

РУМЯНЦЕВ М. К.

СИНТЕЗ КИТАЙСКИХ ТОНОВ

В китайском языке и в других языках изолирующей типологии чрезвычайно большая роль принадлежит слоговому тону. Тон — единица фонологическая: он различает слогоморфемы. Но функции тона не замыкаются только на слогоморфеме. Тоны как просодическое средство организуют всю просодическую систему языка: они создают тоноритмику китайских слов и словоподобных образований, на их основе осуществляются все интонационные дифференциации — синтаксические, коммуникативные, модальные, эмоциональные, стилевые. Все названные просодии модифицируют физическую структуру тонов. В результате тон в речи всегда оказывается единицей интегральной, несущей в себе не только признаки слогового тона как такового, но и всех других просодических наложений. Естественно, что при первичном синтезе тонов нас интересовал, прежде всего, тон слоговой — исходная его структура, дифференцирующая слогоморфемы, а не возможные его варианты с интегральными признаками. Следовательно, в программы синтеза должны быть заложены все функционально значимые (в плане различения морфем) признаки тонов — контурные, регистровые, интервальные, амплитудные, временные.

Но если в естественной речи чистый тон без интонационных накладок получить в принципе нельзя, то в синтезе получение такого (чистого) тона оказывается методически невыгодным. В естественной речи нельзя произнести просто тонированный слог вне интонации называния, поэтому даже реализация тона в слогоморфеме в качестве самоназвания включает в себя некоторые признаки назывной повествовательной интонации. А в синтезе? Кажется, можно ведь не программировать интонацию называния при искусственном создании тона? Можно, но тон без нее не воспринимается как естественный, аудиторам становится трудно оценить его по качеству и собственно тональным признакам, поэтому во многих случаях при работе как с естественной, так и с искусственной речью назывная интонация служит хорошей моделью для реализации собственно тональных признаков: она может быть модально и эмоционально нейтральной, интонационные наложения на слоговой тон можно свести к минимуму. Именно такие интонационно нейтральные искусственные реализации тонов мы и старались получить в процессе первичного синтеза китайских слогов¹. Оценку качества синтезируемых слогов производили два аудитора-китайки. Слоги признавались хорошими, если они правильно и однозначно опознавались с первого предъявления и квалифи-

¹ Синтез осуществлялся в Лаборатории экспериментальной фонетики ИСАА при МГУ на формантном синтезаторе СПИИ-75 (см [1]). Китайские гласные и тоны в синтезе были получены также американским исследователем Дж. М. Хауи [2]. Однако в его публикации рассматривается лишь один параметр тона — F_0 (в том или ином времени звучания), что, безусловно, суживает общую проблематику тона, как естественного, так и синтезированного.

цировались как нормативно китайские слоги. Естественность (человечность) звучания в нашем синтезе, как правило, не достигалась. Явно машинные реализации отсеивались. Слоги с некоторой машинной окраской в звучании (как правило, незначительной), но удовлетворявшие названным требованиям, включались в общий корпус синтезированных слогов.

Первый тон — ровный по частотному контуру (безынтервальный) и высокий по регистру — в нашем синтезе являлся исходным. Все слоги были синтезированы сначала в первом тоне, который создавался, как правило, константным значением F_0 на протяжении звучания всей финали слога².

При синтезе первого тона важно было создать в высоком регистре достаточно продолжительный участок звучания, который на слух воспринимался бы как ровный — безынтервальный. Время звучания этого участка перцептивно значимо; несоблюдение его временной нормы вызывает у аудиторов негативную реакцию. Если даже тон хорошо опознается, то констатируется его ущербность по времени звучания. Самым же неожиданным явилось то, что некоторые синтезированные тоны, получив аудиторскую оценку «хороший тон», были снабжены оговоркой: «но лучше бы чуть продлить его». И эта оговорка относилась к таким программам синтеза, которые предопределяли вполне достаточное время звучания тона. Например, в программе синтеза слога *tōu* было предусмотрено 415 мсек звучания финали. Это время не только достаточное, но и, казалось бы, оптимальное даже для внеконтекстной реализации слога. В естественной речи именно такое, или даже меньшее, время звучания первого тона мы и наблюдаем в оптимальных реализациях. В синтезе же оно оказалось недостаточным для безоговорочного приятия тона аудитором. Объясняется это, видимо, тем, что несмотря на большое общее время звучания финали и константное значение F_0 на протяжении всего звучания (170 Гц), высокий, ровный и, главное, достаточно продолжительный участок звучания в слоге создан все же не был. В значения самой формантной структуры на соответствующих участках, видимо, не были заложены какие-то параметры, которые бы «работали» на тон и вместе с частотой F_0 создавали фонологически значимый участок звучания.

В звучании финали слога первого тона могут быть участки, до некоторой степени нарушающие его общий ровный контур, образуются даже иногда нисходящие и восходящие интервалы, но эти нарушения не должны выходить за пределы допустимых; скольжение частоты F_0 образующее те или иные интервалы (в конце звучания или в начале), не должно приближаться к той критической зоне, в которой начинаются другие — интервальные тоны: второй, четвертый, третий. Кроме того, эти скольжения не должны мешать образованию обязательного участка ровного контура тона.

При аудировании реализаций некоторых программ первого тона аудиторы отметили небольшой «завал» (падение) тона в конце финали. Сравнение аудиторских показаний со значением F_0 в программах этих реализаций показало, что действительно во второй половине звучания финали частота F_0 была несколько меньшей. В слоге *sōng*, например, этот «завал» образовывался разницей в 6 Гц, приходящейся на всю вторую половину звучания финали (166—160; 1,03)³. В слоге *zāi*, признанном

² О синтезе гласных и согласных сегментов слогов см. [3, 4].

³ Здесь и в дальнейшем в скобках показаны частотные значения F_0 (в герцах — Гц) и величина отношения большей частоты к меньшей (интервал).

по качеству его сегментов хорошим, а инициаль даже очень хорошей — естественной, первый тон одним аудитором был забракован, а другой воспринял его не как первый, а как второй. Обращение к программе этого слога показало, что на участке звучания финали была задана частота F_0 с небольшим варьированием значений: 133—130, а полувзвонкая инициаль z начинается с частоты 109 и продолжается с частотой 111 Гц. Таким образом, между началом и концом звучания слога создался восходящий интервал в целую терцию, который и был одним из аудиторов ассоциирован со вторым тоном.

В связи с оценкой аудиторами данной программы синтеза первого тона необходимо поставить проблему граничной интервальной зоны первого тона: какой восходящий (нисходящий) интервал не воспринимается еще как второй (четвертый) тон? Проблема эта сложная и специальными экспериментами в нашем синтезе не решалась, хотя синтез и предоставляет исследователю здесь большие преимущества: любые частотные интервалы с любым шагом их увеличения, начиная от нулевого, можно проверить на восприятие при одном и том же времени звучания, и один и тот же интервал при разном времени звучания. И, наконец, что также очень важно, эти проверки можно осуществить при однотипном и разнотипном распределении амплитудных значений. Естественная речь таких возможностей не предоставляет. А проверки эти в эксперименте синтезом нужны, поскольку известно в настоящее время, что существует некая зависимость, значимая для восприятия, между частотным интервалом и временем, за которое этот интервал образуется.

При выяснении граничных зон интервалов существенно также учитывать амплитудные распределения: образуется ли данный интервал у первого тона при амплитудном распределении, характерном для первого тона (равномерное распределение амплитуд), или амплитуды распределены по-другому — большие амплитуды смещены, например, к концу звучания. При таком смещении меньшие частотные интервалы или даже совсем безынтервальные реализации слогов могут восприниматься как тоны восходящие, а не ровные. Так, в нашем примере синтеза первого тона в слоге *zai* неудача может объясняться не только тем, что в слоге от его начала к концу образовывался восходящий частотный интервал, но и тем, что в программе не был соблюден унисон признаков, характерный для первого тона: ровный частотный контур — равномерное распределение амплитуд. В программе фактически создался унисон, характерный для второго тона: восходящий интервал сопровождается усилением амплитуд от начала звучания слога к концу.

Четыре программы первого тона с финалью *uap* один из аудиторов не воспринял как таковые: тон для него остался неясным, несмотря на то, что программы этих слогов предусматривали постоянную частоту F_0 во всей финали или даже во всем слоге. Объясняется это, по-видимому, тем, что основная частота в этих слогах была задана довольно низкой (139 Гц), а регистр первого тона — высокий, и аудитор по реализациям отдельных слогов не смогла составить для себя общую регистровую шкалу тонов для данного «голоса».

Второй тон — по частотному контуру F_0 — восходящий. Его начало на шкале регистровых уровней помечается цифрой 3, обозначающей средний регистр, а конец — цифрой 5, символизирующей высокий регистр. Следовательно, в синтезе второго тона основная задача заключалась в том, чтобы между началом его звучания и концом создать тот частотный интервал, который воспринимается в данной языковой системе

как восходящий тон. Достигалось это направленным варьированием значений F_0 , такой программой повышающихся частотных значений, которая и образует заданный интервал. Интервал тона задан языковой системой и может варьировать лишь в определенных пределах.

В нашем синтезе не решается задача установления граничных зон интервалов второго тона, но мы получили, однако, такие варианты интервалов, которые весьма близки к минимальным. Так, например, в слогах *ní*, *nín*, признанных аудиторами «естественными» (неотличимыми от человеческих), второй тон создан перепадом частот F_0 от начала звучания к концу всего лишь в 1 секунду (153—170; 1,14). На участке инициали (105 мсек) значение F_0 в этих слогах константно (153), затем за 640 мсек частота F_0 возросла до 170; в слоге *nín* это значение осталось неизменным и на всем участке конечного назального элемента (140 мсек). Второй тон с интервалом в секунду создан и в слоге *háí*. Интервал этот образован в более низком частотном диапазоне (139—155) за практически такое же время, как и в слогах *ní* и *nín* — 660 мсек.

Очень слабый второй тон одна из аудиторов услышала в слогах *láí* и *báí*, в программах которых был задан восходящий интервал F_0 лишь в м.секунду (130—139; 1,06). Другой аудитор тон в слоге *báí* приняла за третий. Этому способствовало заложенное в программе начальное падение частоты F_0 с таким же интервалом, как и последующее повышение: 139—130. Полный контур тона, таким образом, оказался похожим на нисходяще-восходящий контур третьего тона: 139—130—139. В действительности же данные реализации не представляют четко ни второго, ни третьего тонов, поэтому вне контекста они и могут быть восприняты как весьма слабые варианты второго либо третьего тонов. Для акустически четкого второго тона в них мал восходящий интервал, для третьего — недостаточен «нажим» в первой части его звучания.

В слоге *tán*, признанном неотличимым от естественного, за 640 мсек синтезирован интервал второго тона в ум. квинту (1,42). Судя по естественным реализациям, этот интервал и близкие к нему являются оптимальными для внеконтекстных произнесений слогов второго тона.

Восходящий интервал второго тона в естественной речи создается обычно в двух вариантах: в варианте прямого скольжения частот F_0 вверх от какой-то начальной точки и в варианте циркумфлексном (с ложбинкой), когда в начале звучания тона образуется некоторый участок небольшого падения частот, после которого идет прямое скольжение вверх. В фонологическом аспекте начальный участок падения частот избыточен, но фонетически он оказывается нередко необходимым; это как бы разгон, подготовка «стартовой» точки для последующего скольжения частоты вверх. Тон при таком «старте» приобретает особую выразительность. Чаще он встречается в слогах без инициалей или с сонорной инициалью. В синтезе представлены два варианта второго тона: без «ложбинки» и с «ложбинкой». В слоге *á* начальное падение частот F_0 (139—130) происходит в первые 225 мсек звучания, последующее скольжение вверх (130—166) образует восходящий интервал в б. терцию (1,27), воспринятый как «очень хороший» второй тон. В слогах *ní*, *nín*, *yí* на начальном участке тона не происходит никакого падения частот, «стартовая» площадка оказалась ровной, но это не привело к ухудшению воспринимаемых качеств тонов. Все они восприняты аудиторами как «очень хорошие»; тон слога *yí* признан выразительным, а слоги *ní* и *nín* квалифицированы даже как неотличимые от естественных.

В оптимальном своем варианте второй тон, как и другие интервальные

тоны (3-й, 4-й), создается согласованным действием частотных и амплитудных распределений: повышение значений F_0 от начала звучания тона к концу сопровождается повышением амплитудных значений. Реально же, в речи, в зависимости от разных факторов, главным образом, интонационных, возникают и тоны с той или иной степенью рассогласования частотных и амплитудных характеристик. Определить, какая степень рассогласованности частотных и амплитудных характеристик допустима, а во многих случаях необходима, в данной языковой системе, — задача чрезвычайно важная и теоретически, и практически. Теоретический аспект связан здесь с выявлением сложных взаимоотношений тона и интонации (во всех ее проявлениях), а практический — с возможностью тонкого моделирования тонов и их вариантов, пригодных для использования в программном синтезе речи. Интерпретация результатов синтеза второго тона, особенно программ «неудачных», предоставляет поучительный материал. Так, в программе синтеза слога *yi* при очень хорошем сегментном качестве слога второй тон был воспринят аудитором как первый, хотя частотный интервал F_0 был задан для второго тона вполне достаточный (1,2); время звучания также было для второго тона оптимальным (305 мсек). Восприятие второго тона как первого при восходящем частотном интервале, равном б.терции, объясняется амплитудно-частотной рассогласованностью данной программы: частоты F_0 от начала звучания к концу повышаются (от 150 до 181 Гц), а амплитуды F_1 постоянны на всем протяжении звучания слога, т. е. создают модель не второго, а первого тона. В F_2 во второй (бóльшей) половине звучания тона (200 мсек из 305) амплитуды также постоянны. Такое распределение амплитуд и гасит частотный интервал. При изолированном восприятии тона амплитудная информация оказалась для аудитора более сильной, нежели частотная. Второй тон в синтезированном слоге *máo* был воспринят аудитором как первый, хотя частотный интервал был еще бóльшим — равнялся кварте (1,35) при времени звучания 590 мсек. Причина такого восприятия та же: не характерное для второго тона распределение амплитуд. В F_1 на большом протяжении звучания (415 мсек) амплитуды высокие и постоянные, действуют по модели первого тона; эта тенденция поддерживается и амплитудами F_2 . При изолированном восприятии тона и здесь амплитуды гасят интервал, образованный F_0 , причем интервал большой. Меньшие интервалы гасятся легче. Следовательно, когда речь идет об определении граничных частотных интервалов, надо иметь в виду не только сам интервал и время, за которое он образован, но и форму (модель) амплитудного распределения при его образовании.

При психолингвистических экспериментах на восприятие часто приходится удивляться тому, как чутко ухо, воспитанное тональной языковой системой, улавливает внутренний, нередко противоречивый механизм тональных признаков. Хорошим примером этого является программа синтеза слога *qiáng*, которая при восприятии была квалифицирована как нечто среднее между первым и вторым тоном. И действительно, на протяжении 300 мсек в конце звучания тон в этой программе безынтервален: в течение 220 мсек представлен значением 181 Гц, в предшествующие 80 мсек образуется некоторый интервал (170 и 175 Гц), но он не воспринимается как восходящий — гасится равнонаправленными амплитудами F_1 ; они неизменны на всем (300 мсек) участке звучания и поддерживают безынтервальное движение F_0 . Все это и воспринимается аудитором как первый тон. Но в первой части звучания (в течение 160 мсек) программа предусматривает увеличение амплитудных значений, за это время обра-

зуется и частотный интервал (1,25), который работает уже в пользу второго тона. Эту раздвоенность в просодическом звучании аудиторы улавливают, и это безусловно свидетельствует о достаточно надежной соотнесенности восприятия и физических характеристик речевого сигнала. Уместно обратить внимание на то, что раздвоенность характеристик в речевом сигнале и разного рода их несогласованность, которые мы получаем в синтезе, оказываются для лингвиста-экспериментатора чрезвычайно информативными. Способом намеренного рассогласования физических характеристик синтезируемой речевой единицы можно выявить роль каждого ее параметра, выявить возможные пределы этой рассогласованности, как в плане универсальном, так и специфичном для данной звуковой системы. Должно иметь в виду, однако, что работа с тонкой физической структурой речевого сигнала даже на уровне синтеза — дело весьма трудное. Особенно трудно выявить тонкое структурирование одного и того же признака в зависимости от разной функциональной предназначенности составляющих его микроструктур. Амплитуды, например (признак интенсивности), в каких-то своих пропорциях входят в саму структуру изначального спектра звука, образующего то или иное его качество, но это качество может быть реализовано с большей или меньшей силой; образуется, следовательно, своя, другая пропорция этого признака, обеспечивающая не качество как таковое, а реализацию этого качества по шкале громкости. В китайском слоговом тоне амплитуды также работают двунаправленно: они входят в состав спектра, создающего то или иное качество финалей (*yī, wǐ, é* и др.), и модифицируют эти финали в тоновом плане, подстраивая свои значения под значения F_0 . Здесь именно и возникает проблема согласованного и несогласованного действий F_0 и амплитудных характеристик формант.

Двунаправленными являются и действия F_0 . Значения этого параметра как обязательные компоненты входят в изначальный спектр финали слога, определяющий ее качества, и они же (значения F_0), структурируясь определенным образом, представляют это качество в тоновом аспекте, формируют тон как определенную регистрово-контурную единицу.

Взаимосвязь качественного и тонального аспектов в амплитудных параметрах сигнала хорошо прослеживается на примере синтеза слога *ér*, признанного аудиторам неотличимым от естественного. В этой программе за 455 мсек звучания слога образовался оптимальный частотный интервал второго тона (127—172; 1,35). Амплитуды F_1 действовали в унисон с F_0 в первой части звучания на протяжении 235 мсек, затем их значения резко снижались. Это снижение амплитудных значений приходится на участок эризации финали. И здесь остается неясным: работали ли амплитуды на этом участке только на эризацию или же и на тон тоже? F_0 и на этом участке работала на тон. Вполне возможно, что амплитудные характеристики также в этом участвовали, хотя общее движение амплитудных значений по всему звучанию и не было однонаправленным. При качественно однородном звучании финали (без эризации) амплитуды F_1 вслед за F_0 образовывали нормальную для второго тона восходящую линию значений.

Финали китайских слогов в разных тонах на слух воспринимаются как несколько различающиеся по своему качеству. Различия эти в разных тонах и разных финалях проявляются в неодинаковой степени. Но в любом случае важно установить, появляются ли эти различия в качестве финалей в результате различий в их спектральной картине или же значения всех формант остаются неизменными во всех тонах, а изменя-

ются лишь значения F_0 , и, как следствие этого, на разных участках звучания слога возникают разные соотношения составляющих F_0 с параметрами других формант, что и приводит к некоторому общему сдвигу в спектрах финалей и, следовательно, к разным оттенкам их качества в восприятии. Вопрос этот не только академический, поскольку разные оттенки финалей разных тонов создают собственно китайскую их специфику, т. е. для китайского восприятия функциональны и, следовательно, в синтезе должны быть смоделированы⁴. Моделирование же призвано вскрыть самый механизм указанного различия, т. е. ответить на вопрос, нужна ли тонкая коррекция исходных формантных значений слога (полученных в первом тоне) или же эта коррекция происходит автоматически — самим изменением значений форманты F_0 ?

При синтезе второго тона в слове *tán*, признанном неотличимым от естественного, никакой коррекции частотных и амплитудных формантных значений произведено не было: F-картины в первом и втором тонах оказались тождественны. В первом тоне этого слога значение F_0 было константным — 170 Гц на протяжении всего времени звучания, во втором тоне скольжением частот F_0 вверх образовывался интервал в ум. квинту (139—198; 1,42). При создании малого интервала второго тона в слове *hái* (139—155; 1,14) частотных и амплитудных формантных коррекций также не производилось. Можно, по-видимому, предположить, что в принципе и в естественных произнесениях слога первого и второго тонов могут создаваться при относительном тождестве их спектральных картин. Другое дело, что такая ситуация в естественной речи практически может и не реализоваться или реализоваться крайне редко. Человеческие органы речи — синтезатор несравненно лучший, чем любой из ныне существующих машинных, в человеческих произнесениях возможны такие тонкие коррекции спектров слогов, которые мы не можем еще моделировать на машинах. Примеры тонких различий в естественных спектрах финалей разных тонов мы наблюдали при спектральном анализе их произнесений тремя дикторами. В финали *í*, например, второй восходящий тон тянет первую форманту вверх по мере возрастания значений F_0 от начала звучания к концу. У диктора 1 (муж.) финаль *í* во втором тоне начинается со значения F_1 в 250 Гц, ко второй половине звучания значения F_1 возрастают до 300 Гц, и заканчивается финаль частотой F_1 в 350 Гц. У диктора 2 (жен.) повышение значений F_1 от начала звучаний финали *í* к концу еще заметнее: начало — 350 Гц, затем — 400 и конец — 500.

Анализ программ синтезированных слогов с финалью *í*, признанных аудиторами натуральными, — неотличимыми от естественных, показывает, что их натуральность как раз и обусловлена тонкой коррекцией (подстройкой) частотных и амплитудных значений F_1 и F_3 под значения F_0 , согласованным действием всех составляющих спектра. Мера и конкретные пропорции этого согласования не являются универсальными, а определяются языковой системой, и направлены они, прежде всего, на создание нормативного (приемлемого в данной системе) качества звучания. Но достижение этой цели не случайно, по-видимому, оказывается сопряженным и с получением звучания натурального или близкого к таковому.

Впечатление натуральности возникает потому, что в спектре синтезированного гласного снимается монотонность (машинность) звучания,

⁴ Функциональная значимость указанных различий подтверждается, в частности, и тем, что некоторые авторы предлагают рассматривать китайские гласные в разных тонах как разные фонемы.

оно, как и в естественной речи, становится неоднородным по качеству на разных участках звучания, развивается от начала к середине и к концу. F_1 , например, согласуемая по правилам системы с F_0 и частотно, и амплитудно, оказывается представленной не одним значением на протяжении всего звучания, а целым рядом согласованных значений внутри самой форманты и межформантно. В процессе согласования значений на каждом данном участке звучания синтезируемому сигналу фактически задаются частотные и амплитудные «микровариации». Но эти вариации не универсально-человеческие, обусловленные свойствами речевых органов, а системно-языковые, присущие данной фонетической норме.

В естественной речи регистрируется, как известно, и такое микроварьирование, которое не оказывает заметного влияния на качество звуков в системно-языковом плане, но придает им человеческие свойства. В нашем синтезе такая универсально-человеческая вариативность не моделировалась. Существенно, однако, то, что звучание естественное или близкое к таковому в синтезе может быть получено только за счет вариаций характеристик, обеспечивающих пормативное качество звучания.

Т р е т ь и й т о н — нисходяще-восходящий по частотному контуру F_0 . Его начало на шкале регистровых уровней помечается цифрой 2, а конец — цифрой 4. В контуре третьего тона фиксируются обычно три участка: начальное падение частот F_0 , ровный участок и скольжение вверх. Падение частот F_0 большее и более продолжительное, чем во втором тоне, когда последний реализуется в циркумфлексной форме. Возможны также варианты третьего тона без начального падения или с очень небольшим падением. В этих случаях вся первая половина или даже большая часть звучания тона реализуется в низком регистре; во второй половине звучания происходит повышение частот F_0 .

В восприятии третьего тона — его первой половины, ощущается как бы нажим, который в сочетании с последующим повышением частот и создает специфическое качество третьего тона. В синтезе создание этого «нажима» достигалось значительным понижением частот F_0 . В слове *yí*, например, нисходящий интервал составил терцию (149—123; 1,2) и образован он был за 540 мсек, т. е. время, составившее почти две трети всего звучания слога. Интервал образовывался медленно, с затянкой. На участке в 315 мсек возникал практически ровный контур F_0 . Третий — завершающий участок тона с восходящим интервалом в 6. секунду (123—139; 1,12) был создан за 300 мсек конечного звучания. Некоторые авторы полагают, что впечатление нажима в третьем тоне создается максимум мускульной напряженности, приходящейся на низкую часть тона [5]. Следует при этом иметь в виду, что ощущение мускульной напряженности на этом участке третьего тона соотносится не с большими, а наоборот, с меньшими амплитудами интенсивности. Из двух программ третьего тона, тождественных друг другу по всем параметрам, кроме амплитудных, как безусловно лучшая аудиторами была признана программа, в которой на участке ровного контура были меньшие амплитуды, чем на предшествующем участке. Это четко прослеживается по амплитудным значениям и F_1 , и F_3 .

Перцептивно качество финалей в третьем тоне заметно отличается от такового в других тонах. Физическую основу этого отличия составляют не только иные, чем в других тонах, распределения частот F_0 , но и распределение амплитуд. Так, например, если в первом, втором и четвертом тонах разные оттенки в качестве финали в синтезе были созданы только за счет различного распределения частот F_0 , то в третьем тоне к этому

обязательно добавляется еще иная, чем в других тонах, раскладка амплитудных значений, т. е. происходит заметный и закономерный сдвиг в общем спектре финалей.

Аудиторское восприятие синтезированных третьих тонов выявило ряд особенностей в их физической структуре, которые оказывают решающее влияние на их идентификацию. Прежде всего, начало третьего тона и вся последующая его часть — вплоть до конечного повышения частот — должны располагаться в зоне низкого регистра пятиуровневой китайской шкалы. Смещение этих участков тона вверх приводит к тому, что весь частотный контур тона, хотя в целом и оказывается правильным, но как третий тон не опознается, квалифицируется как регистрово ущербный. С такой ситуацией мы сталкиваемся в программе синтеза слога *bī*. При реализации этой программы слог опознается правильно, сегменты признаются нормативными, а тон квалифицируется как плохой — высокий по регистру. Реализация программы слога *hǎi*, в которой тон также смещен по регистру вверх, признана аудитором тоном кантонским, а не пекинским. Завышение начала тона в некоторых наших программах приводило к тому, что это начало третьего тона опознавалось аудитором как четвертый тон, а за этим четвертым тоном шла восходящая часть третьего тона, тон слога раздваивался и в целом признавался плохим. Так именно была квалифицирована реализация программы слога *hǎo*, в которой в нисходящей части первого тона (345 мсек) образовался интервал несколько больший, чем кварта (1,36). Этот интервал поддержан и убывающим распределением амплитуд F_1 ; такое звучание и было воспринято как четвертый тон. Образовавшийся же интервал F_0 последующей части звучания слога (120 мсек), равный м. терции (1,19), производил на аудитора впечатление нормального окончания третьего тона. Все это совмещалось в звучании одного слога. Тона с такими характеристиками в пекинской системе нет, поэтому он и был признан в целом плохим.

Другая опасность синтезировать неправильный третий тон кроется в образовании ненормативного для третьего тона восходящего интервала в конце звучания слога. Так, например, третий тон в слове *nǎ* был воспринят как второй. Способствовала этому такая просодическая структура слога, в которой восходящий интервал F_0 имел все признаки второго тона: он был достаточно большим (1,18), формировался за время намного большее, чем интервал нисходящий (нисходящий — за 320 мсек, восходящий — за 560), и, главное, в движении частоты F_0 в первой части звучания не было предусмотрено большее понижение частоты; понижение с интервалом 1,09 (139 : 127) оказалось для третьего тона недостаточным, в начале звучания нужен больший интервал, а в конце — меньший. Строго говоря, нужен не сам интервал — не его величина важна — необходим более низкий регистр для второй части третьего тона. Выхода же в низкий регистр рассматриваемая программа не обеспечивала, создалось лишь небольшое понижение, напоминающее «ложбинку» в начале второго тона. Все это и предопределило аудиторское восприятие тона.

Восприятие третьего тона как второго свидетельствует о том, что восходящая часть третьего тона по своему качеству (сегментному и тональному) не соответствует в программе модели третьего тона. Подобно тому, как начальная нисходящая часть третьего тона по своему качеству не должна быть похожа на четвертый тон, так же и конечная, восходящая часть третьего тона не должна повторять характеристики второго тона. Начальное падение частоты F_0 в третьем тоне медленное и сдавленное (с нажимом), не такое резкое, как в четвертом тоне, а восходящая часть —

не такая свободная в своем скольжении вверх, как во втором тоне, она тоже сдвленная и как бы закручивается в конце, образуя в завершении нечто похожее на неполный glottal stop. Полная глоттальная смычка образуется иногда в конце первой (нисходящей) части третьего тона. В таких случаях восходящая часть начинается с прорыва этой смычки. Все это и составляет специфику третьего тона, которая в синтезе должна моделироваться. Если не удастся воспроизвести эту специфику, тон получается дефектный.

Третий тон реализуется, как известно, в двух вариантах: в полном (циркумфлексном) и неполном (усеченном). Неполный вариант представлен только первой — нисходящей своей частью. Этот вариант третьего тона реализуется в слогоморфемах, занимающих позицию перед любым другим тоном, кроме третьего. В синтезе рассматриваемый вариант тона получен в позиции перед четвертым тоном в слове *q'iyi*. Тоноритмика слова признана нормативной. Усеченный третий тон характеризуется здесь нисходящим интервалом в м. терцию (134 : 114), образованным за 310 мсек. Интервал этот небольшой, но он затяжной, и выводит тон в низкий регистр. Затяжка же формируется временным фактором: в течение последних 100 мсек частота F_0 практически не меняется (117 и 114 Гц), возникает ровный участок тона. Все это и создает типичную модель неполного третьего тона, воспринимаемую в слове *q'iyi* на фоне сильного четвертого тона с интервалом почти в целую квинту ($162 : 110 = 1,47$).

В программе слога *b'i* неполный третий тон был получен и вне контекстного окружения (вне слова) как самостоятельная просодическая единица. Этот тон вне контекста был и опознан аудитором как таковой — неполный. Следовательно, характеристики его имеют не только относительную, но и абсолютную языковую ценность: не всякий нисходящий интервал воспринимается в данной языковой системе как четвертый тон; важен не только сам интервал (его величина), но и частотный регистр, в котором этот интервал образуется, специфические особенности образования нисходящего интервала. Существенным является, происходит ли свободное и резкое скольжение частоты F_0 вниз или это падение затяжное (замедленное) и сдвленное.

Ч е т в е р т ы й т о н — нисходящий по частотному контуру F_0 . Его начало на шкале регистровых уровней помечается цифрой 5, а конец цифрой 1. Эти пометы определяют четвертый тон как акустическую структуру с большим частотным интервалом между его началом и концом. Действительно, оптимальные для внеконтекстного восприятия и подчеркнuto четкие реализации четвертого тона характеризуются интервалами, близкими к квинте или даже сексте. В его синтезе основная проблема заключалась в том, чтобы создать такой интервал и задать соответствующие этому интервалу амплитудные значения. Четвертый тон по времени звучания самый короткий, и, следовательно, большие интервалы должны создаваться за меньшее, чем в других тонах, время. Так, например, интервал четвертого тона в б. сексту (168—100; 1, 68) был создан в слоге *yi* за 170 мсек, и слог этот был признан неотличимым от естественного. Хорошими признаны также слоги четвертого тона с интервалами большими, чем в кварту. Нисходящие интервалы F_0 в терцию и даже секунду уже производили у аудиторов впечатление четвертого тона. Но такие интервалы, особенно секунда, являются, видимо, граничными либо близкими к таковым. Тон опознается, но всегда сопровождается оговоркой: «не дотянут по интервалу». Самый малый интервал четвертого тона создан нами в слоге *si*, в котором частота F_0 от начала звучания финали к концу изменялась

лишь на 12 Гц, т. е. был образован интервал лишь несколько больший, чем м. секунда.

В синтезе четвертого тона, пожалуй, еще в большей степени, чем в других тонах, необходимо соблюдать унисон (согласованное действие) характеристик. Особо важным фактором является согласование частотных значений F_0 и амплитудных характеристик. Понижение частотных значений F_0 от начала звучания к концу — образование большого нисходящего интервала — должно сопровождаться значительным спадом амплитудных значений. Несоблюдение этой согласованности приводит к ущербу тона: тон начинает восприниматься как инертный, вялый, недостаточный по интервалу, хотя интервал сам по себе вполне достаточный. Так именно воспринимался четвертый тон в слогах *шй*, *би*, *ти*. Частотные интервалы тонов в этих слогах составили: м. сексту и почти квинту. Эти интервалы вполне достаточны для самых оптимальных реализаций четвертого тона, а все они квалифицированы аудитором как недостаточные — инертные. Их инертность с очень большой вероятностью может быть объяснена несогласованным действием амплитудных характеристик. В программе слога *шй* на протяжении 240 мсек звучания (во второй его половине) из 340 мсек амплитуды F_1 оставались неизменными и были по значению своему минимальными. Такими же постоянными минимальными амплитудами они были и в F_2 на протяжении еще большего отрезка звучания. За это время образовывался частотный интервал тона, почти равный ум. квинте (1,39), а амплитуды действовали не по модели четвертого тона, а по модели первого. Переход со средних амплитуд на минимальные после первых 100 мсек звучания слога оказался недостаточным для создания амплитудной модели четвертого тона. В двух программах слога *хй* на всем протяжении звучания слога не было предусмотрено никакого амплитудного скольжения (от больших значений к меньшим), а частотный интервал был задан больший, чем в ум. квинту. И этот интервал оказался погашенным не характерным для четвертого тона амплитудным распределением: тон был признан плохим; аудитору показался недостаточным интервал падения тона.

Инертность (вялость) китайского четвертого (нисходящего) тона объяснялась ранее мною и некоторыми другими авторами недостаточной скоростью образования нисходящего интервала [6]. Новые наши данные показывают, однако, что это объяснение весьма уязвимо. Анализ программ синтеза нормативных и вялых (инертных) тонов свидетельствует, что сам по себе параметр скорости образования частотного интервала не может быть причиной инертности тона. Четвертый тон, хотя и является самым коротким из всех, но и он варьирует по времени звучания в достаточно широких пределах. В разных условиях речи он может звучать 150—200 мсек и даже меньше, но может занимать и время в 400 или даже 500 мсек. А это приводит к тому, что параметр скорости образования частотного интервала оказывается величиной, в очень большой степени зависимой от времени звучания тона. И действительно, в нашем синтезе нормативные тоны были получены при небольших скоростях образования частотных интервалов, а при большей скорости тоны создавались инертные, вялые, хотя, казалось бы, большая скорость образования интервала должна была работать против инертности тона. При одном и том же времени звучания тона и одном и том же частотном интервале, т. е. при одной и той же скорости образования интервала, были получены как плохие (инертные) тоны, так и хорошие — нормативные. Так, в программах слогов *jän* и *pän* четвертый тон реализовался за 420 мсек. За это время в том и другом слоге образовывал-

ся частотный интервал тона равный 1,37. Следовательно, за 1 мсек создавался интервал в 0,0032 (1,37 : 420). Это и есть скорость образования частотного интервала данного тона в единицу времени (1 мсек). Она одинакова и в слове *fàn*, и в слове *pàn*. Однако тон в слове *fàn* признан нормативным, а в слове *pàn* — вялым. Отсюда следует, что не скорость образования частотного интервала тона надо винить в его инертности, по крайней мере, не общую скорость, связанную с временем звучания всего слога. Возможно, какие-то аспекты скорости и могут быть значимы для восприятия тона, но нашими данными они не выявляются.

Предпринятое исследование в целом можно рассматривать как попытку создать акустический алфавит для синтеза китайских тонов, а в принципиальных своих посылах и не только китайских, привлечь внимание к общим проблемам акустического моделирования просодии слога как базисной единицы всего просодического яруса китайского языка.

ЛИТЕРАТУРА

1. Скрипкин Г. Н. Краткое описание синтезатора речи СППИ-75 и принципов программирования // Лингвистическая интерпретация результатов экспериментально-фонетических исследований речевого текста: Тез докл. Минск, 1977.
2. Howie J. M. Acoustical studies of Mandarin vowels and tones Cambridge, 1976.
3. Румянцев М. К. Синтез китайских слогов (финали) // Проблемы восточной филологии М., 1979.
4. Румянцев М. К. Синтез китайских слогов (инициали) // ФН. 1978. № 5.
5. Задосенко Т. П., Хуан Шуан Учебник китайского языка М., 1973.
6. Гань Аошун. Тональные контуры в кантонском диалекте // Вестник МГУ. Востоковедение. 1972. № 1.

