Appendix

Source Code

All of our source code is public available on GitHub[1].

Experiments

Our computers had these specifications:

- 1: Ubuntu 20.04 LTS, AMD Ryzen 7 3800X, Nvidia Titan RTX, CUDA: 11.2
- 2: Ubuntu 18.04 LTS, Intel i5-7500T, Nvidia GeForce GTX 1070 Mobile, CUDA: 10.2
- 3: Pop!_OS 20.04 LTS, Intel i7-7700, Nvidia GeForce RTX 2070 SUPER, CUDA: 10.2

Table 1: Computers on which we ran our experiments including the training durations.

Dataset	Computer	≈ Training duration (h)	(days)
Full	1	117.05	4.88
$Rand_1$	2	175.94	7.33
$Rand_2$	3	89.03	3.71
$Rand_3$	3	89.62	3.73
Greedy	3	91.38	3.81
KLD	2	167.58	6.98

Table 2: Overview of our used seeds.

Type	Seed(s)
Creation of *Rand* sets	3333
Training	1234
Inference 1 to 10	1111 to 1120

[1] `https://github.com/stefantaubert/tacotron`

Table 3: Model parameters.

Parameter	Value
symbols_embedding_dim	512
encoder_kernel_size	5
encoder_n_convolutions	3
n_frames_per_step	1
decoder_rnn_dim	1024
prenet_dim	256
gate_threshold	0.5
p_attention_dropout	0.1
p_decoder_dropout	0.1
attention_rnn_dim	1024
attention_dim	128
attention_location_n_filters	32
attention_location_kernel_size	31
postnet_embedding_dim	512
postnet_kernel_size	5
postnet_n_convolutions	5
max_decoder_steps	3000

Table 4: Mel spectrogram parameters.

Parameter	Value
filter_length	1024
hop_length	256
win_length	1024
window	hann
sampling_rate	22050
n_mel_channels	80
mel_fmin	0.0
mel_fmax	8000.0

Table 5: Training/optimizer parameters.

Parameter	Value
learning_rate	0.0005
weight_decay	10^{-6}
beta1	0.9
beta2	0.999
eps	10^{-8}
grad_clip_thresh	1.0
batch_size	24
mask_padding	True
amsgrad	True
iterations	150000