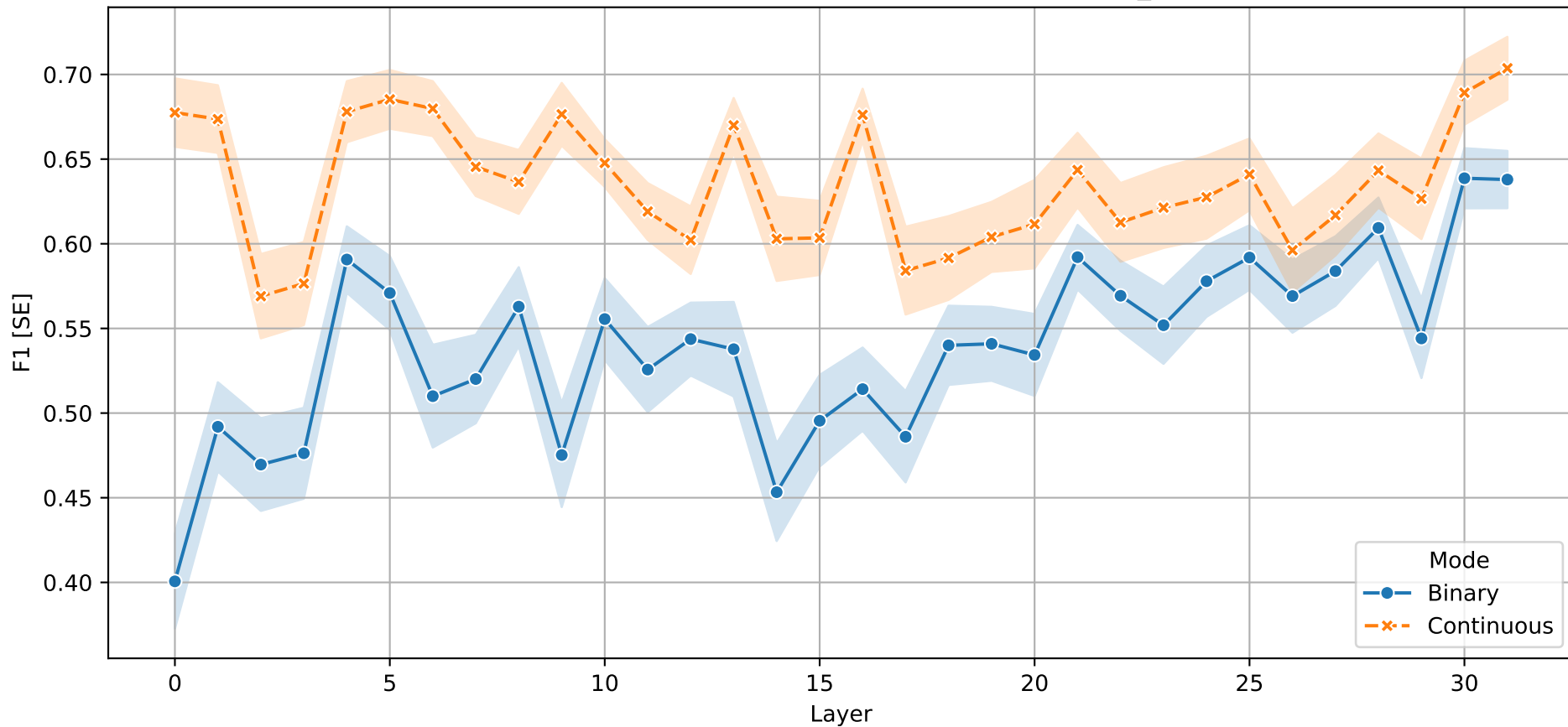
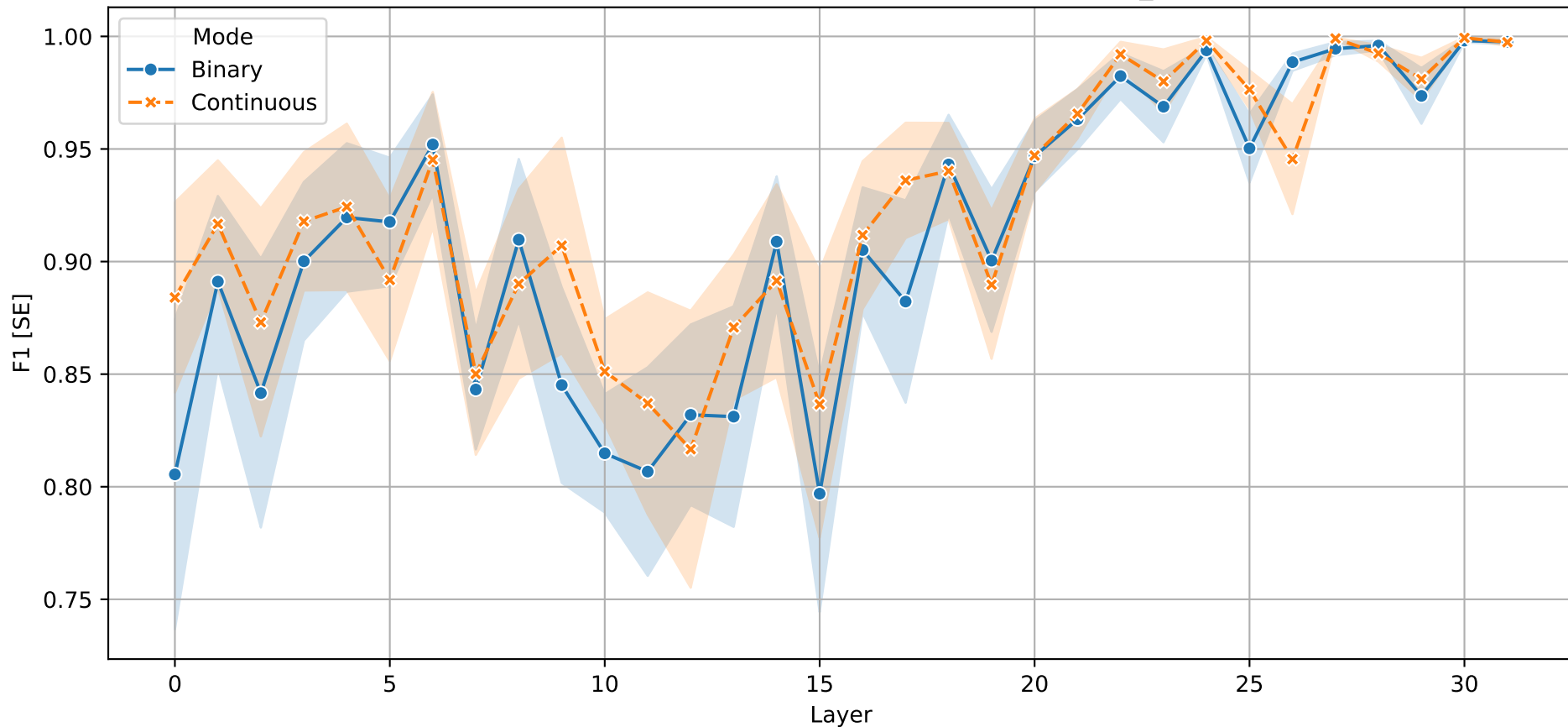


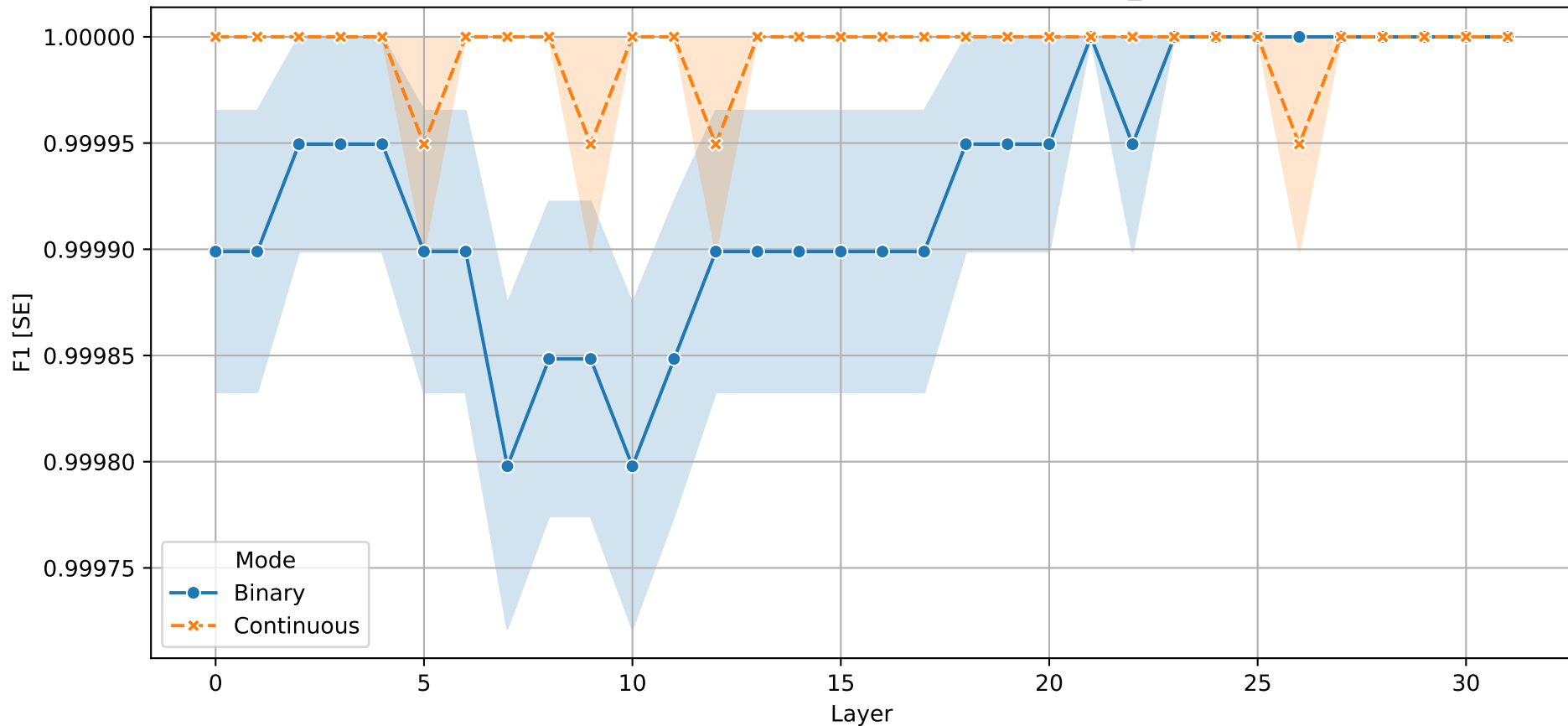
F1 per Layer - Single Neuron Probing for europarl_lang



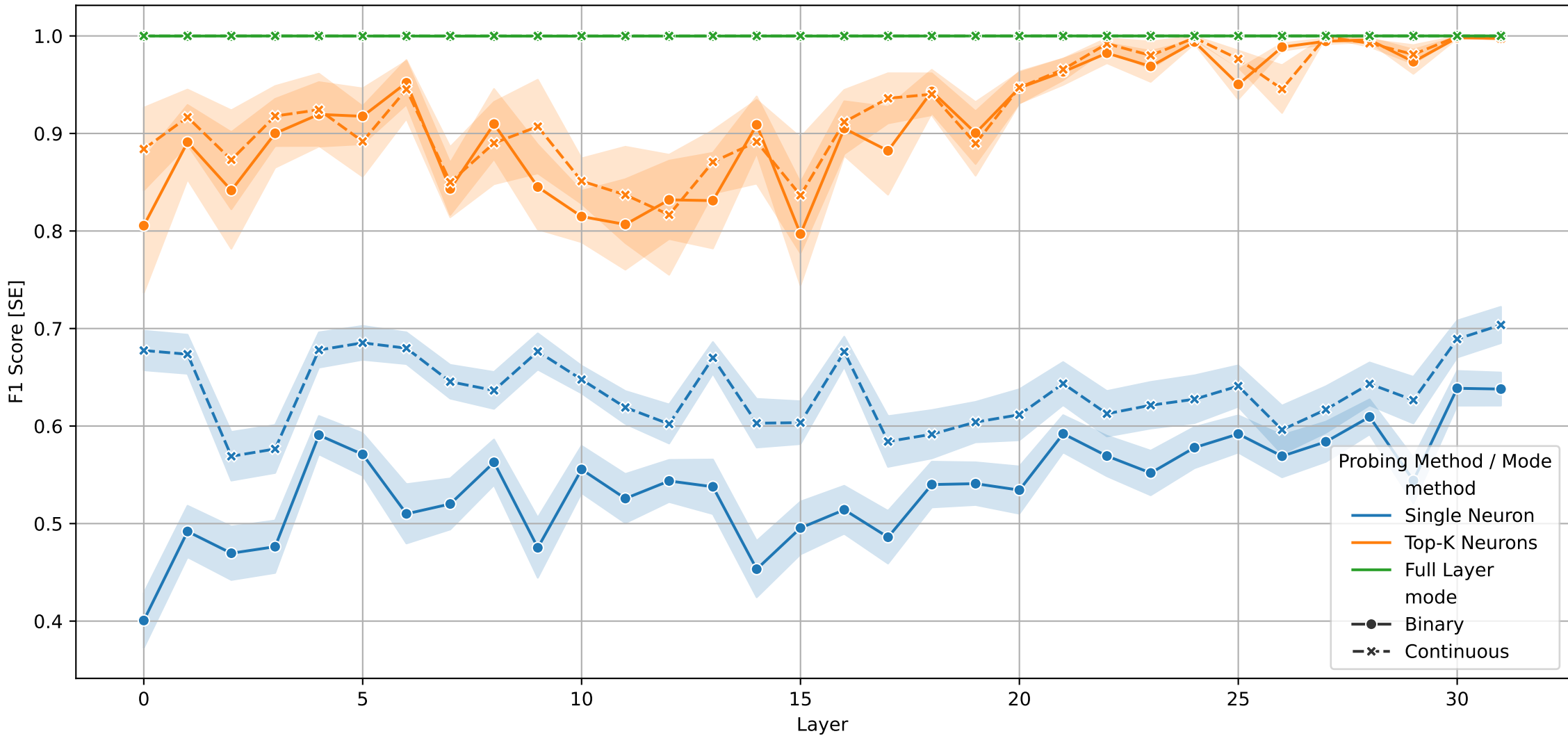
F1 per Layer - Top-K Neurons Probing for europarl_lang



F1 per Layer - Full Layer Probing for europarl_lang



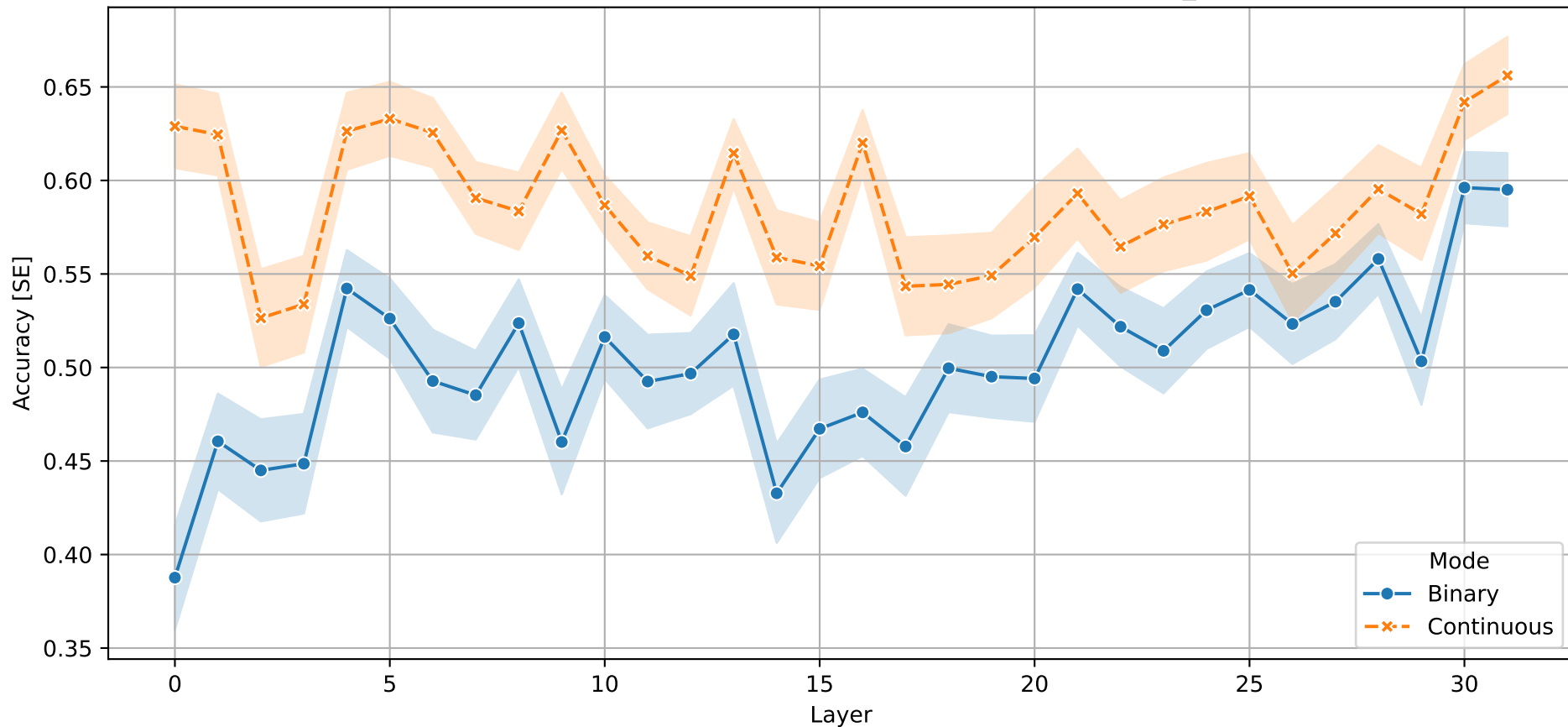
Overall F1 per Layer - All Methods for europarl_lang



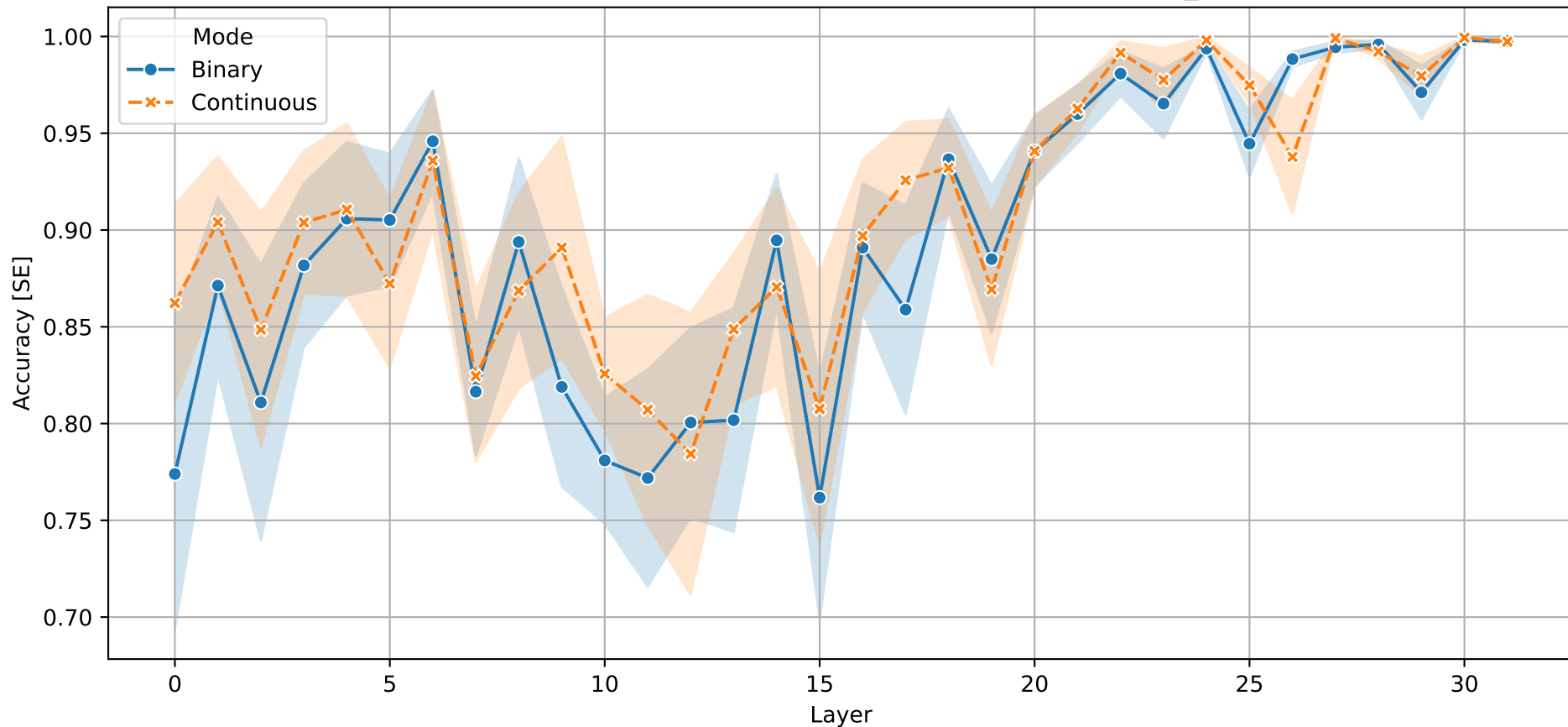
F1 Score Summary by Probing Method

Method	Metric	Binary	Continuous
Full Layer	f1_best_layer	21.0	0.0
Full Layer	f1_max	1.0	1.0
Full Layer	f1_mean	0.9999	1.0
Full Layer	f1_std	0.0002	0.0001
Single Neuron	f1_best_layer	30.0	31.0
Single Neuron	f1_max	1.0	1.0
Single Neuron	f1_mean	0.5394	0.6354
Single Neuron	f1_std	0.219	0.1915
Top-K Neurons	f1_best_layer	30.0	30.0
Top-K Neurons	f1_max	1.0	1.0
Top-K Neurons	f1_mean	0.9125	0.9236
Top-K Neurons	f1_std	0.1074	0.1007

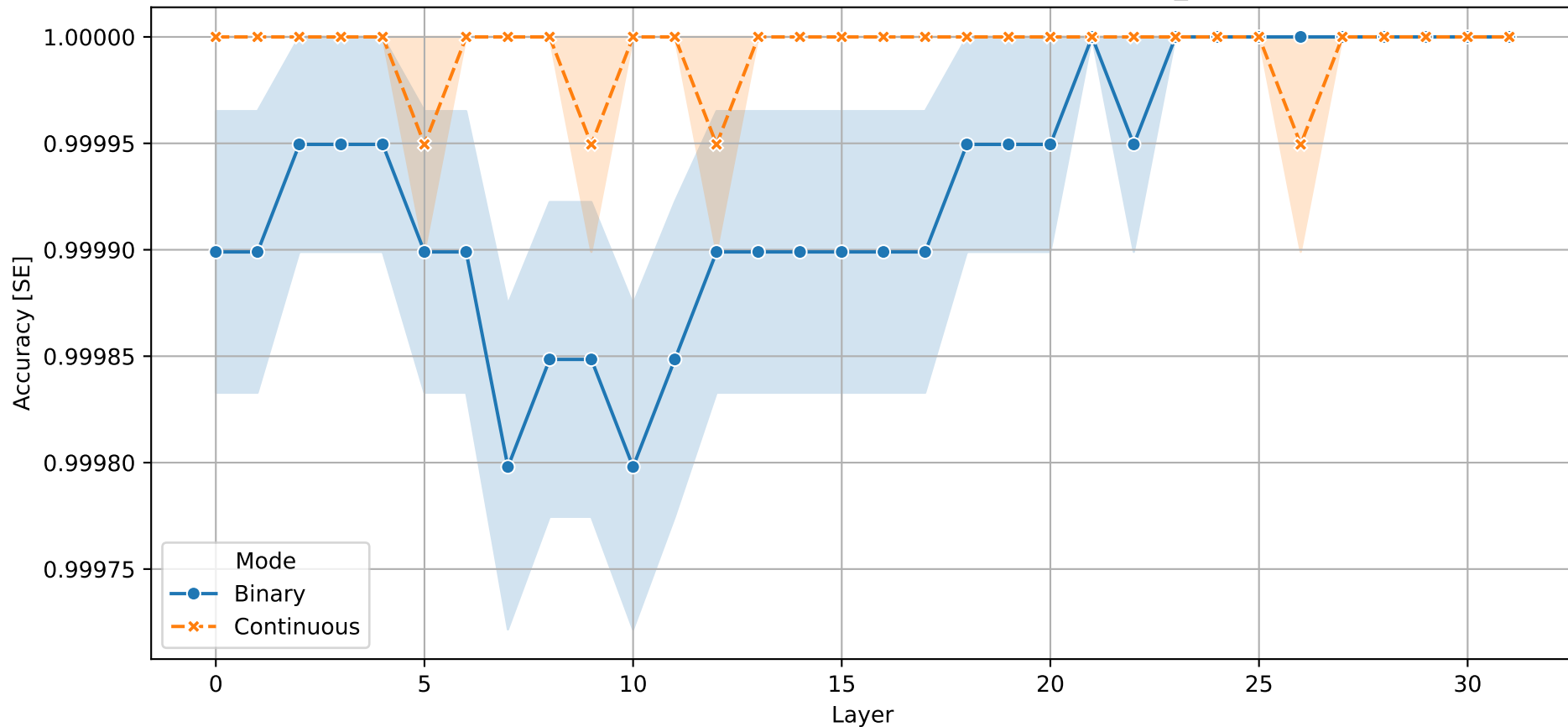
Accuracy per Layer - Single Neuron Probing for europarl_lang



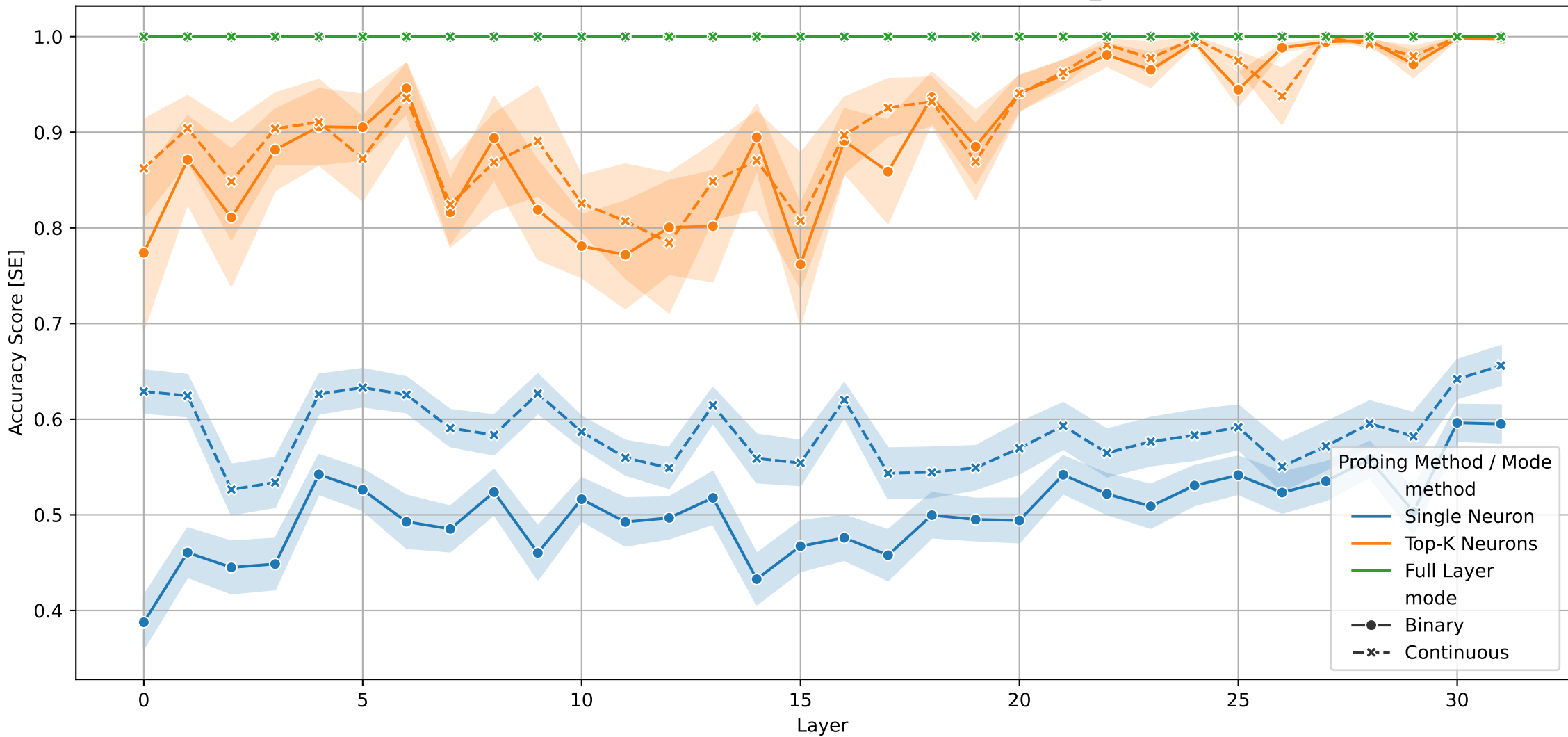
Accuracy per Layer - Top-K Neurons Probing for europarl_lang



Accuracy per Layer - Full Layer Probing for europarl_lang



Overall Accuracy per Layer - All Methods for europarl_lang



Accuracy Score Summary by Probing Method

Method	Metric	Binary	Continuous
Full Layer	accuracy_best_layer	21.0	0.0
Full Layer	accuracy_max	1.0	1.0
Full Layer	accuracy_mean	0.9999	1.0
Full Layer	accuracy_std	0.0002	0.0001
Single Neuron	accuracy_best_layer	30.0	31.0
Single Neuron	accuracy_max	1.0	1.0
Single Neuron	accuracy_mean	0.5023	0.5861
Single Neuron	accuracy_std	0.2128	0.2043
Top-K Neurons	accuracy_best_layer	30.0	30.0
Top-K Neurons	accuracy_max	1.0	1.0
Top-K Neurons	accuracy_mean	0.898	0.9106
Top-K Neurons	accuracy_std	0.1278	0.1203