

Table S1 part 1

Complexify	Sin-cos encoding	LPF cutoff Freq.	Gate	Combine freqs	Loss
0	1	100	1	1	0.1389
0	1	100	0	0	0.1414
1	0	100	1	0	0.1421
0	0	210	0	0	0.1423
1	1	100	1	0	0.1426
0	1	210	1	0	0.1435
0	1	100	1	0	0.1447
0	0	100	0	0	0.1448
0	0	210	0	1	0.1449
0	1	100	0	1	0.1451
0	0	210	1	1	0.1457
0	0	210	1	0	0.1459
0	0	210	1	1	0.1462
0	1	210	0	0	0.1465
0	0	100	1	1	0.1473
1	1	210	1	0	0.1476
0	0	210	0	1	0.1485
1	0	210	1	0	0.1485
:	:	:	:	:	:
:	:	:	:	:	:

Table S1 (continued)

Complexity	Sin-cos encoding	LPF cutoff Freq.	Gate	Combine freq	Loss
0	0	100	0	1	0.1499
0	0	210	0	0	0.1515
0	1	210	1	0	0.1518
0	1	210	0	1	0.1533
0	0	210	1	0	0.1557
0	1	20	1	1	0.1574
0	0	20	1	1	0.1614
0	1	20	1	0	0.1620
0	1	20	0	1	0.1631
1	0	210	1	0	0.1633
1	0	210	1	0	0.1633
0	0	20	0	0	0.1645
1	1	20	1	0	0.1655
0	0	20	0	1	0.1663
1	0	20	1	0	0.1667
1	0	20	1	0	0.1667
0	1	20	0	0	0.1676
0	0	20	1	0	0.1737
0	0	100	1	0	0.1491

Table S1: Ablation Study. In a series of tests, we examined the importance of each building block in the Adaptive-Spectrum-Layer (ASL). The results are sorted by the validation loss. Each column represents one building block with 1=used, 0=absent, or its specific value (LPF cutoff frequency). **Complexify** represents a module that learns new complex vectors from the given complex FFT. **sin-cos-encoding** represents using the sine and cosine values of the phase or just an angle. **Gate** represents a GLU-like mechanism in Fourier space (or a simple FC instead). **Combine freqs** represents using a separate projection layer on each FFT bin, or a single one

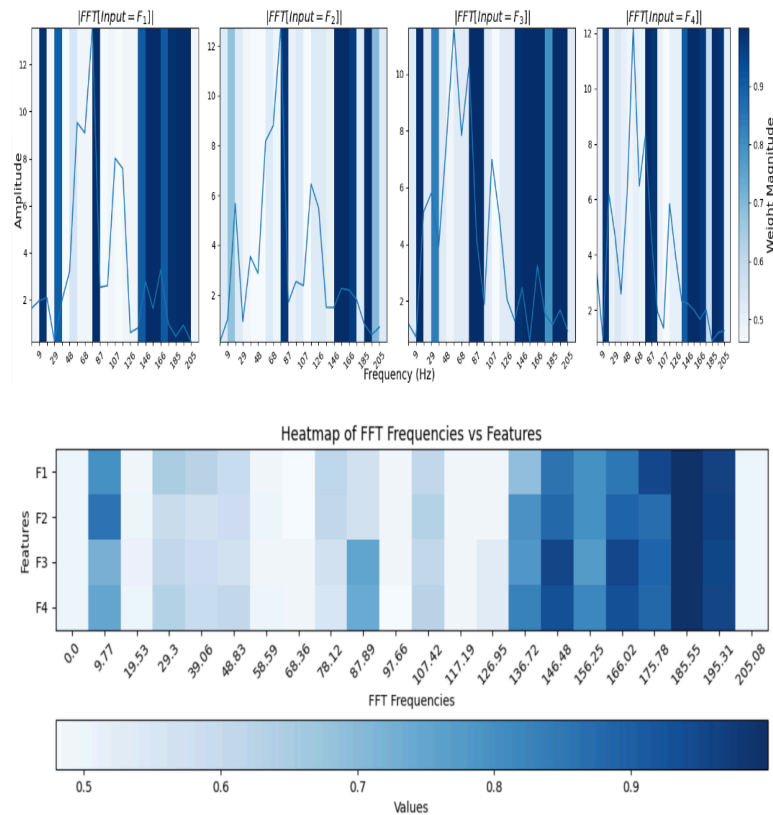
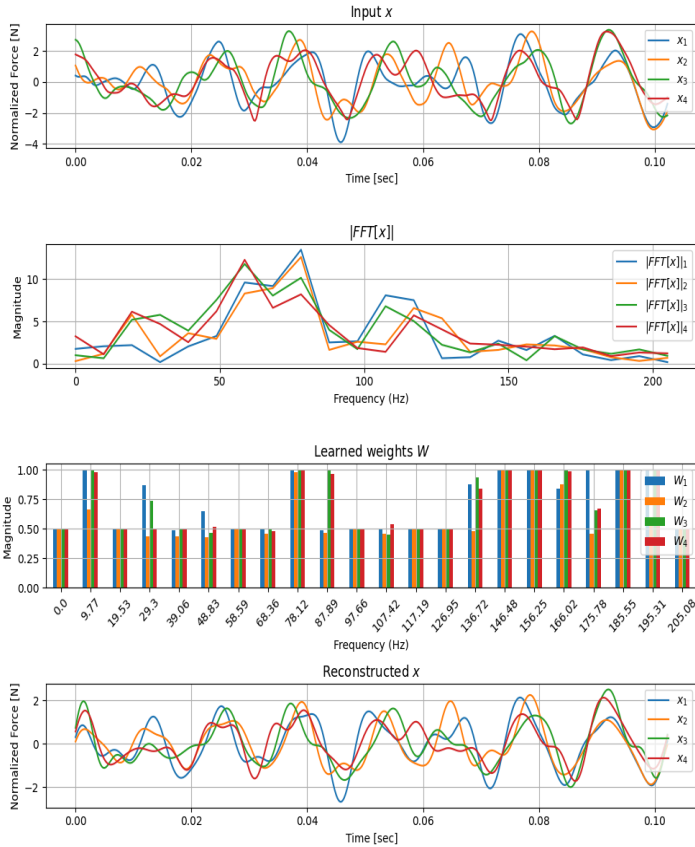


Figure S1: Fourier Bins and Attention in Fourier Space. (Left) An example of input propagation through a pre-trained Seq2Seq with ASL - Each bin in FFT space is multiplied by a bounded weight, and the reconstructed input is the inverse-FFT of the new representation. (Right) An example of per-feature attention weights in Fourier space of a pre-trained Seq2seq with ASL, alongside the averaged weight in a series of sliding windows. In this example, we identify three regions of interest - Low (~10Hz). Medium (~90Hz) and High (150-200Hz)