

Speechmatics – речь в текст с переводом в режиме реального времени

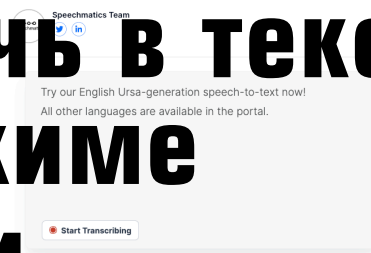
Дамилола Омотозо

В телевидении за кадром остаются многие операции и процедуры, необходимые тем не менее для создания высококачественного контента. К таковым относятся, в частности, преобразование речи в текст, что необходимо для различных целей – описания контента, перевода, субтитрования и др. Параллельно зачастую требуется и перевод исходного текста на разные языки. В настоящее время инструментов для этого достаточно много, но далеко не все они в полной мере продуктивны.

Недавно компания Speechmatics – разработчик эффективной технологии распознавания речи, раскрыла планы по расширению своих возможностей преобразования речи в текст за счет добавления функции перевода в режиме реального времени и интеграции всех возможностей в единый API. Преодоление языковых барьеров позволяет большему числу людей потреблять контент вне зависимости от того, на каком языке он был изначально произведен. Кроме того, появляется возможность автоматического перевода контента в режиме реального времени, транслируемого из разных регионов планеты. Такая комбинация функций позволяет потребителям использовать один из наиболее точных инструментов преобразования речи в текст и выполнять перевод речи для 69 языковых пар.

Функция перевода речи в режиме реального времени появилась спустя месяц после запуска компанией Speechmatics ядра Urса, обеспечивающего максимально точный перевод речи в текст. Точность здесь на 25% выше, чем у Whisper от OpenAI, и на 38% – чем у Google. Функционал Urса удвоился благодаря тому, что Speechmatics добавила возможность перевода текста в режиме реального времени не только для английского языка, но и для других языков ему в пару. Например, для немецкого, испанского, вьетнамского и др. Единый API дает возможность перевода на несколько языков одним запросом. Например, на базе одного звукового потока можно получить преобразование англоязычной речи в текст в режиме реального времени и параллельно перевод на японский, французский, хинди, китайский и корейский языки.

Транскрипция, а теперь и перевод, обеспечиваемые в режиме реального времени решением Speechmatics, предоставляют такой же уровень точности, как и при аналогичных процедурах для предварительно записанного аудио. Есть также возможность управления балансом между точностью и скоростью (задержкой) выполнения операции, чтобы пользователь мог настроить параметры сервиса в соответствии со своими потребностями. Единый API позволяет повысить эффективность и оперативность рабочих процессов за счет того, что в нем сочетаются и транскрипция, и перевод.



Urса provides the world's most accurate speech-to-text and is a breakthrough in accessibility, reducing the digital divide for voices that other speech-to-text systems struggle to recognize. We've been blown away by how Urса can accurately transcribe challenging speech - in particular, speech that would have been impossible for a human to grasp without first reading the transcript. We found Urса can transcribe singing very well too, even though we didn't specifically train for this use case. You may like to try and understand the following clips without showing the transcripts first.

Компании из самых разных отраслей получают возможность охватить широкую аудиторию за счет перевода речи в режиме реального времени, что раньше было довольно сложной и дорогостоящей процедурой, поскольку все делалось вручную силами специального персонала. В частности, для вещательной индустрии США, объем которой в 2022 году оценивался в сумму более 300 млрд долларов, быстрое и точное выполнение переводов, сконцентрированные в едином API, открывает возможность титрования транслируемого вживую потокового контента для зрителей по всему миру. Аналогично, контактные центры, где охват абонентов имеет значение, могут расширить сферу своей деятельности за счет поддержки нескольких языков, используя экономически эффективную технологию автоматизации и предоставляя клиентам более высокое качество обслуживания за счет общения с ними на их родных языках.

Вот что сказал руководитель отдела продаж и разработки Speechmatics Дамир Дерд: «*Это знаковое событие в сфере технологии распознавания речи, и мы гордимся тем, что остаемся на острие прогресса и демонстрируем наше стремление к тому, чтобы каждый голос был услышан и понят. Это новое предложение открывает для наших клиентов по-настоящему глобальный рынок, поскольку позволяет практически мгновенно переводить каждое сказанное слово. Поскольку спрос со стороны зрителей в разных регионах планеты на телепрограммы, спортивные и развлекательные трансляции, подкасты и игровые стриминги, видео на YouTube и в других соцсетях растет, потребность в титровании видео на разных языках тоже увеличивается. Мы рады, что запускаем свой новый сервис, и продолжим работать, расширяя спектр поддерживаемых языков и придавая платформе способность использовать их для перевода, в том числе и для случаев, когда английский не является базовым.*

«*Speechmatics обеспечивает наиболее высокую на рынке точность преобразования речи в текст для предварительно записанных файлов и живых потоков, – отметил президент ENCO Кен Фроммерт. – Добавление перевода в режиме реального времени в их единый API кардинально меняет все в плане титрования прямых ТВ-трансляций, давая возможность не только транскрибировать, но и применять Speechmatics для точного перевода текста в режиме реального времени, чтобы формировать титры при вещании на глобальную аудиторию.*

Впервые перевод в режиме реального времени был продемонстрирован компанией на выставке NAB 2023. ▶