

<b>Parameter</b>	<b>Single-Speaker</b>	<b>VCTK</b>	<b>LibriSpeech</b>
FFT Size	4096	4096	4096
FFT Window Size / Shift	2400 / 600	2400 / 600	1600 / 400
Audio Sample Rate	48000	48000	16000
Reduction Factor $r$	4	4	4
Mel Bands	80	80	80
Sharpening Factor	1.4	1.4	1.4
Character Embedding Dim.	256	256	256
Encoder Layers / Conv. Width / Channels	7 / 5 / 64	7 / 5 / 128	7 / 5 / 256
Decoder Affine Size	128, 256	128, 256	128, 256
Decoder Layers / Conv. Width	4 / 5	6 / 5	8 / 5
Attention Hidden Size	128	256	256
Position Weight / Initial Rate	1.0 / 6.3	0.1 / 7.6	0.1 / 2.6
Converter Layers / Conv. Width / Channels	5 / 5 / 256	6 / 5 / 256	8 / 5 / 256
Dropout Keep Probability	0.95	0.95	0.99
Number of Speakers	1	108	2484
Speaker Embedding Dim.	-	16	512
ADAM Learning Rate	0.001	0.0005	0.0005
Anneal Rate / Anneal Interval	-	0.98 / 30000	0.95 / 30000
Batch Size	16	16	16
Max Gradient Norm	100	100	50.0
Gradient Clipping Max. Value	5	5	5