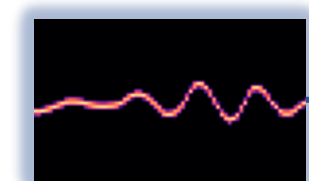
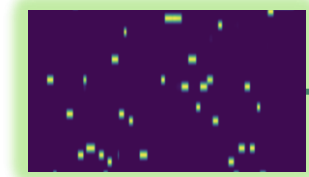


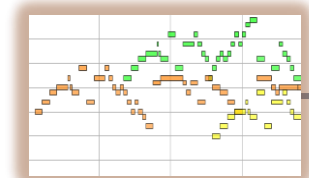
Time-varying Conditions



f_0



PPG



Piano Roll

CAT

Encoder

FiLM x N

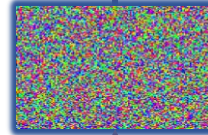
Fuse Conditions

C.A.

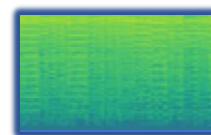
Decoder

FiLM x N

Estimate noise

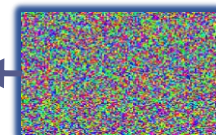


L_1



Mel

+



Noise

Performer Condition
Noise Level



Diffusion
Timestep

t

S.E.

Scale,
Shift

FiLM

Binary Descriptor

Audio Embedding

MLP