

Е. Б. Карневская

ПРИНЦИПЫ И ЕДИНИЦЫ ФОНЕТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Фонологический принцип, приписывающий языковой статус звуковым элементам, имеющим смыслоразличительную функцию или обладающим собственной семантикой, стал общепринятым методологическим положением фонетических исследований. В основе данного принципа лежит признание существования различных фонетических форм и способности слушателя воспринимать их как таковые независимо от их смыслоразличительного потенциала. Автономность фонетической формы предопределила саму возможность создания общefonетических описаний и получила отражение, в частности, в международной фонетической транскрипции (IPA).

Кардинальный вопрос фонетики/фонологии о том, насколько велики должны быть различия, чтобы стать языковыми, и наоборот, как близки должны быть формы, чтобы функционировать как тождественные, решается, как известно, при обращении к семантическим оппозициям или дистрибутивным характеристикам [1]. При этом действительная роль фонетического сходства/несходства как основы для возникновения или, напротив, отсутствия фонологически дискуссионных вопросов как на сегментном, так и надсегментном уровнях часто не получает должного внимания.

Важность вопросов фонетического сходства в полной мере осознается при межъязыковом сопоставлении. Традиционно в центре внимания находятся фонетические особенности фонологически сходных звукотипов. Существенно, что последние прежде всего являются фонетически сходными. Иными словами, фонологическое сходство предполагает наличие фонетического подобия, в то время как обратной зависимости нет (ср., например, статус звонкого межзубного согласного [ð] в английском и испанском языках – фонема и аллофон соответственно). Тем не менее подчеркнем, что совпадение артикуляторно-перцептивных характеристик этих звуков фиксируется сознанием носителей

обоих языков, независимо от степени их осведомленности относительно фонологических систем изучаемого (или родного) языка. Очевидно в этой связи, что овладение особенностями варьирования фонем при их реализации в связной речи, как обусловленного структурно (т.е. коартикуляционными и позиционными факторами), так и диафонического (т.е. свободного, немотивированного), возможно лишь на основе осознанной перцептивной идентификации конститутивных признаков звуковой единицы и соотнесения их с артикуляторными признаками.

Использование понятия «звуковая единица» в данном случае соответствует его толкованию в работах Л. В. Бондарко, Л. А. Вербицкой, Л. В. Златоустовой и др. [2; 3; 4] как некоторой единицы, отличной и от фонемы, и от ее манифестации в речи. Это же значение «промежуточной единицы» присуще и понятию «аллофон» [5]. Специфика аллофона состоит в том, что он совмещает в себе признаки обобщенности, свойственные фонеме, и контекстуальной обусловленности, характеризующей ее реализацию. Именно через эту единицу и осуществляется воплощение фонемы в реальный звук речи. Иными словами, соотношение между «звуком языка» и «звуком речи» является опосредованным и выражается при помощи тройственной схемы: фонема → аллофон → звук, подчеркивающей, в силу этого, неправомерность отождествления аллофона с конкретным звуком.

Недистинктивность присущих аллофонам различий, вытекающая из невозможности их парадигматического противопоставления, предопределяет примарность фонетического сходства для их идентификации. Вместе с тем именно вопрос степени фонетического сходства/различия звуков, реализующих разные аллофоны одной фонемы, является наименее изученным, хотя для решения этой задачи имеется целый ряд теоретических положений, относящихся к общепринятым. Так, общеизвестно, что не все артикуляторно-акустические модификации, вызываемые коартикуляцией, перцептивны и что последние должны быть в принципе акустически более ярко выраженными [6]. Исследователям ясно также, что должны быть найдены количественные и качественные пороги указанного разграничения. К настоящему времени в нормативных фонетических описаниях оно базируется, главным образом, на интуитивно-эмпирических представлениях. Важно в этой связи отметить, что, поскольку при фонетически одинаковых условиях эффект взаимного влияния звуков в разных языках различен (вследствие своеобразия их фонологических систем и артикуляционно-перцептивных баз), ранжирование аллофонов в указанном выше плане имеет как лингвистически инвариантные, так и специфически языковые черты.

Признание фонетического неравенства аллофонов между собой имеет ряд важных теоретических и практических последствий, в частности для обучения произношению и для фонетического моделирования. В первом случае результатом ранжирования аллофонов является отбор вариантов фонем (звуковых единиц), подлежащих специальной тренировке. Примерами подобных вариантов в английском языке могут быть оглушенный фрикативный [r] после глухих взрывных, неаспирированные [t], [p], [k] после щелевого [s], дентальные варианты альвеолярных согласных перед межзубными [θ], [ð] и др.

Несмотря на существенность наблюдаемых в этих случаях артикуляторно-перцептивных изменений, обучение и овладение данными вариантами воз-

можно лишь в соответствующих звуковых сочетаниях, в отличие от «изолированного» обучения и овладения т. н. «основными аллофонами» фонем [7], выполняющими роль репрезентантов фонем на субстанциональном уровне.

Следует указать, что изолированное воспроизведение названных выше вариантов фонем в принципе достижимо, и это, в конечном счете, определяет их особое место в системе аллофонов, приближая наблюдаемую здесь вариативность к диафонической [6].

Отмеченные особенности ни в коей мере не меняют существа вопроса: обязательные для произносительной нормы языка варианты фонемных реализаций ассоциируются, как и все аллофоны, со строго определенными условиями функционирования. В то же время от прочих комбинаторно-позиционных реализаций тех же фонем их отличает изменение самого «качества» звука, предполагающее, что изменения в его артикуляторно-акустической структуре перцептивно значимы (не будучи дистинктивными).

Из сказанного следует, что для практических целей обучения языку нет необходимости «знать» общее число аллофонов каждой фонемы, выводимое, как известно, из фонотактических свойств данного языка. Именно такой избирательный подход отражен в нормативных описаниях звукового строя, необходимой частью которых является выделение регулярных аллофонов-вариантов фонем. Фонетическая специфика последних раскрывается через артикуляторные изменения, отличающиеся от обычных коартикуляционных процессов осознанностью и контролируемостью со стороны говорящего (обучаемого) и, следовательно, слушающего.

Несколько по-иному решается вопрос аллофонии с точки зрения прикладной задачи синтеза речи, неосуществимой без реализации коартикуляционных явлений. Первоочередным здесь становится выбор подхода к определению оптимальной аллофонной базы [8]. Сама возможность множественных решений относительно числа и состава аллофонов, безусловно, возникает в силу различной степени перцептивно-акустических модификаций, вызываемых коартикуляционными процессами. В этой связи важно отметить, что принципы, на которых строится моделирование в известных в настоящее время программах синтеза, в частности синтеза русской речи [9; 10; 11; 12], демонстрируют, при достаточно широком диапазоне вариантов, общую тенденцию к большей дискретизации по сравнению с моделями, которые стремились к максимальной сжатости и экономичности репрезентации фонетической системы языка. Поскольку, со всей очевидностью, большая дискретизация не тождественна максимальной, она обусловлена концептуальным подходом к соотношению коартикуляционных факторов по значимости, основанном как на известных из литературы, так и на экспериментальных данных.

При определении состава единиц моделирования для синтеза английской речи, разрабатываемой компанией «Сакрамент», одним из концептуальных решений было признание асимметрии в коартикуляции типов СГ и ГС с точки зрения как качественных, так и количественных модификаций каждой из смежных сегментных единиц. Указанная асимметрия состоит в превалировании влияния предшествующего согласного на качественные характеристики, а последующего согласного – на количественные характеристики гласного. Как результат такого подхода, число аллофонов гласных фонем задается числом

возможных слогоначальных сочетаний СГ. Тем самым фактором аллофонического варьирования гласного становится не только направленность переходов по FII, т. е. зависимость от места образования, но и другие составляющие коартикуляционного перехода, обнаруживающие зависимость от способа образования согласного и его фортичности/леничности.

Отражением неравнозначности левостороннего (ГС) и правостороннего (СГ) влияния в рассматриваемой аллофонной базе является отсутствие ориентированности гласного на тип последующего согласного. В то же время учитывается наличие правого консонантного контекста как таковое, благодаря тому что исходный звуковой материал состоит из слов типа ССГСС. Данное соображение, на наш взгляд, имеет принципиальный характер, поскольку меньшая степень модифицированности гласного справа в этом случае получает конкретное содержание, а именно, слабую возможность идентификации и дифференциации последующего согласного в зависимости от качества (а не количества) гласного. Вместе с тем возможность более заметного влияния гласного на последующий согласный, а именно на словоконечные [l], [r], [ŋ], учитывается путем введения трех типов конечных аллофонов для них: после а-образных, i-образных и u-образных гласных. Интервокальные сонорные [l, ŋ, n, m, w, r, j] рассматриваются как элементы соединения СГ.

В то же время нужно признать, что и без применения специальных приемов сглаживания разборчивость и легкость восприятия как основные составляющие качества синтезированной речи достаточно высоки. То есть можно говорить о способности восприятия «воссоздать» непрерывность артикуляции целостной фонетической единицы – слова – при определенной акустической дискретности его реализации. Фактически и в случае восприятия правостороннего стыка гласного и согласного (ГС) при отсутствии реального участка перехода имеет место виртуальная непрерывность, обеспечиваемая восприятием. В обсуждаемой программе участок гласного, примыкающего к последующему согласному, помимо своей недифференцированности относительно типа согласного, оказывается величиной непостоянной по протяженности. Причина этого заключается в упомянутой выше специфике коартикуляции типов СГ и ГС и превалировании роли количественных модификаций гласного для последних. Данные модификации осуществляются за счет сокращения или удлинения правостороннего участка гласного и не затрагивают так называемой стационарной части, которая является носителем качественных различий аллофонов.

Изменения в длительности гласного в принципе могут быть заданы априорно в соответствии с установленными в многочисленных исследованиях закономерностями ритмической организации речи. Предпочтение тем не менее отдавалось экспериментальному определению ритмических коэффициентов, позволяющему конкретизировать степень изменения с учетом длительности аллофонов в базе, а также с учетом специфики временных изменений для кратких, долгих и дифтонгов. Последний класс гласных фонем представляет особую трудность для осуществления варьирования по длительности: сокращение правой части дифтонга более чем на 10-15% не может не затронуть само качество второго элемента дифтонга и, таким образом, разрушить структуру дифтонга, предполагающую определенное соотношение первого и второго элементов по длительности. Стратегия, избранная для данного класса гласных, учитывала

необходимость сохранения данного соотношения (2:1), и сокращению/удлинению подвергалась срединная треть общей длительности дифтонга (ср. известные из нормативной фонетики указания на то, что модификации позиционной долготы гласных происходят на первом элементе дифтонга).

С целью определения диапазона варьирования длительности гласного, все аллофоны, согласно избранным группам, были реализованы диктором в двух фразовых позициях: начальной и конечной, предполагающей характерную для терминальной акцентной единицы временную растяжку наряду с изменением ЧОТ.

В реализации консонантных сегментных единиц одной из важнейших проблем является включение в модель ассимилятивных изменений в сочетаниях согласных. В обсуждаемой модели ассимиляция, ведущая к наиболее заметным (воспринимаемым) модификациям английских согласных, отражена путем введения дополнительного числа согласных «фонем» (tr, dr, kl, st, sp и целый ряд других). Иными словами, неравенство аллофонов по степени и характеру модифицированности по сравнению с основным репрезентантом фонемы приводит к необходимости их дифференцированной трактовки и усложнения общей классификации. Введение сдвоенных «квазифонем» не решает всей проблемы внутрислоговых, межслововых и межсловных консонантных стыков, разнообразие и частотность которых относится к специфическим чертам английской речи. Отражение этих особенностей в модели синтеза может существенно повысить его качество.

ЛИТЕРАТУРА

1. *Clark, J., Yallop, C. An Introduction to Phonetics and Phonology / J. Clark, C. Yallop. – Oxford & Cambridge, MA: Basil Blackwell, 1994. – 400 p.*
2. *Бондарко, Л. В. Фонетическое описание языка и фонологическое описание речи / Л. В. Бондарко. – Ленинград : Изд-во ЛГУ, 1981. – 199 с.*
3. *Вербицкая, Л. А. Звуковые единицы русской речи и их соотношение с оттенками и фонемами : автореф. дисс. ... канд. филол. наук / Л. А. Вербицкая. – Ленинград : Изд-во ЛГУ, 1965. – 19 с.*
4. *Златоустова, Л. В. Фонетические единицы русской речи / Л. В. Златоустова. – М.: МГУ, 1981. – 108 с.*
5. *Зиндер, Л. Р. Общая фонетика : учеб. пос. / Л. Р. Зиндер. – 2-е изд., перераб. и доп. – М. : Высшая школа, 1979. – 312 с.*
6. *Торсуев, Г. П. Константность и вариативность в фонетической системе / Г. П. Торсуев. – М. : Наука, 1977. – 125 с.*
7. *Щерба, Л. В. Избранные работы по русскому языку / Л. В. Щерба. – М. : Учпедгиз, 1957. – 188 с.*
8. *Zinovieva, N. Phonetically Sufficient Allophonic Database for Concatenation Synthesis of Russian Speech // Proc. of XIIIth International Congress of Phonetic Sciences. – Stockholm, 1995. – Vol. 2. – P. 358–362.*
9. *Зиновьева, Н. В. Программный синтез русской речи (синтезатор «Агафон») / Н. В. Зиновьева, О. Ф. Кривнова, Л. М. Захаров // Труды международного семинара «Диалог'95» по компьютерной лингвистике и ее приложениям. – Казань: Полиграф, 1995. – С. 121–127.*
10. *Лобанов, Б. М. Синтезатор речи по тексту как компьютерное средство «клонирования» персонального голоса / Б. М. Лобанов, Е. Б. Карневская, Т. В. Левковская // Труды Международной конференции «Диалог-2001». – М., 2001. – С. 265–272.*

11. *Lobanov, B.* TTS-Synthesizer as a Computer Means for Personal Voice Cloning (On the example of Russian) / B. Lobanov, H. Karnevskaia // *Phonetics and its Applications* / P. Sojka, I. Kopecek, L. Pala (Eds.). – Stuttgart : Steiner, 2002. – P. 445–452.
12. *Skrelin, P.* Allophone- and suballophone-based speech synthesis system for Russian // *Proceedings of the Third International Workshop on Text, Speech and Dialogue*. – Brno, 2000. – P. 271–276.