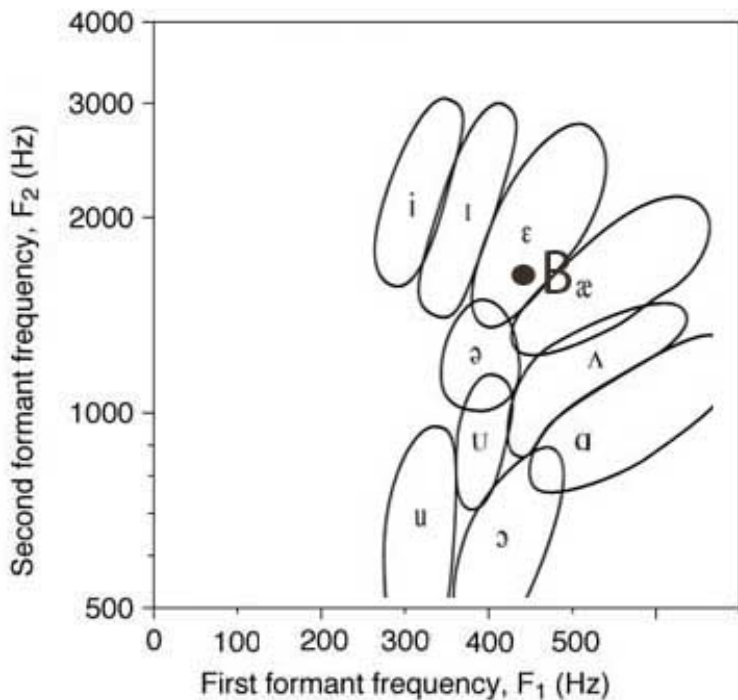


Рис. 2. Тестовое слово **В** в пространстве фонетических категорий относительно вводной фразы 2.

ветствующие гласным звукам категории, обладающие внутренней структурой [9]. Наилучший экземпляр этой структуры в слуховом пространстве действует подобно магниту. Этот эффект для гласных родного языка проявляется уже у шестимесячных младенцев. Хотя к этому времени младенец способен произносить звуки практически всех языков мира [4]. Эксперименты на обезьянах не обнаружили магнитного эффекта восприятия [9].

Учитывая этот факт, было решено смоделировать в пространстве  $F_1, F_2$  ситуацию, соответствующую двум исследуемым произнесениям, с использованием областей для английских гласных [11], приспособив их путём трансформации соответствующих областей в соответствии с вариантами 2 и 3 из таблицы 1. Трансформация в данном случае сводилась к сжатию и сдвигу областей. Результат изображен на рисунках 2 и 3.

Из рисунков видно, что неподвижный в формантном пространстве объект **В** соответствует областям разных гласных звуков. Другими словами,

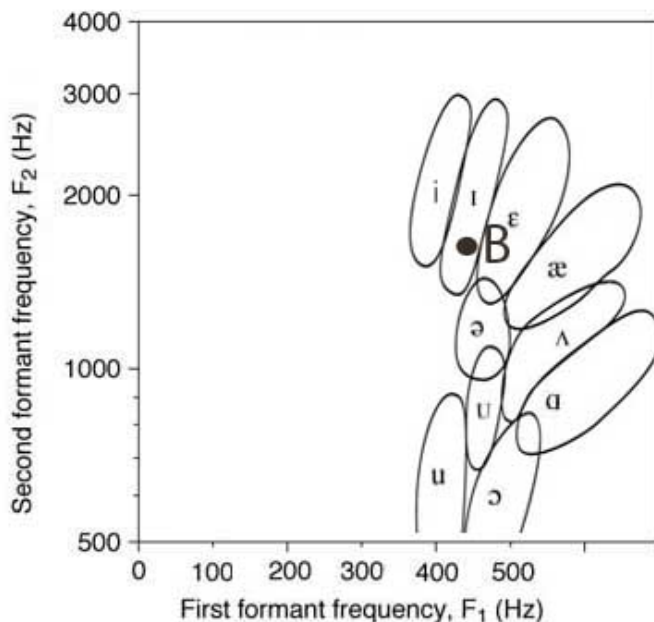


Рис. 3. Тестовое слово В в пространстве фонетических категорий относительно вводной фразы З.

пространство, в котором происходит идентификация тестового слова, есть подвижное пространство категорий, определяемых вокалическим контекстом.

Следовательно, любой гласный звук не является независимым ни от чего объектом, а существует как элемент набора гласных данного языка с их взаимным расположением в формантном пространстве. Это замечание полезно при освоении произношения на иностранном языке, а также при обучении речи глухих, поскольку размеры речевого тракта у всех разные, а пространство артикуляций напрямую связано с фонетическими категориями.

Обучаясь речи, ребёнок физически не может воспроизвести голос своей матери, т.е. воспроизвести такие же значения формантных частот. Но усваивая относительное расположение категорий в перцептивном пространстве, он учится правильно артикулировать гласные звуки.

Сохраняя взаимное расположение категорий в артикуляционном пространстве, различные голоса могут проявлять различие в количественных отношениях внутри семейства гласных. Этим характеризуется вокалический идиолект, т.е. индивидуальный вокализм человека, что может быть использовано при идентификации личности по голосу.

Рассматривая остальные комбинации из работы Ладефогед, следует обратить внимание на комбинацию *Вводная фраза 5 – Тестовое слово С*. Обсуждая результат идентификации для этого случая (48 идентификаций как *bet* и 12 – *bat*), Ладефогед пишет, что результат для него не совсем понятен. По его мнению, часть идентификаций должна соответствовать слову *but*.


Однако если посмотреть на числовые данные таблиц 2 и 3, то видно, что тестовое слово находится вне определяемого контекстом подвижного пространства, и идентификация во многом опирается на индивидуальные свойства самого тестового слова. Это касается и тестового слова D в комбинации с любой из вводных фраз.

Продолжая обсуждение, Ладефогед пишет, что необходимы дополнительные исследования для определения длины контекста, влияющего на категорию тестового слова.


Если вернуться к рисунку 1, то видно, что гласные не являются точками в пространстве категорий, а демонстрируют изменчивость. Эта изменчивость связана с коартикуляцией, а также с тем, что в слитной речи практически отсутствуют стационарно неизменные участки. И в реальности на многих звуках наблюдается континуум значений формантных параметров [5]. Переходные сегменты от согласного к гласному, дифтонги также могут служить вокалическим контекстом. В ситуации, связанной с рисунком 3, вводная фраза сокращалась до «Please» или «is». При этом тестовое слово продолжало идентифицироваться русскоязычными слушателями как *bit* в 50 % случаев.

Далее был проведён эксперимент с пересадкой звука в чужой контекст на материале естественной речи. В эксперименте использованы фрагменты интервью, размещенного на сайте радиостанции «Эхо Москвы». Был исследован мужской голос [3].


Ниже приведено звучание слова «глупости» с ударным звуком [y]\*.

Звучание № 1: 

Далее – звучание фрагмента «что-то противопоставить».

Звучание № 2: 

В последнем фрагменте первый звук [o] был заменен на [y].

Звучание № 3: 

---

\* Для прослушивания звуковых фрагментов файл статьи необходимо открыть в программе Adobe Acrobat Reader (<https://get.adobe.com/ru/reader/>).

Аудирование на нескольких десятках слушателей показывает устойчивое восприятие [y] как [o].

Интересно отметить, что в данном случае контекст не является предшествующим для идентифицируемого гласного. Значит, действие контекста может распространяться как вперёд, так и назад.

Интересно также, что контекстный сдвиг произошёл на голосе одного человека в пределах одного интервью, и расстояние между заменяемыми звуками находится в пределах двух-трёх фраз. Это говорит о том, что контекст, определяемый гласными звуками одного голоса, нестабилен. Количественные параметры этой нестабильности также могут характеризовать человека по его голосу.

В заключение можно сказать, что человечество неосознанно использует замечательный механизм подвижной системы координат, участвующий в идентификации гласных. Этот механизм позволяет нам адаптироваться к любому голосу и в ряде случаев помогает адаптироваться в ситуациях искажения речевого сигнала в некоторых пределах. Эти пределы задаются сохранением взаимного расположения алфавита гласных.

#### Литература:

1. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. М.: РГГУ, 2001. 592 с.
2. Кузнецов В.Б. Спектральная динамика и классификация русских гласных // Акустический журнал. 2002. Т. 48. № 6. С. 849–853.
3. Радиостанция «Эхо Москвы». URL: [www.echo.msk.ru](http://www.echo.msk.ru) (дата обращения: 27.05.2015)
4. Рау Ф.Ф. Устная речь глухих. М., 1973. 302 с.
5. Щерба Л.В. Субъективный и объективный метод в фонетике: Языковая система и речевая деятельность. М.: Едиториал УРСС, 2004. 432 с.
6. Acoustic characteristics of American English vowels / Hillenbrand J., Getty L.A., Clark M.J., Wheeler K. // The Journal of the Acoustical society of America. 1995. Vol. 97. № 5. Pt. 1. 3099–3110 pp.
7. Hawkins S. Betting on bits: Contextual influences on the perception of “phonetic categories” // The Journal of the Acoustical Society of America. 2005. V. 118. URL: <http://www.linguistics.ucla.edu/faciliti/PLBday/hawkins.pdf>. (дата обращения: 02.11.2010).
8. Hawkins S. Phonetic Categories – UCLA Department of Linguistics. URL: <http://www.linguistics.ucla.edu/faciliti/PLBday/hawkins.ppt> (дата обращения: 10.01.2009).
9. Kuhl P.K. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not // Perception & Psychophysics. 1991. Vol. 50(2). 93–107 pp.
10. Ladefoged P., Broadbent D.E. Information Conveyed by Vowels // The Journal of the acoustical society of America. 1957. V. 29. № 1. 98–104pp.
11. Peterson G.E., Barney H.L. Control methods used in study of vowels // The Journal of the acoustical society of America. 1952. Vol. 24. 3099–3110 pp.
12. Strange W. Evolving theories of vowel perception // The Journal of the acoustical society of America. 1989. Vol. 85. № 5. 2081–2087 pp.