Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2018»

Москва, 30 мая — 2 июня 2018 г.

ФОНЕТИЧЕСКИЙ КОМПОНЕНТ В ИНТЕГРАЛЬНОЙ МОДЕЛИ ЯЗЫКА: МОДЕЛИРОВАНИЕ ПРОЦЕССОВ КОАРТИКУЛЯЦИИ В РУССКОЙ РЕЧИ НА ОСНОВЕ ЛОКУС УРАВНЕНИЙ

Кузнецов В. Б. (kuvlad2007@yandex.ru)

Московский государственный лингвистический университет, Москва, Россия

В статье рассматривается сложный вопрос исследования количественных характеристик СГ коартикуляции в русской речи методом локус уравнений. Показано, что в ударном слоге F2 переход различается у твёрдых согласных и не зависит от способа образования согласных. В безударном слоге наблюдается последовательное противопоставление твёрдых велярных губным. Не различаются уравнения регрессии только у безударных и ударных велярных.

F2 переход мягких согласных, находящихся в ударном и безударном слоге, не зависит от места и способа образования согласного. Установлено, что у безударных и ударных мягких согласных уравнения регрессии значимо различаются: у безударных мягких y-пересечение более низкое, а наклон линии регрессии более крутой.

Ключевые слова: форманта; формантные треки; формантные переходы; линейная регрессия; локус; локус уравнение; y-пересечение; коартикуляция

PHONETIC COMPONENT OF INTEGRAL LANGUAGE MODEL: MODELING COARTICULATION PROCESSES IN RUSSIAN BY MEANS OF LOCUS EQUATIONS

Kouznetsov V. B. (kuvlad2007@yandex.ru)

Moscow state linguistic university, Moscow, Russia

The aim of the present paper is to describe CV coarticulation in Russian by means of locus equations. The effects of such factors as place and manner of consonant articulation, its palatalization, position in stressed or unstressed syllable on properties of F2 transition were examined. It was shown that locus equations for nonpalatalized stressed consonants were different. The role of manner of articulation in this case was insignificant. F2 transition of palatalized stop and fricative consonants were adequately described by two regression equations—one in stressed syllable and the other for the consonants unstressed syllable.

Key words: formant; formant transition; linear regression; locus; locus equation; intercept; slope; coarticulation; formant tracks

1. Введение

Со времён первых спектрографических исследований речи [Potter et al. 1947] сложилось представление о том, что формантные переходы гласного содержат информацию, в частности, о месте образования предшествующего согласного. Анализируя спектрограммы сочетаний взрывных согласных с различными гласными, Поттер и др. [Potter et al. 1947] высказали предположение, что в качестве инвариантного акустического коррелята места образования согласного (МОБР.) может служить частота F2[1] того гласного, у которого после данного согласного отсутствует формантный переход. Так, для согласны [b] и [d] такими гласными-индикаторами являются гласные [u] и [æ], соответственно.

В пользу предположений Поттера и др. свидетельствовали результаты экспериментов Хаскинских лабораторий [Liberman et al. 1954] по восприятию двухформатных гласных, синтезированных по нарисованным от руки спектрограммам. Было показано, что наилучшее восприятие [b] и [d] достигалось, если F2 переход начинался или «указывал» на частоту 720 Гц и 1800 Гц, соответственно. В случае с [g] этот звук воспринимался удовлетворительно только перед незадними гласными при начале F2 с 3000 Гц. Эти характерные частоты авторы назвали локусом (locus) согласного. Последующие эксперименты показали, что МОБР, например, [d] однозначно идентифицируется в контексте различных гласных, если F2 переход только ***указывает***, но не начинается с 1800 Гц [Delattre et al. 1955].

Однако исследование в последующие годы формантных переходов в реальной речи [Lehiste, Peterson 1961] не подтвердило статус локуса как инвариантной характеристики МОБР согласных. Отмечалась лишь относительная стабильность F2 переходов для альвеолярного [d]. По мнению Л. А. Чистович и В. А. Кожевникова [Chistovich, Kozhevnikov 1972], понятие виртуального локуса совершенно несостоятельно с точки зрения обработки речевого сигнала слуховой системой человека.

В 90-е годы прошлого столетия Суссман провел серию исследований [Sussman et al. 1991]; [Sussman, Shore 1996] для изучения информативности формантных переходов, используя так называемые локусные уравнения (locus

[1] Характерную для гласных и согласных частоту F2 авторы называли хаб (HUB).

equations). Идею локус уравнений предложил Б. Линдблом [Lindblom 1963]. Анализируя редукцию гласных, он заметил, что между начальным значением частоты второй форманты гласного и её значением в момент достижения гласным целевого состояния существует линейная зависимость.

В математической статистике разработан метод парной линейной регрессии, который позволяет по исходным данным (в нашем случае пары значений F2, характеризующих формантный переход) подобрать прямую линию (уравнение прямой линии), которая в некотором смысле наилучшим образом описывает искомую зависимость. Начальные значения F2 откладываются на оси Y (зависимая переменная), целевые значения F2 — на оси X (независимая переменная, предиктор). Соответственно локус уравнение имеет следующий вид для F2:

$$F2\text{старт} = \alpha F2\text{центр} + c,$$

где α — угол наклона линии регрессии, c — точка пересечения линии регрессии с осью Y (y-пересечение). Величина угла наклона характеризует степень коартикуляции МОБР согласного с последующим гласным: чем меньше наклон, тем меньше влияет гласный на МОБР согласного.

Принципиальное отличие «виртуального локуса» Хаскинских лабораторий от метрики локус уравнения Суссман видит в том, что в первом случае речь идёт об абстрактном фиксированном акустическом корреляте МОБР, не зависящим от последующего гласного, тогда как локус уравнение предназначено отразить систематический характер контекстно-обусловленной взаимосвязи между начальным и целевым значением F2 перехода.

Безошибочное распознавание МОБР [b, d, g] на основе угла наклона и пересечения позволило Суссману [Sussman et al. 1991] заключить, что локус уравнения обладают реляционной (relational) инвариантностью акустического коррелята МОБР взрывных согласных.

В отечественной фонетике исследование коартикуляционных процессов в большинстве случаев, за редким исключением [Krivnova, Kuznetsova 1980], ограничивается отдельными наблюдениями, преимущественно качественного характера, несмотря на использование инструментальных методов [Bondarko 1977]. В работе [Lobanov et al. 1982] приведена формула для расчета формантных переходов гласных в целях синтеза русской речи по тексту. Можно предположить, что расчетная формула основана на ограниченных данных.

В настоящей работе методика локус уравнения используется как способ количественного описания в русской речи свойств F2 перехода в структуре СГ в зависимости от таких факторов как МОБР согласного, его твердость-мягкость, способ образования и ударность-безударность последующего гласного.

2. Методика

2.1. Речевой материал

Характер разработанного речевого материала должен был позволить исследование влияния на СГ коартикулящию таких факторов, как МОБР

согласного, его твёрдость-мягкость, способ его образования (взрывной-фрикативный), нахождение в ударном или безударном слоге. Учитывая предшествующий опыт применения методика локус уравнения, использовалась бессмысленная структура СГСГ, в которой оба согласных и гласных реализовывались одним и тем же звуком, например [тата], а ударение приходилось на второй слог. В качестве согласных использовались твёрдые и мягкие взрывные и фрикативные согласные [п, п', т, т', к, к', ф, ф', с, с', х, х'] и в качестве гласных — соответствующие аллофоны гласных фонем /а, и, у/. При записи речевого материала дикторам было предложено по возможности не редуцировать качество гласного в безударном слоге. Введение этого ограничения и наличие в речевом материале звукосочетаний, нетипичных для русского языка, продиктовано стремлением получить на начальном этапе исследования более полную репрезентацию артикуляторного и акустического пространства гласных.

Исследуемые структуры произносились в рамочном предложении «Вырос СГСГ сильным».

Для чтения был составлен список из 108 предложений (12 согласных х 3 гласных х 3 повторения). Предложения были объединены в 18 блоков, в которых использовался один и тот же гласный и все согласные были твёрдыми или мягкими. Последовательность блоков была случайной. Подобная организация материала позволила существенно сократить число ошибок при чтении исследуемых структур.

2.2. Дикторы и запись речевого материала

Были записаны шесть дикторов мужчин без дефектов речи и диалектных черт. Запись проводилась в безэховом помещении с помощью микрофона Philips SBC MD 110, внешнего АЦП E-MU 0204 и компьютера. Частота дискретизации — 44100 Гц, разрядность — 16 бит.

2.3. Измерение формантных частот

До проведения измерений исходная частота дискретизации была снижена до 11025 Гц. Точки измерения частоты F2 определялись по спектрограмме и осциллограмме. Значение F2 на границе согласного с гласным измерялось на спектре, вычисленном на интервале первого полноценного периода колебания голосовых связок. Второе значение F2 измерялось в точке, где траектория F2 стабилизировалась или достигала экстремального значения.

В обоих точках длина окна анализа, задаваемая вручную, совпадала с одним периодом. Спектральные разрезы вычислялись двумя методами: быстрое преобразование Фурье (БПФ) и линейное предиктивное кодирование (ЛПК). Порядок модели ЛПК подбирался таким образом, чтобы добиться максимального совпадения с БПФ спектром. Порядок модели мог принимать значения от 10 до 14. Частота F2 измерялась вручную на ЛПК спектре.

3. Результаты

Оценивая в целом качество построенных уравнений линейной регрессии, следует сказать, что все уравнения статистически значимы при $\alpha = 0{,}05$, коэффициенты всех уравнений достоверно отличны от нуля, за исключением двух случаев: значений пересечения в уравнении для твёрдых [к, х] в ударном слоге.

3.1. Мягкие взрывные и фрикативные согласные

Как и можно было ожидать [Bondarko 1977], ни у ударных, ни у безударных мягких согласных (имеются ввиду согласные в ударном или безударном слоге) МОБР согласного и способ его образования не оказывают влияние на начальное значение F2 перехода.

Нахождение в ударном или безударном слогах является статистически значимым фактором. Для ударных и безударных мягких согласных локус уравнения имеют следующий вид:

ударные мягкие: $\mathbf{Y = 0{,}45x} + 1118{,}2$ ($R^2 = 0{,}63$);

безударные мягкие: $\mathbf{Y = 0{,}74x + 511{,}5}$ ($R^2 = 0{,}71$).

Можно видеть, что у безударного мягкого согласного начальное значение F2 перехода расположено фактически на 500 Гц ниже, чем у согласного в ударном слоге, и оно в значительно большей степени зависит от последующего гласного.

Наше исследование перцептивной значимости формантных переходов гласного для идентификации МОБР глухих взрывных [Kouznetsov 1999] показало, что по переходу СГ МОБР мягкого согласного распознается на уровне случайного угадывания. Более высокий результат был характерен для распознавания по ГС переходу. Спектральный анализ F2 перехода в последовательности ГС свидетельствует о том, что МОБР согласного является вторым по значимости после ряда гласного фактором, определяющим конечное значение F2 перехода [Kouznetsov 1995]; [Purcell 1979].

3.2. Твёрдые взрывные и фрикативные согласные

Твёрдые ударные взрывные различаются либо по обоим параметрам (пересечению и углу наклона линии регрессии) соответствующих уравнений, либо по одному из них. Различие состоит в том, что 95 % доверительные интервалы значений сравниваемых параметров не пересекаются. Подобранные линии регрессии и уравнения, представленные на рис. 1 подтверждают современные представления о СГ коартикуляции. Как известно, велярный согласный легко коартикулирует с последующим гласным. Отсюда наибольший угол наклона линии регрессии. Наименьший угол характерен для губного согласного. Артикуляторному взаимодействию гласных с этим звуком противодействует использование разных артикуляторов: губ и языка. Уравнение регрессии для [п] способно объяснить только 60 % вариативности начального значения F2 перехода. R2 для [т] и [к] составляет 0,85 и 0,90, соответственно.

Способ образования твёрдых ударных и безударных существенной роли в формировании F2 перехода не играет: уравнения регрессии фрикативных согласных не отличаются от соответствующих взрывных.

Среди безударных твёрдых последовательно противопоставляются губные велярным. Безударный переднеязычный по некоторым параметрам совпадает с губным или с губным и велярным.

Сравнение уравнений для твердых ударных и безударных показывает, что только в случае велярных наблюдается чёткое совпадение. В тех случаях, когда ударные и безударные различаются, у последних отмечаются более низкие значения пересечения и более крутой наклон линии регрессии.

4. Выводы

К основным результатам настоящего исследования следует отнести следующее:

- продемонстрирована высокая эффективность инструмента локус уравнения при описании явлений коартикуляции в русском языке;
- установлено, что способ образования согласного не влияет на характеристики F2 перехода ни у твёрдых, ни у мягких согласных в ударном или безударном слоге;
- для мягких согласных продемонстрировано отсутствие влияния МОБР согласного и способа его образования на начальное значение F2 перехода, что подтверждает предположение Кривновой и Кузнецовой [Krivnova, Kuznetsova 1980];
- впервые установлено различие F2 переходов у мягких согласных. находящихся в ударном и безударном слогах.

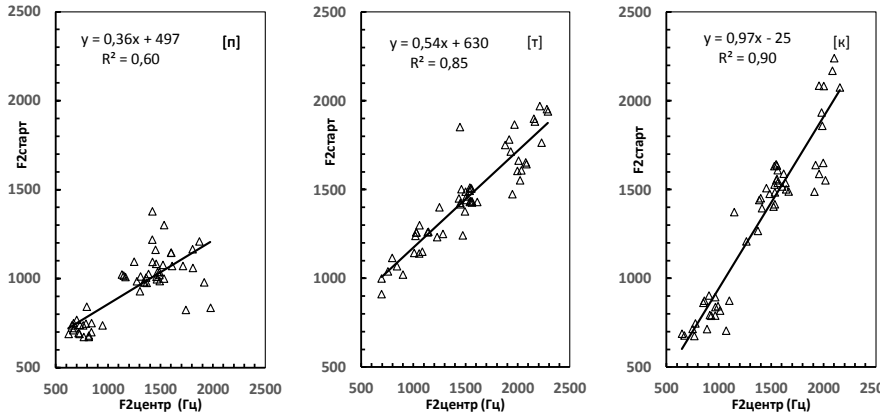

Рис. 1. Диаграммы рассеяния данных с подобранными линиями регрессии и соответствующие уравнения для твёрдых взрывных в ударном слоге. В уравнении регрессии для [к] значение y-пересечения не отличается от нуля при $\alpha = 0{,}05$

Литература

1. *Bondarko L. V.* (1977), Sound pattern of modern Russian. [Zvukovoj-stroj-sovremennogo-russkogo-yazyka], Education [Prosveshchenie], Moscow.
2. *Chistovihch L. A., Kozhevnikov V. A.* (1972), Speech perception, [Vosprijatie rechi] Phisiology of sensor systems [Fiziologiya-sensornyh-sistem], Nauka, Leningrad, pp. 427–515.
3. *Delattre P. C., Liberman A. M., Cooper F. S.* (1955), Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America. Vol. 27. pp. 769–773.
4. *Kouznetsov V. B.* (1995), Coarticulation effects at formant transition from vowel to soft consonant in Russian [Ehffekty koartikulyacii pri perekhode ot glasnogo k myagkomu soglasnomu v russkoj rechi], Phonetic issues [Problemy fonetiki]. Vol. 2, Nauka, Moscow, pp. 84–100.
5. *Kouznetsov V. B.* (1999), Role of linguistic and perceptual factors in identification of voiceless plosives using vowel context [Rol lingvisticheskih I perceptivnyh faktorov v identifikaci — gluhih smychnyh po vokalicheskomu kontekstu], Phonetic issues [Problemy fonetiki]. Vol. 3, Nauka, Moscow, pp. 48–60.
6. *Krivnova O. F., Kuznetsova V. B.* (1980) On phonetic features of hard vs. soft consonants in Russian (based on comparative analysis of consonant coarticulation in Russian and French) [O foneticheskoj prirode priznaka tverdost` myagkost` soglasny`x v russkom yazy`ka (na osnove sravnitel`nogo analiza koartikulyacii soglasny`x v russkom i franczuzskom yazy`kax)] // Theoretical and historical issues of Russian [Problemy` teorii i istorii russkogo yazy`ka.].MGU, Moscow, pp. 37–67.
7. *Lehiste I., Peterson G.* (1961) Transitions, glides, and diphthongs. Journal of the Acoustical Society of America, Vol. 33. pp. 268–277.
8. *Liberman A. M., Delattre P. C., Cooper F. S., Gerstman L. J.* (1954)The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs,. Vol. 68. pp. 1–13.
9. *Lindblom B.* (1963) On Vowel Reduction Royal Institute of Technology. Report N 29. Stockholm.
10. *Lobanov B. M., Marchenkov M. A.* (1982) Algorithm of calculating formant parameters in a text to speech synthesis. [Algoritm sinteza formantnyh parametrov po tekstu], Proceedings of 12th allunion workshop 'Automatic recognition of perceptual images' ['Tezisy dokladov I soobshchenij 12-go vsesoyuznogo seminara avtomatcheskoe raspoznavanie sluhovyh obrazov], Kiev, pp. 414–417.
11. *Potter R. K., Kopp G. A., Kopp H. C. G.*(1947) Visible speech. Van Nostrand Co. New York.
12. *Purcell E. T.* (1979) Formant frequency patterns in Russian VCV utterances Journal of the Acoustical Society of America. Vol. 66(6). pp. 1691–1702.
13. *Sussman H. M., McCaffrey H. A., Matthews S. A.*(1991), An investigation of locus equations as a source of relational invariance for stop place categorization Journal of the Acoustical Society of America., Vol. 90, pp. 1309–1325.
14. *Sussman H. M., Shore J.* (1996), Locus equations as phonetic descriptors of consonantal place of articulation. Perception and Psychophysics Vol. 58, pp. 936–946.