Поговори со мною, робот!

тест редакции

Продолжение. Начало в №5/2021

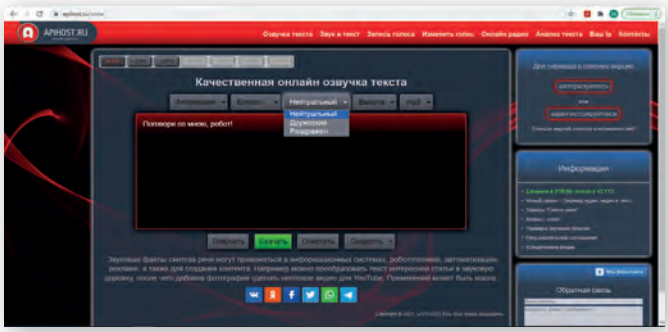
Арсений Ворошилов

Интернет-просторы оказались на удивление многоголосыми в том смысле, что в сети есть довольно много онлайновых сервисов для преобразования текста в речь. Но, как показывает практика, много – далеко не всегда значит хорошо. Забегая вперед, скажу, что бывает и очень даже неплохо. Но об этом ниже.

В предыдущей – первой – части цикла речь шла о сервисе Polly, который предоставляет AWS (Amazon Web Services). Но есть подобные ресурсы и в русскоязычном сегменте Всемирной паутины, причем заметно превосходящие по эффективности сервис AWS – с куда более простыми инструментами настройки синтезированной речи по ударениям, паузам и т. д. К примеру, на многих таких сервисах можно четко поставить ударение в слове, вставив знак «+» перед слогом, на который надо сделать ударение: теле+видение. А чтобы добавить паузу, используется дефис. Чтобы удлинить паузу, надо добавить еще один дефис. Очевидно, что это гораздо проще, чем программирование в AWS.

Теперь, собственно, к ресурсам преобразования текста в речь. Довольно долго я пользовался ресурсом Apihost.ru, где есть как платный, так и бесплатный варианты. А бесплатный вариант содержит три режима, различающиеся функциями и возможностями.

В режиме V1 TTS объем текста ограничен 1 тыс. символов. Причем учитываются как значащие символы, так и пробелы. Преобразование текста в речь выполняется довольно неплохо, но, если можно так сказать, без учета контекста. Это выражается в неправильной расстановке ударений даже, казалось бы, в самых простых и очевидных конструкциях. К примеру, при озвучивании словосочетания «соотношение качества и цены» робот в слове цены почему-то ставит ударение на «е». Исправить это можно знаком «+», вставленном в слово в соответствующем месте, но, во-первых, доводка текста до правильного произнесения занимает довольно много времени, а во-вторых, добавление каждого корректирующего символа – это уменьшение полезного объема текста. А потеря в результате таких операций одного слова в предложении означает потерю всего предложения.

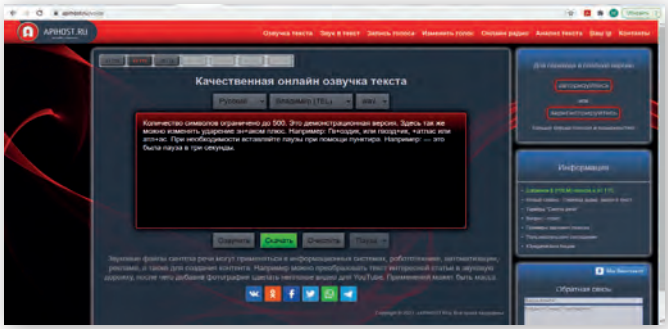


Режим V1 TTS

Еще часть текста могут «украсть» аббревиатуры, как на русском, так и на английском. Робот в большинстве случаев произнесет их просто как набор букв: AWS – авс, вместо «эй дабл ю эс». Чтобы это исправить, придется заменить аббревиатуру транскрипцией.

Зато в режиме V1 TTS есть широкий выбор мужских и женских голосов, для которых можно задать эмоциональную окраску (нейтральная, дружественная, раздраженная) и скорректировать высоту тона, повышая или понижая ее в соответствии с выпадающим списком. Эффект получается разный, порой довольно забавный. Есть также четыре языка на выбор: русский, украинский, турецкий и английский. Присутствие турецкого, честно говоря, немного удивило. Интересно, что сервис даже смог озвучить русский текст на украинском языке. Правда, сделать это на английском или турецком отказался.

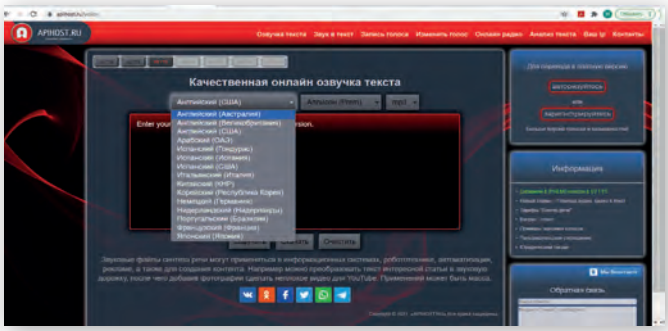
Для формирования трека есть два файловых формата – mp3 и wav. И в том и в другом случае трек формируется монофонический. А чтобы сделать из него стерео, надо в аудиоредакторе дублировать его в оба канала. Я использую редактор Audacity, который с этим отлично справляется.



Режим V2 TTS

Режим V2 TTS до недавнего времени был применимым с более качественными голосовыми характеристиками, поскольку варианты голосов подразделяются здесь на стандартные и премиум. Премиум изначально были недоступны в бесплатной версии, а вот стандартные можно было использовать. Правда, объем текста здесь ограничен значением 500 символов. Кроме того, спектр голосов ограничен всего четырьмя – один мужской и три женских (в версиях «стандарт» и «премиум»), отсутствуют выбор эмоциональной окраски и коррекция высоты тона, вывод трека только в формате wav, а языков всего два – русский и казахский.

Но теперь этот режим стал вообще не актуален, поскольку даже если выбрана стандартная версия, робот сообщает, что премиум-версия голоса недоступна. Вот такая странность.



Режим V5 TTS

Ну а режим V5 TTS для озвучивания текста на русском языке совсем не годится, поскольку этот язык тут не поддерживается. Зато он может быть полезен, если нужно сформировать трек на одном из иностранных языков, даже на корейском. Выбор настроек тут более чем скуден: язык и голос. Формат трека – mp3.

В интерфейсе видно, что есть еще четыре режима, но в демоверсии они неактивны и относятся к полной, то есть платной версии, которую автор не тестировал.

На мой взгляд, даже в таком усеченном виде этот ресурс работает пока лучше, чем AWS Polly, и пользоваться им легче. Но времени на то, чтобы довести текст не то что до желаемого, а до приемлемого звучания, придется потратить немало. Особенно если текст большой.

Продолжение следует