

О моделях

1. Vosk — это легкая и быстрая модель, подходит для приложений с ограниченными ресурсами.
2. Jasper — более сложная модель, баланс между скоростью и качеством.
3. Whisper — это мощная модель, но требует много ресурсов.

Сравнительный обзор метрик

Модель	Время распознавания (VR), сек	Длина файла, сек	RTF	WER (%)	Особенности распознавания
Vosk	1.93	5.43	0.354	140.00	Быстрый, низкая скорость, очень высокий WER
Jasper	3.00	5.43	0.552	100.00	Средняя скорость, высокий WER
Whisper	185.20	5.43	34.084	50.00	Очень медленный, лучше качество, средний WER

Анализ и выводы

Время распознавания и RTF

1. Vosk — самый быстрый, время около 2 секунд, $RTF \approx 0.354$. Это говорит о высокой скорости обработки, что подходит для приложений в реальном времени или с ограниченными ресурсами.
2. Jasper — чуть медленнее, около 3 секунд, $RTF \approx 0.552$. В целом, работает быстрее, чем Whisper, но медленнее Vosk.
3. Whisper — значительно медленнее, более 185 секунд, $RTF \approx 34.084$. Этот показатель указывает на высокую вычислительную сложность и низкую эффективность для коротких аудиофайлов.

Ошибки WER

Важно: в $WER > 100\%$ — это означает, что количество ошибок превышает количество слов в эталонном тексте, что говорит о плохой точности.

1. Vosk — WER 140% — очень высокий, что указывает на низкое качество распознавания в данном конкретном случае.
2. Jasper — WER 100% — тоже плохой результат, полностью не справляется с задачей.
3. Whisper — WER 50% — результат лучше, но все равно оставляет много ошибок.

Выводы

1. Выводы по скорости: модель Vosk значительно быстрее Jasper и намного быстрее Whisper. Это делает его подходящим для задач, использующихся на устройствах персонала заводов ТМК.
2. Качество распознавания: В данном случае все модели показали плохой или средний результат по WER, что, скорее всего, связано с недостаточной подготовкой модели. WER должен быть значительно ниже 100%, чтобы говорить о приемлемом качестве.
3. Важно отметить, что высокую вычислительную сложность модели Whisper — позволила улучшить в качестве в данном тексте.