

Е. Б. Карневская

ВЗАИМОДЕЙСТВИЕ ФОНОЛОГИЧЕСКОГО И ФОНЕТИЧЕСКОГО В МОДЕЛИРОВАНИИ ПРОСОДИИ

(из опыта исследований по лингвистическому обеспечению синтеза речи)

В статье подчеркивается актуальность теоретико-методологических вопросов для современных экспериментально-фонетических исследований. Отмечается принципиальное сходство новых стратегий просодического моделирования с традиционной британской школой анализа и описания фразовой просодии (интонации), главным образом с точки зрения критериев выделения и внутренней организации тонального контура. Разграничиваются понятия единиц просодического анализа и единиц моделирования просодии как структур (типов), противопоставленных другим типам в инвентаре единиц для лингвистического обеспечения синтеза речи, а также для лингводидактических целей. Приводятся примеры функциональной значимости фонетической дискретизации в рамках фонологически тождественных типов и категорий.

Ключевые слова: просодическая единица; тональный контур; акцентная группа; фонетическая дискретизация; фонологические различия; инвариантность; поликомпонентная структура.

INTERACTION OF PHONETIC AND PHONOLOGICAL FEATURES IN MODELING UTTERANCE PROSODY

(with reference to the experience of prosodic research for speech syntheses)

The paper is concerned with the fundamental issues of prosodic analysis. The new theories and strategies of prosodic modeling are claimed to share common grounds with the traditional British school of intonation (utterance prosody) in the principles of the internal organization of prosodic units and their invariant and variable components. Examples of phonetic-phonological interplay are produced referring to the author's experience of prosodic modeling for speech synthesis.

Key words: prosodic unit; pitch contour; accentual group; phonetic discretization; phonological distinctions; invariant features; polycomponential structure.

Вопросы таксономии, функций и моделей описания фразовой просодии (интонации) в последние десятилетия вновь стали актуальной самостоятельной научной проблемой, направленной на создание современной лингвометодологической платформы для экспериментально-фонетических исследований. Помимо научно-теоретических, такая платформа необходима и для прикладных целей, в том числе для создания дидактически ориентированных 'баз данных', расширяющих и обновляющих имеющийся объем знаний современном состоянии произносительной нормы языка (включая просодию). В решении этих задач полезным может оказаться опыт просодического моделирования как части лингвистического обеспечения многоязычного синтеза речи. Исследования в названной области проводились фонетистами МГЛУ (при непосредственном участии автора доклада) в течение более трех десятилетий в сотрудничестве с создателями программ автоматического синтеза речи [1; 2; 3; 4]. Подчеркнем, что полезными в лингводидактическом отношении могут быть не только положительные результаты, но и отрицательные 'уроки' синтеза [5; 6].

Ознакомление с рядом наиболее известных современных теорий просодического моделирования [7; 8; 9; 10] и сравнение с ними британского подхода

к описанию интонации, служившего концептуальной основой упомянутых выше исследований, выявляют их общность в отношении исходных методологических принципов. К ним, во-первых, относится примарность просодико-смысловой сегментации речевого потока в определении его дискретных фрагментов, совместимых со слоговыми последовательностями разной протяженности, включая односложные. Во-вторых, это роль акцентуации в определении опорных точек высотно-мелодической линии в рамках каждого из названных фрагментов. Постоянным признаком последних является наличие когнитивно-перцептивного высотного изменения, ассоциируемого с комплексом значений: коммуникативно-прагматических, логико- и эмоционально-модальных. Благодаря этому данные речевые отрезки представляют собой минимальные коммуникативно-значимые единицы – фразу или ее часть (синтагму), обладающую внутренней смысловой связанностью. Принципиально важно уточнить, на наш взгляд, что фраза/ синтагма, выделяемая в результате модификаций просодических характеристик (ч.о.т., интенсивности, длительности), одновременно служит для просодии ‘фиксированной вербальной основой’ [10], единицей анализа просодии, поскольку данная слоговая последовательность является носителем цельнооформленной просодической структуры – тонального контура как такового. С другой стороны, фраза/ синтагма выступает носителем тонального контура как категориальной единицы, противопоставленной по форме и содержанию идентичным по статусу единицам в просодической подсистеме конкретного языка. Соответственно, каждая фраза/ синтагма, выделяемая из речевого потока, репрезентирует тот или иной тип контура через определенный структурно-позиционный вариант одной из его функциональных (конфигурационных или высотно-диапазонных) разновидностей.

Как известно, инвентарь тональных контуров для английского, русского, белорусского и ряда других языков к настоящему времени, в целом, установлен. В то же время остаются разногласия относительно общего числа контуров, зависящего от принятой авторами степени дискретизации высотно-уровневых различий.

Нужно указать, что при моделировании просодии речи, как и вообще при описании просодических характеристик, тональный (высотно-мелодический) контур определяет классификацию просодических контуров (структур) как комплексных полипараметрических единиц. Такая трактовка обусловлена не только ведущей ролью высотно-тонального компонента в дифференциации и идентификации фразово-просодических различий; она учитывает вовлеченность двух других параметров – интенсивности и длительности – в формирование тонального контура, и прежде всего, его обязательного элемента – терминального (ядерного) изменения ч.о.т., – которое отличается от неядерных высотно-изменений увеличением длительности ударного и безударных слогов слова-носителя ядерного тонального акцента. Интегрированность амплитудно-временных модификаций в высотно-мелодические, однако, не нивелирует восприятия их ‘отдельности’, и, таким образом, просодический контур, как уже отмечалось, реально представляет собой поликомпонентную перцептивно-акустическую структуру. Ее динамичность по протяженности задана объективно-субъективной смысловой сегментацией линейно-речевого потока, в то время как инвариантность этой структуры обусловлена функциональной спецификой составляющих ее частей

(предшкалы, шкалы, ядра, заядра) изначально предназначенных для противопоставления акцентуации и дезакцентуации как основы для экспликации информационной структуры высказывания.

Линейность просодического контура раскрывается в последовательности акцентных групп, которые, как и контур, полипараметричны и вбирают в себя (по форме и по функции) три совмещенные на временной оси единицы: высотномелодическую (тон, тональный акцент), динамическую (фразовое ударение, акцент) и темпоральную (ритмогруппа, ритмический такт, стопа). Минимальная длина акцентной группы – односложное слово, выделенное полным фразовым ударением. В случае последовательности акцентированных односложных слов их разграничение не вызывает опасности нарушения целостности слов. Такая опасность возникает при словесно-слоговой развернутости смежных акцентных групп. Маркирование границ акцентных групп при моделировании определяется в этом случае избранным подходом к роли безударных слогов в формировании акцентной группы. Направленность на достижение смысловой разборчивости, что особенно важно для синтезированной речи, предполагает предпочтительность семантического подхода с подразделением безударных слогов на проклитики и энклитики и, кроме того, на внутрисловные и межсловные соединения слогов внутри акцентной группы. Эти два разграничения объясняются не только целесообразностью подчеркивания семантико-синтаксической связанности компонентов акцентной группы. Они опираются на установленные в экспериментальных исследованиях достоверные акустические различия между безударными слогами в зависимости от их позиции в слове и морфологической характеристики афикса или служебного слова.

Тот факт, что конституэнты контура – акцентные группы – не идентичны по своей форме и статусу, позволяет говорить о стирании качественных различий между признаками ядерной акцентной единицы и просодического контура. Последний отличается от нее, по существу, только сегментивно-интегративными свойствами. Правомерно поэтому рассматривать просодический контур и ядерную акцентную группу как единицы одной иерархической ступени, различая их как распространенный и нераспространенный контур соответственно. Более низкой по иерархии единицей следует признать неядерную акцентную группу, неспособную в силу своей структурной недостаточности к автономизации. Выполняя второстепенную роль с точки зрения конституирования контура, неядерная акцентная группа в то же время может выступать в качестве его дистинктивного элемента при более детализированной высотно-уровневой дискретизации предъядерной части и расширении на этом основании имеющегося инвентаря контуров в целом. Следует указать и на участие акцентной группы в экспликации информационного объема высказывания.

Следует уточнить, что тезис о нечленимости акцентной группы для русского, белорусского и английского языков, на наш взгляд, можно принять только в отношении корреляции информационной и акцентной структур. Причина того, почему акцентная группа не может быть признана в полном смысле неделимой, заключается в ее внутренней организации, допускающей, как было указано выше, межсловное соединение элементов, причем не только неравноценных по признаку автосемантичности, но и лексически полнозначных, с вытекающими отсюда особенностями реализации.

Отмеченная сложность структуры акцентной группы оказалась в ходе эксперимента перцептивно существенной характеристикой, влияющей на оценку качества синтезируемой речи. С целью устранения негативных эффектов различия между внутриакцентным межсловным и внутриакцентным внутрисловным соединением сегментных единиц были реализованы путем 'отмены' коартикуляционных модификаций согласных в сочетаниях типа СГ на межсловных стыках (как, например, 'от Оли', 'at all'), а именно, путем изменения аллофонической репрезентации согласных и гласных фонем в зависимости от характера их стыка. В то же время на внутрисловных соединениях этого типа коартикуляция сохраняется в полном объеме [11; 12]. Подобное разграничение достигалось при компилятивном синтезе.

Наиболее сложными для реализации оказываются акцентные группы, представляющие собой соединение полнозначных лексических единиц, одна из которых выделена частичным ударением. Моделирование этого типа ударения, т.е. ударного слога частично выделенного слова, при компилятивном методе синтеза осталось практически нерешенной задачей. Теоретически представлялось возможным использовать, например, для частичного ударения в предшкале ядерные просодические дифоны (сочетания СГ) из низкого восходящего контура, поскольку в обоих случаях основной характеристикой является ровный низкий высотный уровень. Тем не менее, такая подстановка, при ее фонологической обоснованности, оказалась фонетически неприемлемой. Аналогичным образом оценивалось использование заядерных просодических дифонов для частичного ударения в предикате. Фонетическая несочетаемость фонологически 'адекватных' элементов является, на наш взгляд, убедительным доказательством функциональности градиентных высотных различий.

Сохранение качественных (а именно, направление, форма, и конфигурация) и количественных (регистр, интервал) характеристик типа тонального акцента независимо от слоговой 'длины' акцентной группы и, следовательно, реального распределения в ней движения ч.о.т. по слогам, обеспечило успешность применения данной модели при разных методах синтеза речи. Влияние слогового состава акцентной группы на поведение ч.о.т. на самом ударном слоге привело к необходимости введения двух разных просодических дифонов – фактических единиц синтеза. Таким образом, модель воспроизводит структуру продуцирования тона как при ее компрессированном, так и при развернутом варианте, соотносимую с идентификацией типа мелодического изменения как целостной фигуры. Ее воспроизведение и восприятие и является целью (target), в то время как слоги, составляющие ее, функционально изоморфны фонемному составу слова. Аналогичный процесс происходит в естественной речи при идентификации и воспроизведении тонального акцента. Так, например, реализация низкого восходящего тона в английском языке (*Low Rise*) и ИК-4 в русском языке в случае, когда ударный слог не является последним в интонационной группе / фразе, несмотря на низкий уровень ч.о.т. и отсутствие перцептивно значимого подъема ч.о.т. на ударном слоге никак не ассоциируется на уровне восприятия с низким статическим тоном. Таким образом, когнитивный эталон низкого восходящего тона 'предусматривает' сдвиг высотного изменения на заударные слоги, который не влияет на идентификацию локализации тонального акцента. Иными словами, мы наблюдаем здесь случаи просодических омонимов: совпадение формы и отсутствие такового в семантике тональных единиц.

Высотно-уровневые характеристики всех предъядерных акцентных групп сходны в своей противопоставленности характеристикам ядерной акцентной группы по признаку автономности, который на фонетическом уровне, как уже было отмечено, совмещает в себе присутствие изменения ч.о.т., т.е. 'кинетичности', и увеличения длительности всех слогов-конституэнтов данной акцентной группы, наиболее заметном на ядерном и заударном слогах. Отсутствие данной характеристики у предъядерных акцентных групп не означает отсутствия перцептивных высотно-уровневых модификаций на них. Мелодическая структура неядерных акцентных единиц может отклоняться от статичности (ср. скользящую, скандентную, нисходящую и восходящую 'шкалы'). При этом тип шкалы относительно независим от ядерного тона, т.е. в принципе, возможна любая сочетаемость тональных акцентов. С другой стороны, нельзя не признать, что смежные предъядерные акцентные группы в контуре могут различаться между собой по данному признаку, представляя таким образом гетерогенную шкалу. При моделировании просодии для синтеза речи случаи гетерогенной предъядерной высотно-мелодической структуры рассматривались как сочетания отдельных контуров. В теоретическом плане и в практике реализации просодической структуры такие 'шкалы' являются показателем тонкой просодико-смысловой сегментации речевого отрезка: свободной межакцентной связи или неполного (промежуточного) членения.

В естественной речи акцентное членение, как и синтагматическое, обусловлено смысловыми факторами, а его мелодические модификации определяются, главным образом, прагматическими и субъективно-модальными коннотациями. В то же время в моделировании для синтеза опора в выборе мелодического изменения ограничена формальными факторами, а именно, позицией акцентной группы в контуре: начальной, срединной, конечной. При этом, выбор образца (экземпляра) из естественной речи определяется статистической предпочтительностью и объемом модально-семантического содержания типа контура.

Позиционное маркирование предъядерных акцентных групп при любом мелодическом типе шкалы подтверждается экспериментальными данными, полученными на материале как естественной, так и синтезированной речи. Оценка качества просодии синтезированной речи, созданной на основе обсуждаемой модели, показала, что ударные просодические дифоны предъядерных акцентных единиц, отобранные из естественной русской и английской речи носителей языка без учета позиции акцентной единицы в названной части контура, стали причиной низкой оценки качества просодического аспекта синтезируемой речи и ее качества в целом. Информанты-носители языка и аудиторы-эксперты оценили реализации, в которых ударный слог начальной акцентной единицы был представлен срединным, более низким по высотному уровню дифоном, как неадекватные речевой ситуации с точки зрения передаваемой информации и звучащие 'странно' с точки зрения их представления о норме интонационной структуры. Таким образом, фонологически тождественные реализации, принадлежащие к одной категориальной единице – средне-высокому предъядерному статическому тону, обладают функционально значимыми фонетическими различиями. Тем самым экспериментально подтверждается необходимость двух этапов в анализе и репрезентации просодических характеристик высказывания, что и достигается фонетико-фонологическим анализом.

ЛИТЕРАТУРА

1. Карневская Е. Б. Тональный контур как единица просодической организации связной речи // Системные характеристики устной и письменной речи : сб. науч. ст. / Мин. гос. пед. ин-т иностр. яз. ; отв. ред. А. В. Зубов. Минск, 1985. С. 33–39.
2. Lobanov, B., Karnevskaya, E. MW speech synthesis from text // Proceedings of the XII. International Congress of Phonetic Sciences, Aix-en-Provence, August, 19-24 1991 : in 5 vol. Aix-en-Provence, 1991. Vol. 4. P. 406–409.
3. Lobanov, B., Karnevskaya, E. TTS-Synthesizer as a computer means for personal voice 'cloning' // Phonetics and its Applications : Festschrift for Jens-Peter Koester on the Occasion of his 60th Birthday. Stuttgart, 2002. P. 445–452.
4. Karnevskaya, E. B. Prosodic modelling in TTS-speech synthesis // Prosodie, traduction, fonctions : Actes du XXXII^e Colloque International de Linguistique Fonctionnelle, Minsk, 7-10 Oct. 2009 / L. Morozova, E. Weider (eds.). Fernelmont (Belgique), 2010. P. 59–62.
5. Карневская, Е. Б., Лобанов Б. М. Модели синтеза мелодического контура русских и английских фраз // Автоматическое распознавание и синтез речи : тез. докл. Всесоюз. науч. школы-семинара АРСО-12, Киев, 9-16 июня 1982 г. / АН УССР. Киев ; Одесса, 1982. С. 399–402.
6. Karnevskaya, E. B. The linguistic aspect of multi-language speech synthesis // Proc. of the 11th Intern. Congr. of Phonetic Sciences, Tallinn, August 1-7, 1987 : in 6 vol. / Acad. of Sciences of Estonian SSR, Inst. Of Lang. And Lit. Tallinn, 1987. Vol. 2. P. 168–171.
7. Pierrehumbert, J. Tonal Elements and their Alignment // Prosody: Theory and Experiment : studies presented to Gosta Bruce / ed. by M. Horneo. Dordrecht, 2000. P. 11–36.
8. Ladd, R. Intonational phonology. Cambridge : Cambridge Univ. Press, 2001. 334 p.
9. Gussenhoven C. The phonology of tone and intonation. Cambridge : Cambridge Univ. Press, 2004. 355 p.
10. Xu, Y. Speech prosody – theories, models and analysis // Courses on Speech prosody / ed. A. R. Meireles. Cambridge : Cambridge Scholars Publ. 2015. P. 146–187.
11. Карневская, Е. Б., Долматова, Е. Д. Словоразграничительный потенциал аллофонического варьирования английских согласных в речи // Вариативность в языке и речи : тез. докл. междунар. науч. конф., Минск, 4-5 апр. 2019 г. / Мин. гос. лингвист. ун-т ; редкол.: Л. М. Лещева (отв. ред.) [и др.]. Минск, 2019. С. 98–100.
12. Карневская, Е. Б., Долматова, Е. Д. Просодические ключи к распознаванию межсловного и внутрисловного соединения фонем // Фонетика в современной лингвистической парадигме : сб. науч. ст. / Мин. гос. лингвист. ун-т ; редкол.: Е. Б. Карневская (отв. ред.) [и др.]. Минск, 2021. С. 71–76.