

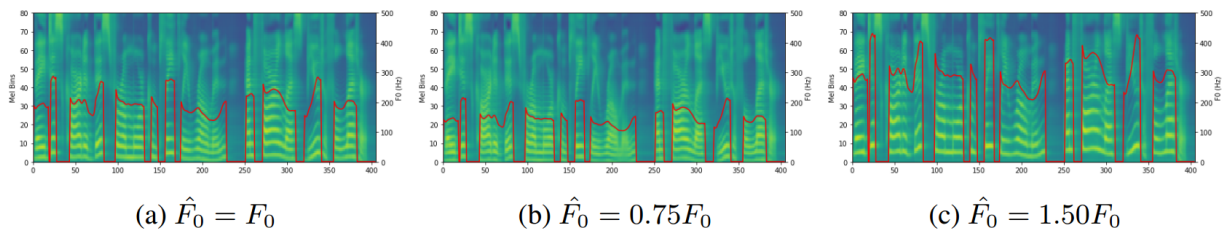
Details in Variance Control

Note: Because FastSpeech has already demonstrated the ability to control speed and breaks, and FastSpeech 2 and 2s inherits its advantages of robust and controllable speech synthesis, we only demonstrate examples for controlling pitch and energy (volume) in synthesized speech.

Pitch Control

FastSpeech 2 and 2s can control the pitch of each frame arbitrarily.

The following picture demonstrates: the pitch value of each frame is multiplied by a factor, i.e., 0.75 and 1.50.

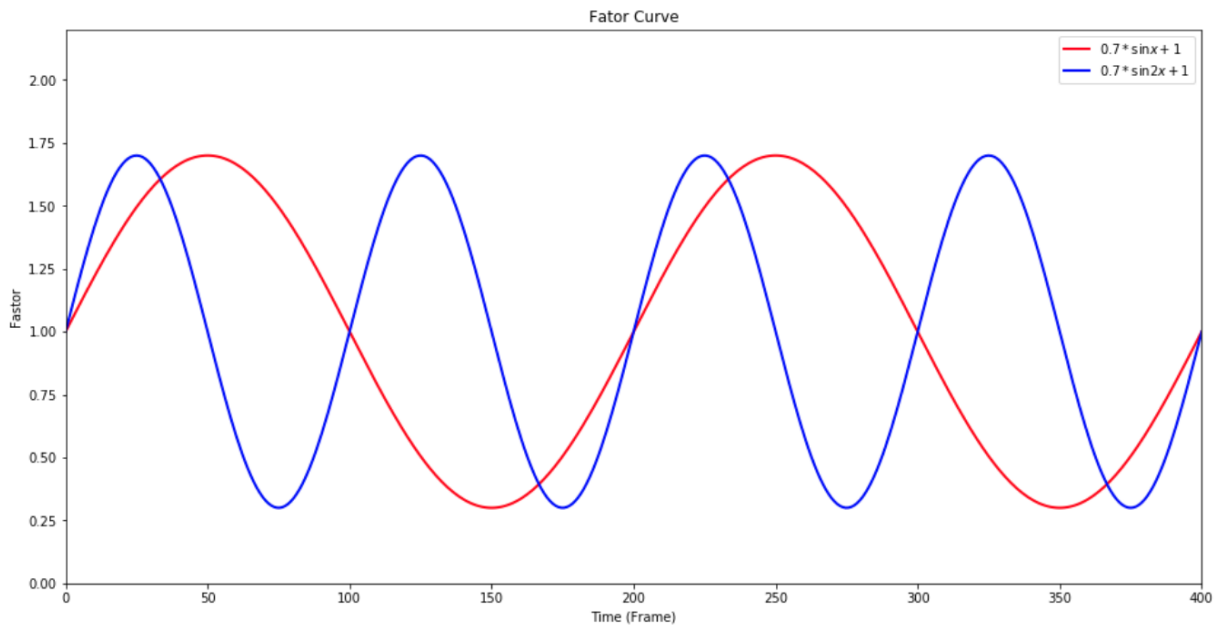


Energy (Volume) Control

FastSpeech 2 and 2s can control the energy of each frame arbitrarily.

We demonstrate using a factor curve to control the energy of each frame.

Use sinusoidal curves of different periods to multiply with the predicted energy, which results in periodic volume change in synthesized speech.



Directory Structure

```
├─ fs2_energy    control energy in FastSpeech 2
|   └─ audio1    amp0.7_period100/amp0.7_period200 indicates factor curve is the blue/red one
|                  respectively
|   |   └─ pred_LJ001-0001_amp0.7_period100.png
|   |   └─ pred_LJ001-0001_amp0.7_period100.wav
```

```
| | └─ pred_LJ001-0001_amp0.7_period200.png
| | └─ pred_LJ001-0001_amp0.7_period200.wav
| | └─ pred_LJ001-0001.png
| | └─ pred_LJ001-0001.wav
| | └─ transcripts.txt
| └─ audio2
|   | ...
└─ fs2_pitch    control pitch in FastSpeech 2
  | └─ audio1    pitch0.75 means the factor is 0.75
  | | └─ pred_LJ001-0035_pitch0.75.png
  | | └─ pred_LJ001-0035_pitch0.75.wav
  | | └─ pred_LJ001-0035_pitch1.0.png
  | | └─ pred_LJ001-0035_pitch1.0.wav
  | | └─ pred_LJ001-0035_pitch1.5.png
  | | └─ pred_LJ001-0035_pitch1.5.wav
  | | └─ transcripts.txt
  | └─ audio2
  |   | ...
└─ fs2s_energy  control energy in FastSpeech 2s
  | ...
└─ fs2s_pitch   control pitch in FastSpeech 2s
  | ...
└─ README.md
```