

КОНФЕРЕНЦИИ
И СОВЕЩАНИЯ

УДК 612.821.8:413.41

СПЕКТРАЛЬНАЯ ДИНАМИКА
И КЛАССИФИКАЦИЯ РУССКИХ ГЛАСНЫХ

© 2002 г. В. Б. Кузнецов

Московский государственный лингвистический университет
119034 Москва, ул. Остоженка, 38
E-mail: kvlad@ccas.ru

Представлены результаты экспериментального исследования спектральных характеристик русских ударных гласных в зависимости от твердости–мягкости окружающих согласных, локального темпа речи и индивидуальности диктора.

На основе полученных данных рассматривается возможность классификации гласных при использовании статических и динамических характеристик спектра, анализируется роль индивидуальных артикуляторных стратегий и темпа речи.

Изучение функционирования в речи гласных звуков является одной из сложнейших проблем, не нашедшей своего удовлетворительного решения до настоящего времени. Известный фонетист Джон О'Хала, резюмируя историю исследований звучащей речи, приходит, в частности, к неутешительному выводу о том, что до сих пор нет единого мнения о том, каковы акустико-слуховые корреляты фонетического качества гласного.

Столкнувшись с антиномией инвариантности фонологических единиц и их устрашающей реализационной вариативностью в речи, некоторые лингвисты впали в фонетический агностицизм, отрицая роль артикуляторно-акустических свойств звуковых единиц в их классификации (фонетическом декодировании речевого потока). Так, авторитетный исследователь русской фонетики Л.В. Бондарко пришла к выводу, что свойства психологического пространства гласных фонем не вытекают из фонетических (материальных) свойств их звуковых реализаций, а определяются, в частности, такими их функциональными свойствами, как частота встречаемости в речи или возможность чередования в слове [1, стр. 184].

На наш взгляд, реализационная вариативность гласных в русской речи действительно имеет архи сложный, но систематический характер, который может и должен быть раскрыт в результате систематических и кропотливых исследований всех аспектов функционирования звуков в речевой коммуникации.

В фокусе настоящей работы – спектральная классификация русских ударных гласных, когда их вариативность обусловлена такими факторами, как твердость/мягкость окружающих согласных, длительность гласного, обусловленная не темпом речи, а лингвистическими переменными (длина предложения, число слогов в слове и т.д.),

и в определенной степени дикторская индивидуальность.

Методика. Речевой материал и дикторы. В соответствии с задачами исследования использовались три вида речевого материала: 1) односложное псевдослово со структурой CVC, в которой согласные симметричны и реализуются твердым или мягким фрикативным [s, s'], а гласный представлен соответствующими 10 аллофонами пяти гласных фонем: между твердыми согласными (твердый контекст) – [a, e, o, u, ɨ] (твердые аллофоны) и между мягкими (мягкий контекст) – [ja, je, jo, ju, i] (мягкие аллофоны). Для записи речевого материала был составлен список, в котором каждое из 10 псевдослов повторялось 33 раза в случайном порядке; 2) вышеописанное тестовое псевдослово было включено в состав рамочного предложения “Повтори... опять”. Как и в предыдущем случае список для прочтения состоял из 330 предложений; 3) рамочное предложение “Мне сообщили, что старейшая... компания обанкротилась” содержало псевдослово “бестѳурная”, в котором согласный слева и справа от ударного гласного V реализовывался твердым или мягким смычным [t, t'], а в качестве V выступали соответствующие 10 аллофонов. Список для чтения состоял из 110 переложений с 11 кратным повторением в случайном порядке 10 исходных предложений.

Для краткости первый тип речевого материала будем называть в дальнейшем “слово–фраза”, второй тип – “короткая фраза” и третий тип – “длинная фраза”. Как показали последующие измерения [], структура этих фраз позволила получить тестовые гласные разной длительности (средние значения соответственно – 154, 99 и 76 мс) при нормальном темпе чтения.

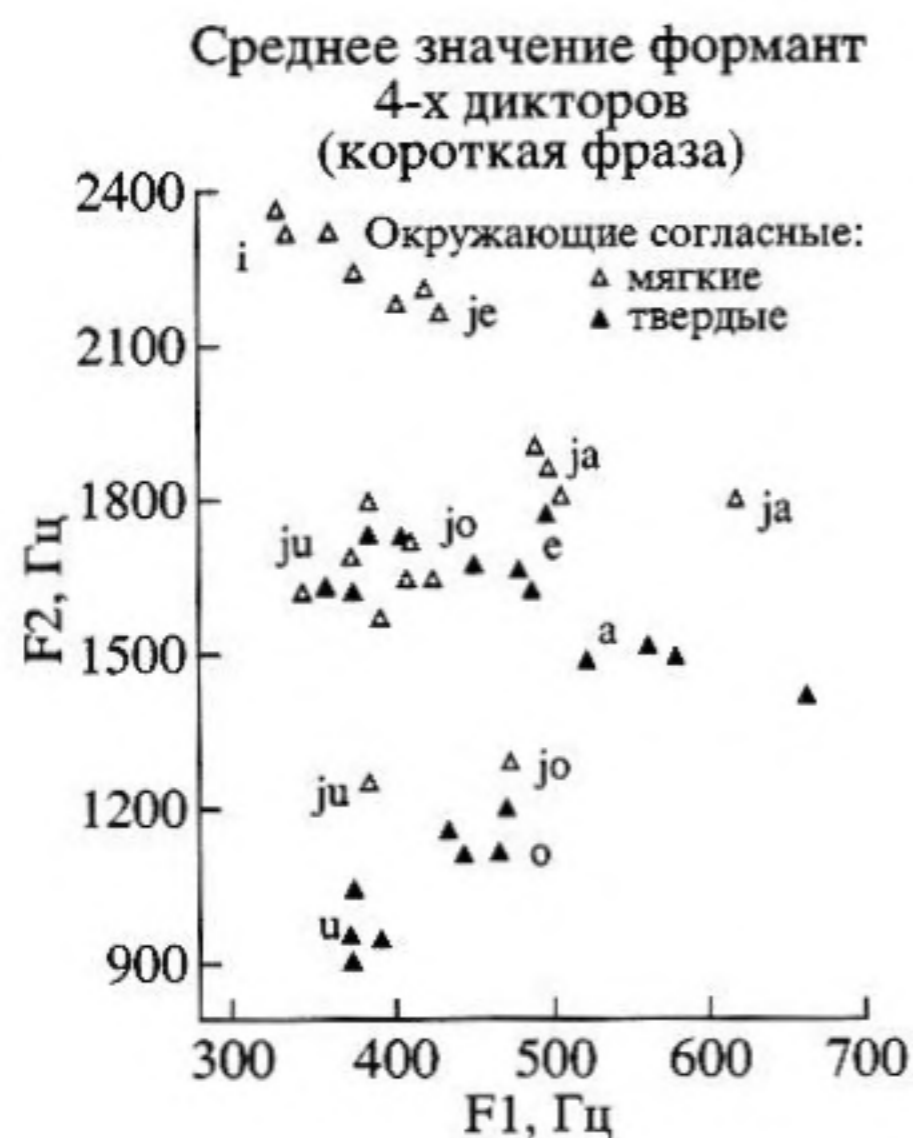


Рис. 1.

Речевой материал “короткая фраза” был записан от четырех дикторов мужчин в возрасте 25–50 лет, уроженцев города Москвы, владеющих литературной нормой произношения и не имеющих дефектов речи. Остальной речевой материал был прочитан одним из этих четырех дикторов (диктор С).

Спектр гласного измерялся на границе с предшествующим согласным (переходное значение) и в точке, где траектория второй форманты достигала экстремума, или в середине гласного (целевое значение). Вычисление частот и амплитуд трех первых формант обычно производилось автоматически, но в трудных случаях – вручную с использованием ЛПК-анализа, контролируемого

по FFT-спектрограммам. Более детальное описание методологических аспектов исследования можно найти в [2, 3].

Результаты спектральных измерений традиционно приводятся в пространстве F1–F2 в виде точек – средних значений параметров (рис. 1) или эллипсов (остальные рисунки), представляющих собой 95%-ые совместные доверительные области значений параметров для данного гласного. Ориентация эллипса определяется знаком коэффициента корреляции между F1 и F2, а его главная ось лежит на прямой линейной регрессии F2 по F1.

На рис. 1 приведены для 4-х дикторов средние целевые частоты формант. Можно видеть, что одни и те же гласные разных дикторов расположены на плоскости F1–F2 довольно компактно, за исключением четырех гласных диктора З: [ju, jo, ja, a]. На рис. 1 эти гласные помечены отдельно. Для более детального анализа полученных результатов на рис. 2 и 3 представлены данные соответственно диктора З и типичного представителя остальных дикторов диктора С. Начнем анализ данных со сравнения спектральной вариативности аллофонов одной и той же фонемы. Из рисунков видно, что у обоих дикторов целевые значения частоты F1 у мягких гласных неверхнего подъема систематически больше, чем у соответствующих твердых аллофонов, однако это различие не достигает уровня статистической значимости.

Что же касается частоты F2, то у мягких и твердых аллофонов одной и той же фонемы разность между целевыми значениями лежит в диапазоне 300–700 Гц и, за исключением одного случая, является статистически достоверной у обоих

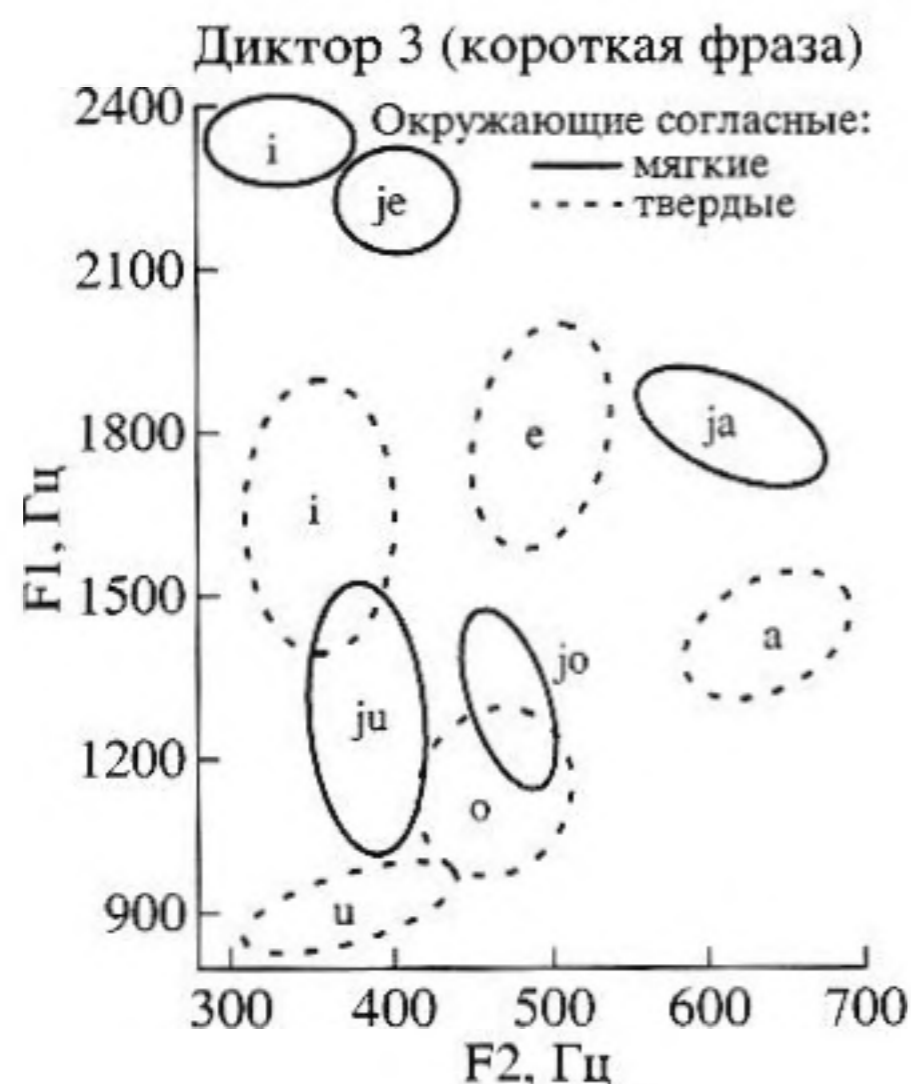


Рис. 2.

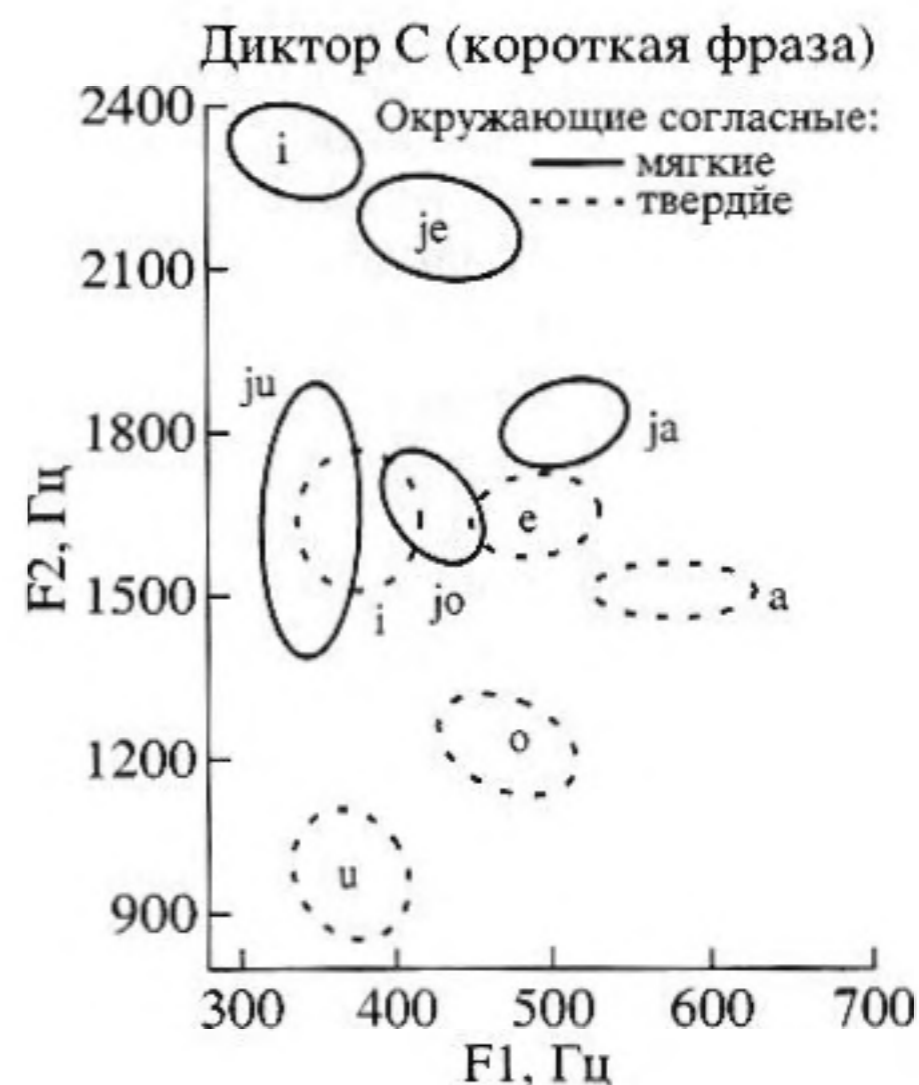


Рис. 3.

дикторов. Отклонение от этой закономерности связано с пересечением эллипсов гласных [o] и [jo] диктора З (рис. 2). Заметим, что у аллофонов фонемы /u/ того же диктора доверительные области почти соприкасаются.

Оценивая различия между твердыми и мягкими аллофонами, мы приходим к убеждению, что они не могут быть полностью отнесены на счет эффекта коартикуляции, когда наблюдаемую F-структуру в контексте мягких согласных трактуют как не достижение акустической цели, за которую принимается F-структура твердого аллофона. Мы предполагаем, что для реализации фонемы в твердом и мягком контексте используются разные артикуляторные программы. В пользу такого предположения можно привести несколько доводов. Во-первых, зафиксированные в исследовании максимальные величины "недолета" F2 (см., например, на рис. 3 аллофоны фонем /u/ и /e/) явно выходят за рамки того, что обычно, связывается с коартикуляцией звуков. В наших данных примером вариативности гласных, связанной с коартикуляцией, может служить аллофоническая реализация фонемы /o/ и в некоторой степени /u/ диктором З (рис. 2).

Во-вторых, мягкие аллофоны занимают на плоскости F1–F2 области, которые приписываются Международной фонетической ассоциацией гласным, отличным от основного варианта фонемы (твердого аллофона). По своим формантным характеристикам, измеренным в положении максимально близком к целевому, гласные [ju] и [jo] диктора С (рис. 3) могут классифицироваться как передние огубленные в отличие от своих "визави", задних огубленных.

В-третьих, считается, что степень коартикуляции обратно пропорциональна времени, имеющемуся для артикуляции звука. Однако формантные характеристики мягких аллофонов диктора С измеренные в трех типах фраз, обеспечивающих разную длительность гласных, отличаются друг от друга незначительно. Это можно видеть на рис. 3 и 4, а также в таблице, где приведены целевые частоты трех первых формант для 4 гласных диктора С. Так, хотя средняя длительность гласного в длинной фразе была в два раза больше, чем в короткой, различие в значениях F2 не превышает 150 Гц. Обращает на себя внимание также тот факт, что средние длительности гласных [jo] и [ju] у дикторов С и З практически совпадают, а формантные характеристики гласных различаются довольно значительно (ср. рис. 2 и 3), что ставит под сомнение использование ими одних и тех же артикуляторных программ.

Подытоживая, можно заключить, что все дикторы при произнесении твердых и мягких аллофонов реализовывали разные артикуляторные программы. Диктор З в случае фонемы /o/ и, воз-

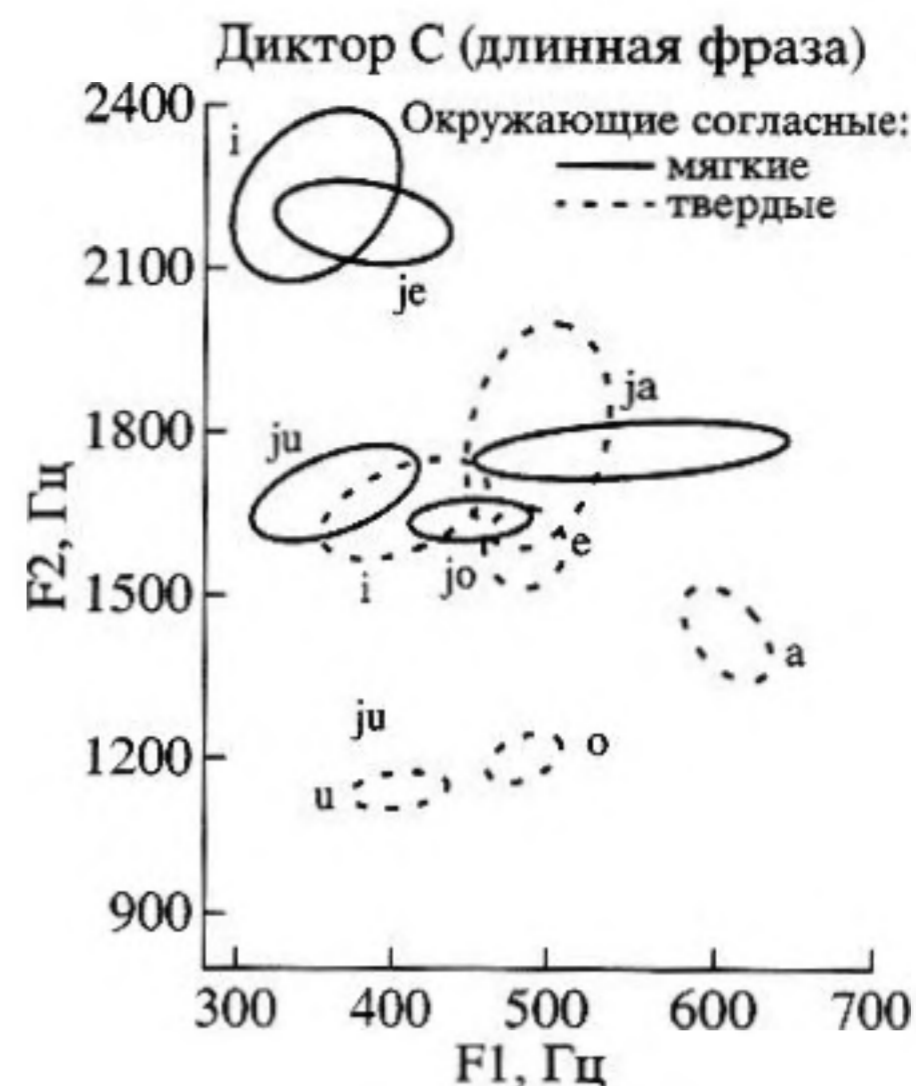


Рис. 4.

можно, /u/ использовал одну и ту же программу; спектральные же различия твердых и мягких аллофонов в данном случае связаны с коартикуляцией.

Перейдем теперь к анализу расположения гласных в пространстве F1–F2 по отношению к друг другу, независимо от их фонемной принадлежности. Из рис. 1, 2 и 3 видно, что гласные расположены в виде треугольника, в вершинах которого находятся гласные [i, a, u]. Этот треугольник распадается на два меньших треугольника, образуемых твердыми и мягкими аллофонами. В их вершинах располагаются соответственно гласные [i, a, u] и [i, ja, ju]. Одна из сторон этих треугольников совпадает: она лежит в диапазоне частот F2 1500–1800 Гц. Отметим, что это не относится к конфигурации гласных диктора З. Общая сторона треугольников образуется в результате

Динамика формантных частот в зависимости от длительности гласного (диктор, С, V_i – начало гласного, V_t – цель гласного)

Гласн.	Слово-фраза			Короткая фраза			Длинная фраза		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
i	418	1360	2535	394	1483	2445	420	1654	2555
t	403	1504	2503	375	1627	2364	407	1658	2520
E_i	462	1490	2603	451	1575	2542	458	1607	2565
E_t	534	1598	2634	497	1633	2518	474	1633	2614
JO_i	410	1779	2600	385	1855	2621	398	1680	2572
JO_t	479	1570	2522	424	1652	2510	451	1632	2599
JU_i	389	1802	2531	385	1873	2511	385	1715	2525
JU_t	386	1533	2406	346	1624	2391	363	1684	2520

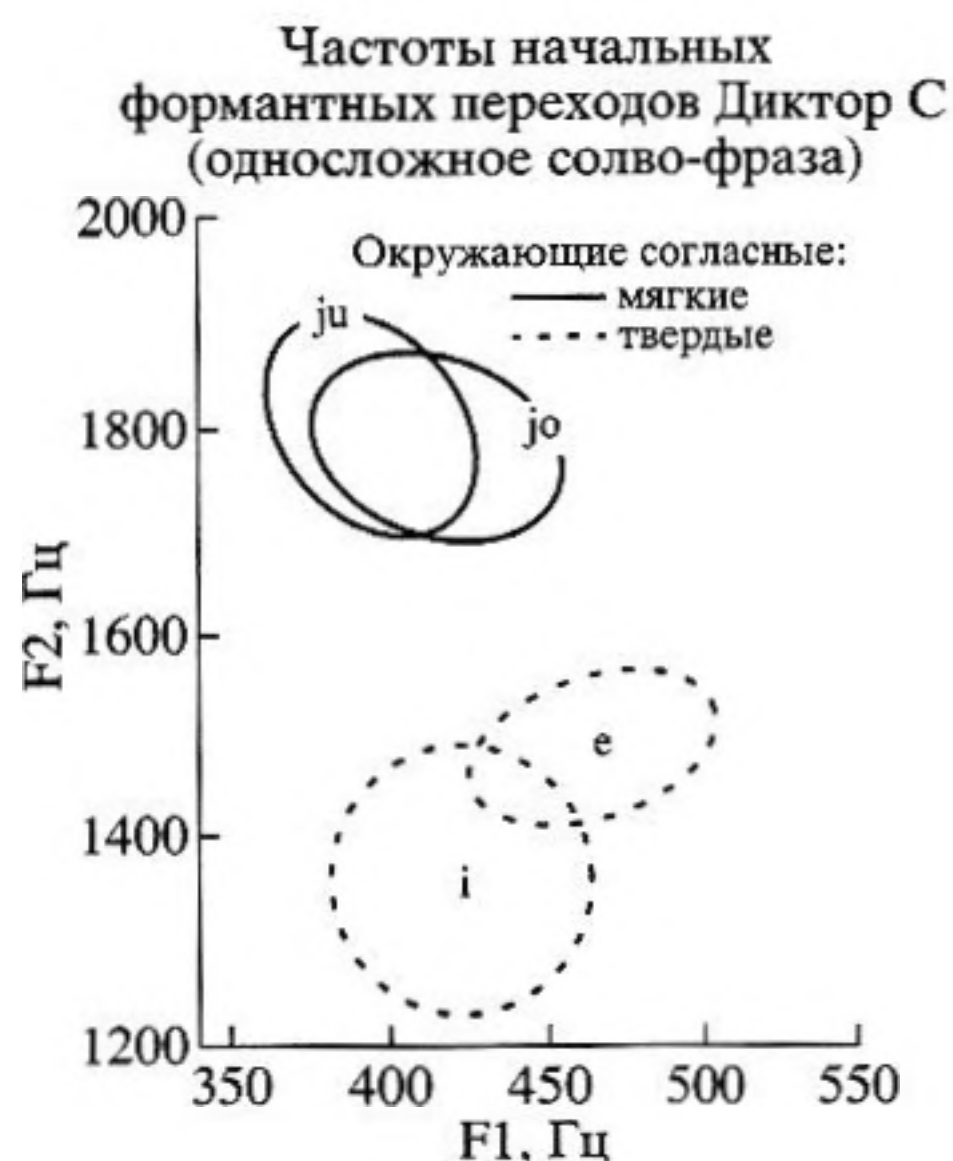


Рис. 5.



Рис. 6.

пересечения доверительных областей гласных [ju] и [jo] с [i] и [e] (см. рис. 3). У этих гласных совпадают частоты F2, и они дифференцируются только по значению F1. Это означает, что статическая спектральная характеристика ударных гласных даже одного диктора в принципе не обеспечивает их полного различия. Снятие неоднозначности статической классификации гласных легко достигается, если мы можем определить, произнесен ли данный гласный после твердого или мягкого согласного [4]. Как известно, о характере предшествующего согласного можно судить по формантному переходу: после мягкого согласного частота F2 не опускается ниже 1700 Гц.

Проведенные нами измерения F-структуры на границе тестовых гласных с предшествующими согласными представлены для 4-х вышеуказанных гласных на рис. 5 и в таблице. Можно видеть, что начальные значения формантных переходов у твердых и мягких аллофонов хорошо разнесены по F2. Таким образом, объединив данные рис. 3 и 5 (что позволяет нам отследить динамику спектральных изменений на отрезке гласного), мы получаем полную дифференциацию тестовых гласных.

В случае диктора З для классификации гласных можно ограничиться статическими спектральными характеристиками.

В завершение статьи рассмотрим, насколько все вышесказанное остается справедливым в ситуации максимального сокращения длительности тестовых гласных (длинная фраза). Что же касается двух других исследованных градаций длительности, то с точки зрения спектральных характеристик мы не обнаружили существенных

расхождений между гласными из слова-фразы и из короткой фразы.

Анализ рис. 4 показывает, что по сравнению с гласными короткой фразы (рис. 3) расположение аллофонов стало более компактным, и они разбились на три группы: хорошо различимые [u, o, a], тесно сбившиеся [ju, jo, e] с приближающимся к ним [ja] и ставшие практически неразличимыми [je] и [i]. Большинство эллипсов ориентировано горизонтально, что означает малую вариативность у гласных частоты F2. Как и в случае с гласными короткой фразы статическая спектральная информация явно не достаточна для дискриминации всех аллофонов. Ситуация стала даже хуже, так как для гласных [je] и [i] привлечение данных о переходных значениях формант бессмысленно – оба гласные мягкие аллофоны. Измерения переходных значений для гласных [ju, jo, e], представленные на рис. 6, показывают, что ни по F1, ни по F2 различия между твердыми и мягкими аллофонами не достигают статической значимости. Можно отметить, что только область совместных доверительных значений для [e] не пересекается с доверительными областями мягких аллофонов. Таким образом, дискриминация гласных [ju, jo, e] на основе спектральной информации, содержащейся только на отрезке гласного, оказывается невозможной, так как формантный переход не показывает твердость/мягкость предыдущего согласного.

Так как человек не путает между собой эти гласные (в нашей предыдущей работе [5] мы провели эксперимент по распознаванию слушателями тестовых гласных в длинной фразе), остается предположить, что человек сначала распознает не отдельные звуки, а более крупные единицы ти-

па слога, из которых затем извлекаются фонетически единицы меньшего размера. Заметим, что эта гипотеза выдвигалась исследователями раньше, но по другому поводу: было установлено, что один и тот же фрикативный шум взрыва идентифицируется по разному в зависимости от фонетического качества последующего гласного [6].

СПИСОК ЛИТЕРАТУРЫ

1. *Бондарко Л.В.* Фонетическое описание языка и фонологическое описание речи. Л., 1981. С. 198.
2. *Kouznetsov V., Ott A.* Inherent vowel duration in Russian // *Estonian Papers in Phonetics (EPP 1984–1985)*. Tallinn, 1986. P. 67–95.
3. *Kouznetsov V.B., Ott A.* Spectral properties of Russian stressed vowels in the context of palatalized and nonpalatalized consonants // *Proc. XI-th Intern. Congress of Phonetic Sciences*. V. 3. Tallinn, 1987. P. 117–120.
4. *Кузнецов В.Б.* К вопросу о представлении фонетических знаний в моделях речевой деятельности. Опыт описания аллофонической структуры ударного вокализма в русском языке // *Лингвистические аспекты проблемы различительных признаков в системах автоматического распознавания и синтеза речи (Сборник научных трудов / МГИИЯ им. М. Тореза; вып. 329)*. М., 1989. С. 38–49.
5. *Кузнецов В.Б., Ott A.* Автоматический синтез речи. Алгоритмы преобразования "буква-звук" и управление длительностью речевых сегментов. Таллинн, 1989. С. 135.
6. *Cooper F.S. et al.* Some experiments on the perception of synthetic speech // *JASA*. 1952. V. 24. P. 597–606.

Spectral Dynamics and Classification of Russian Vowels

V. B. Kouznetsov

Moscow State Linguistic University, ul. Ostozhenka 38, Moscow, 119034 Russia
e-mail: kvlad@ccas.ru

Abstract—The spectral characteristics of stressed Russian vowels are studied experimentally in relation to the palatalized or nonpalatalized feature of the consonant environment, the local speech tempo, and the speaker effect. The experimental data are used as a basis to discuss the possibility of the vowel classification by their static and dynamic spectral characteristics. The role of individual articulatory strategies and speech tempo is analyzed.