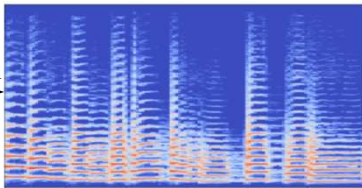


Input

Short-time
Fourier transform



Wave encoder

$f_{wav}(\cdot)$

0	1	1
2	1	3
3	5	7

1D CNN

**Time-frequency
encoder**

$f_{spec}(\cdot)$

1	2	2	6
3	2	4	4
7	2	3	5
4	3	8	4

2D CNN