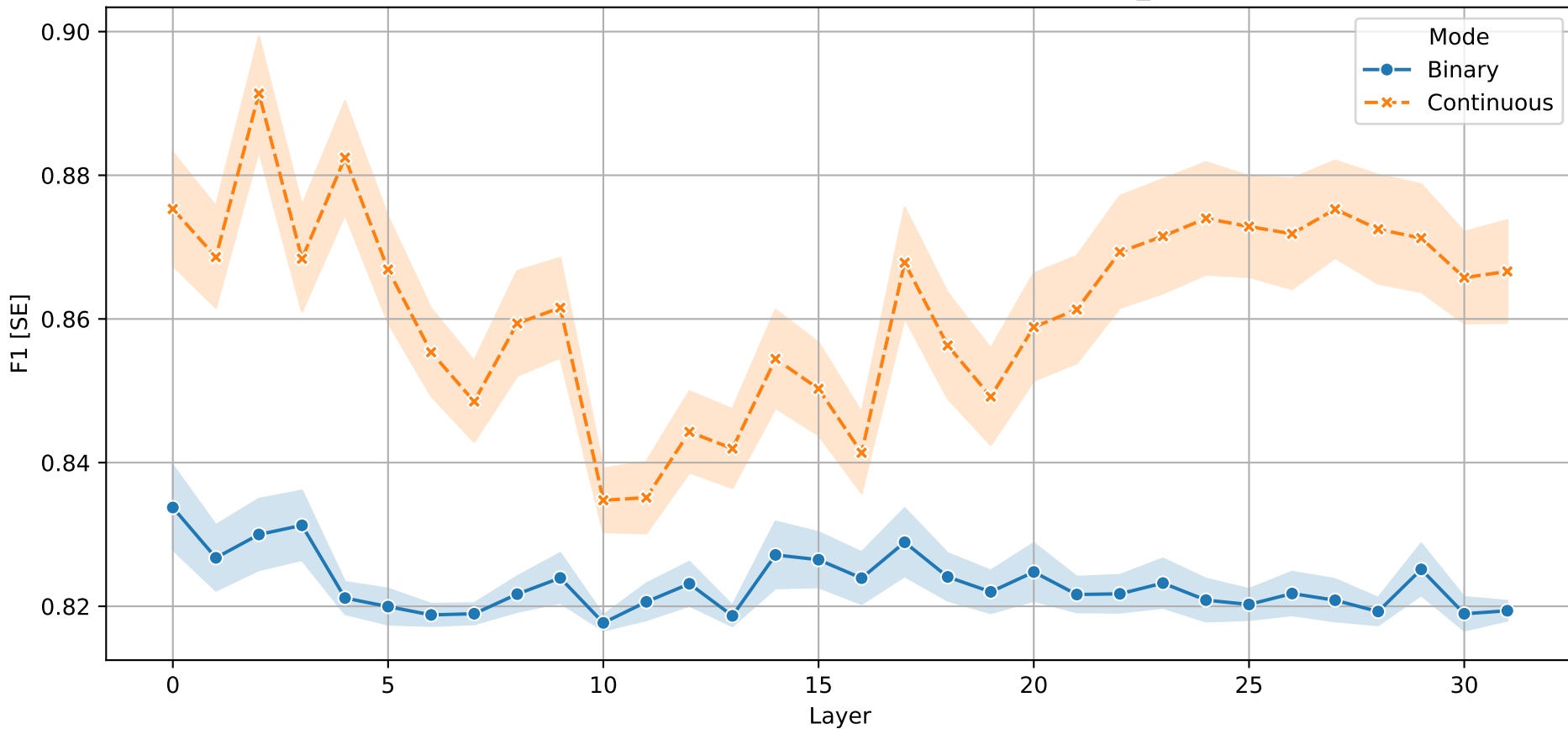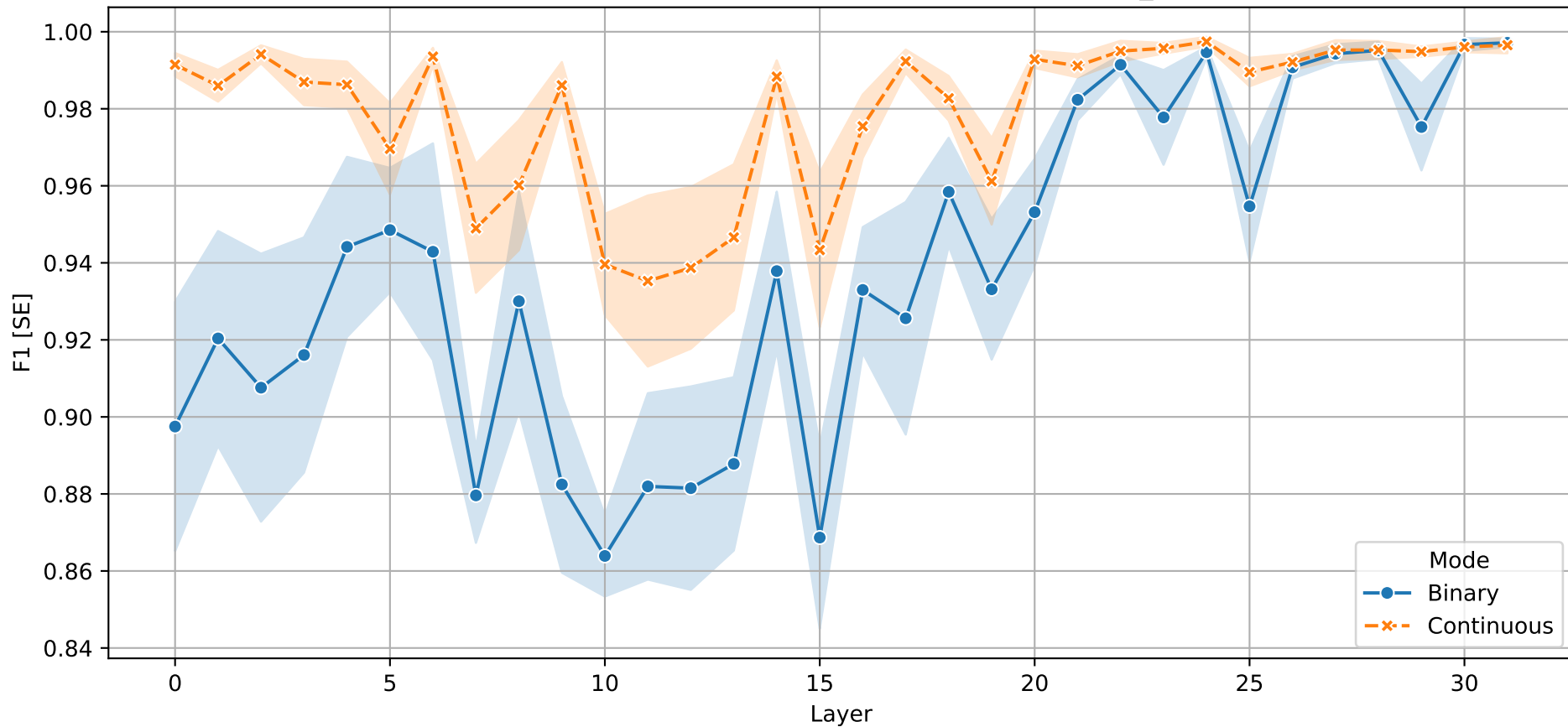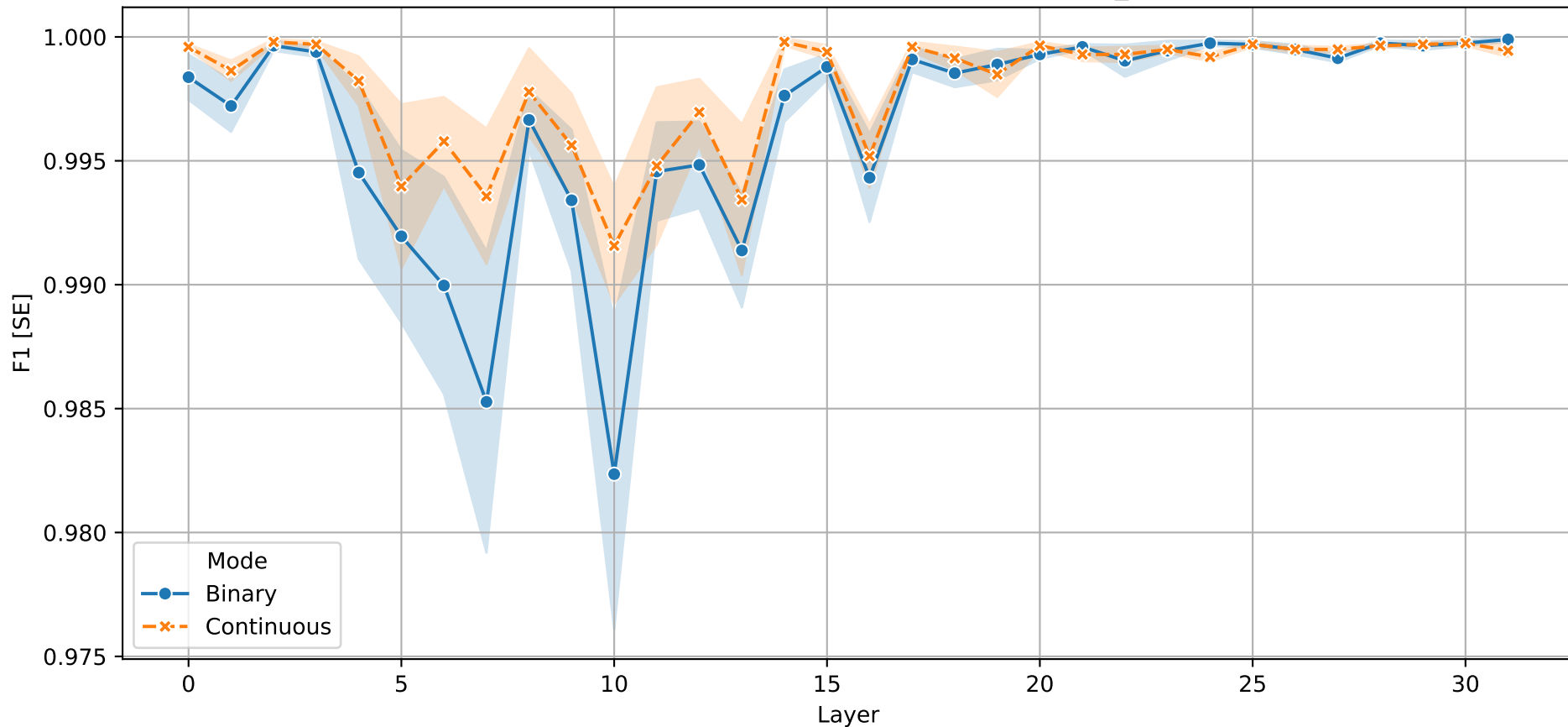F1 per Layer – Single Neuron Probing for europarl_lang
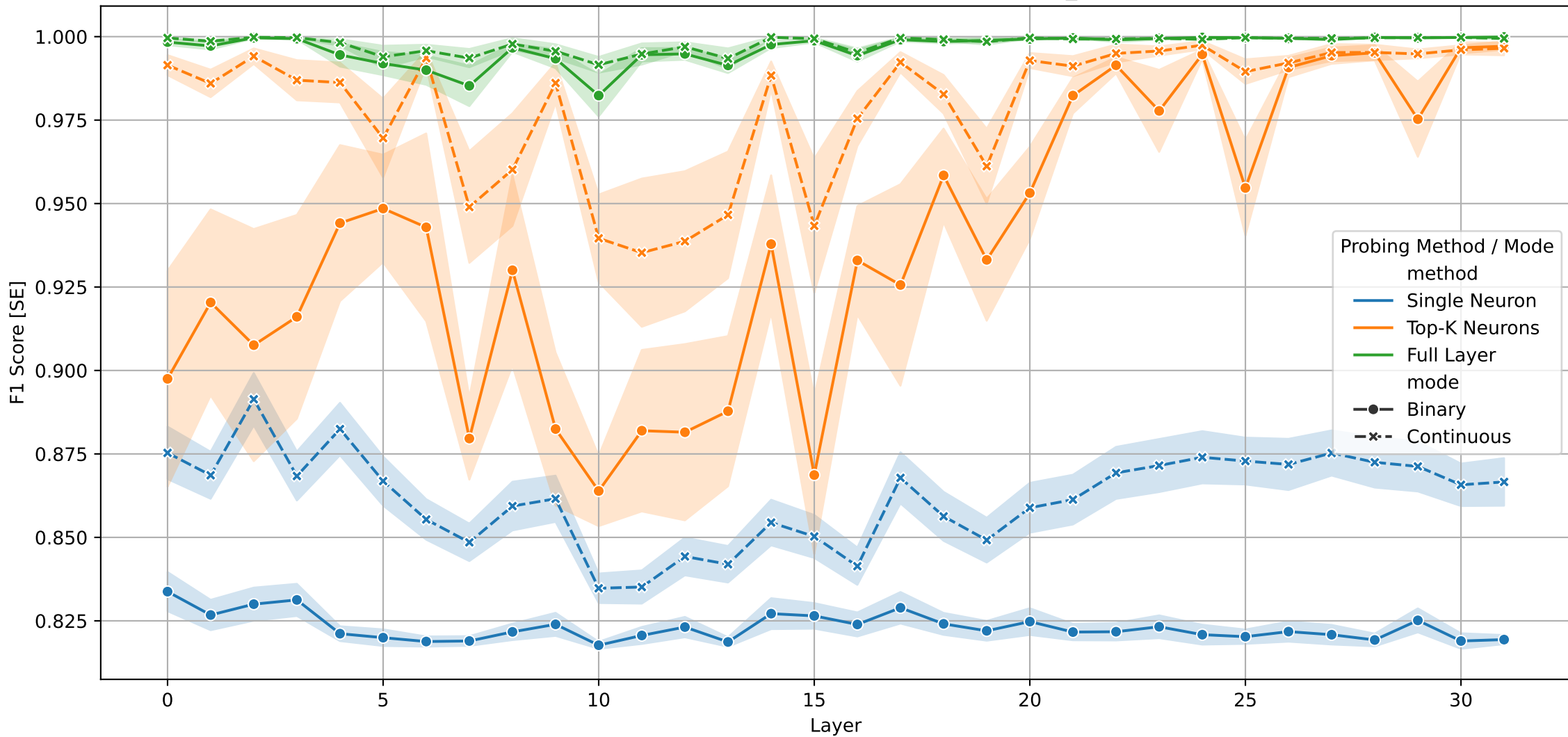
F1 per Layer – Top-K Neurons Probing for europarl_lang

F1 per Layer – Full Layer Probing for europarl_lang
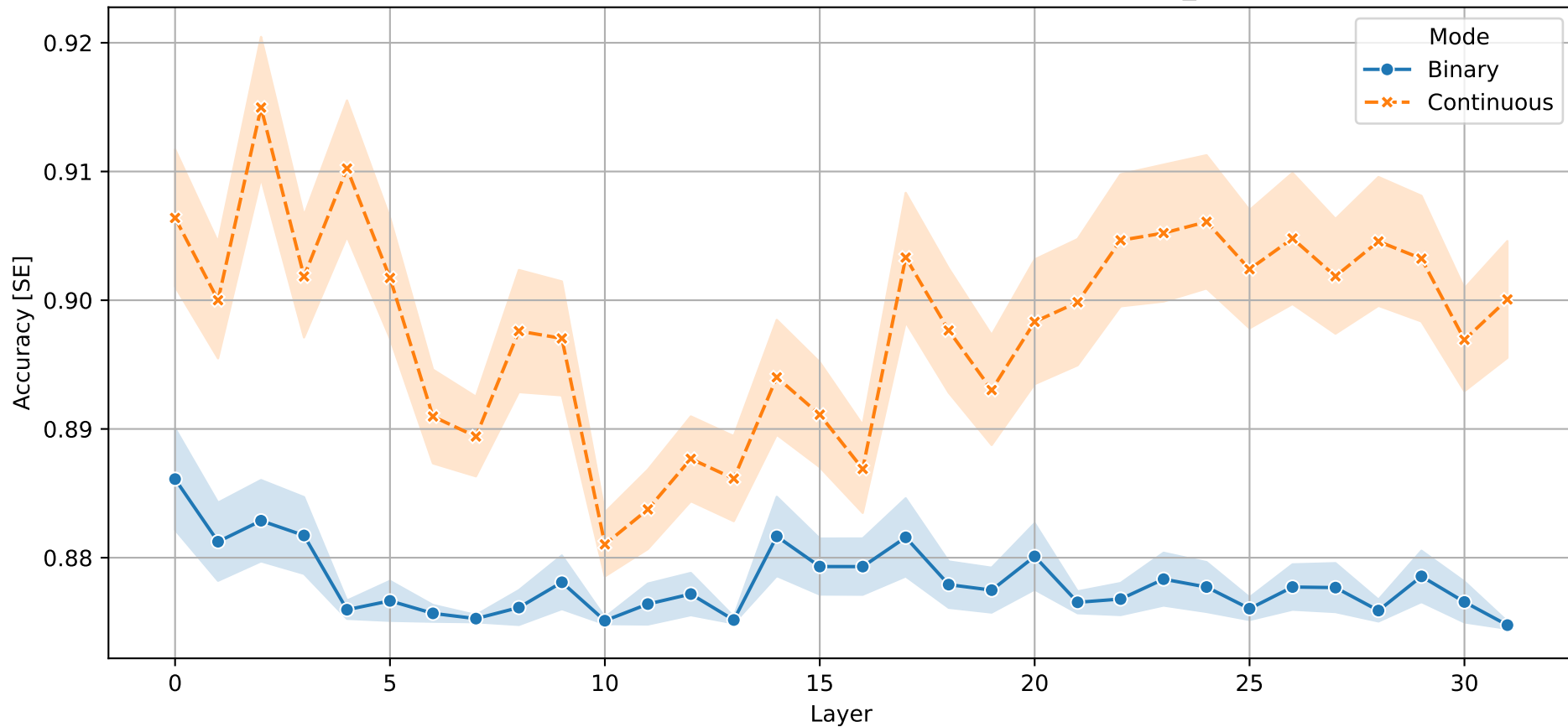
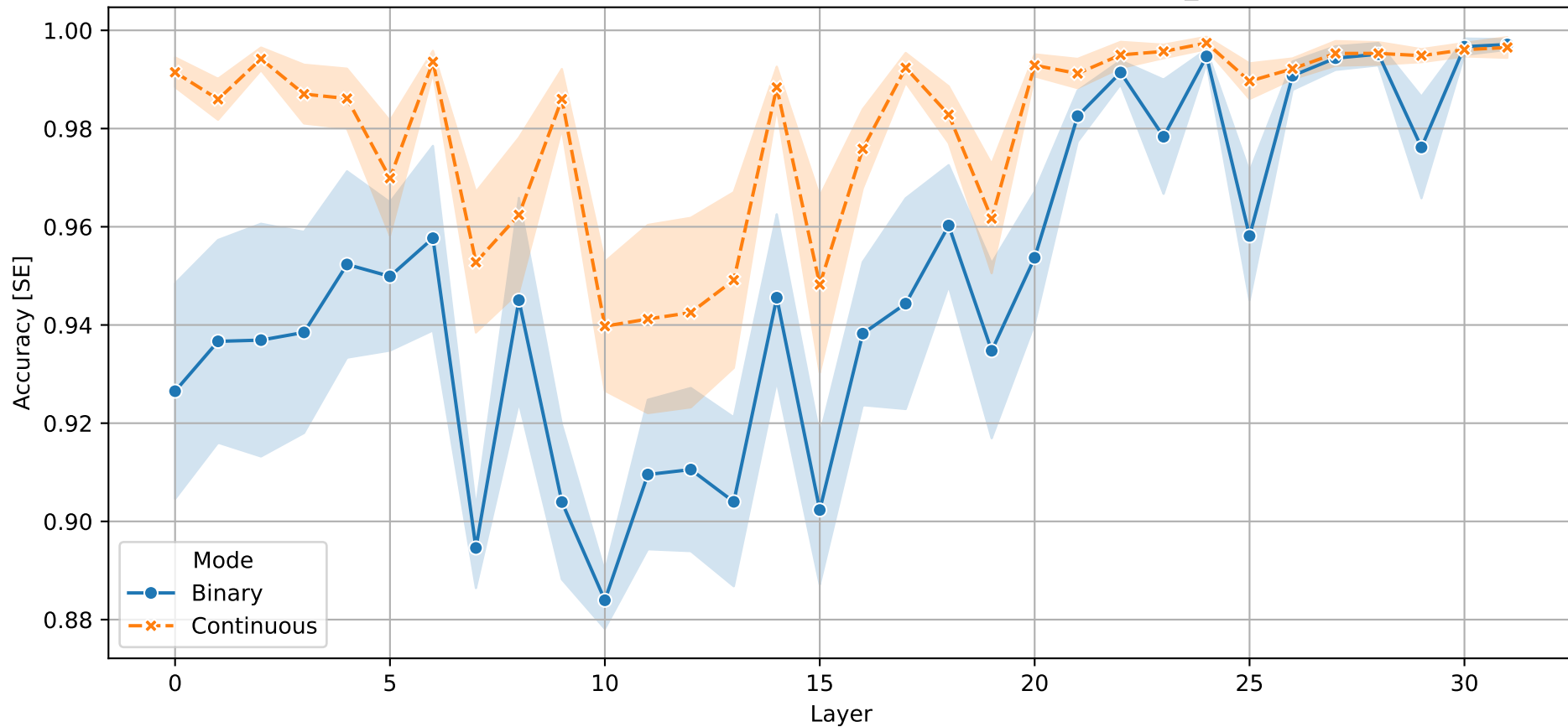Overall F1 per Layer – All Methods for europarl_lang

F1 Score Summary by Probing Method

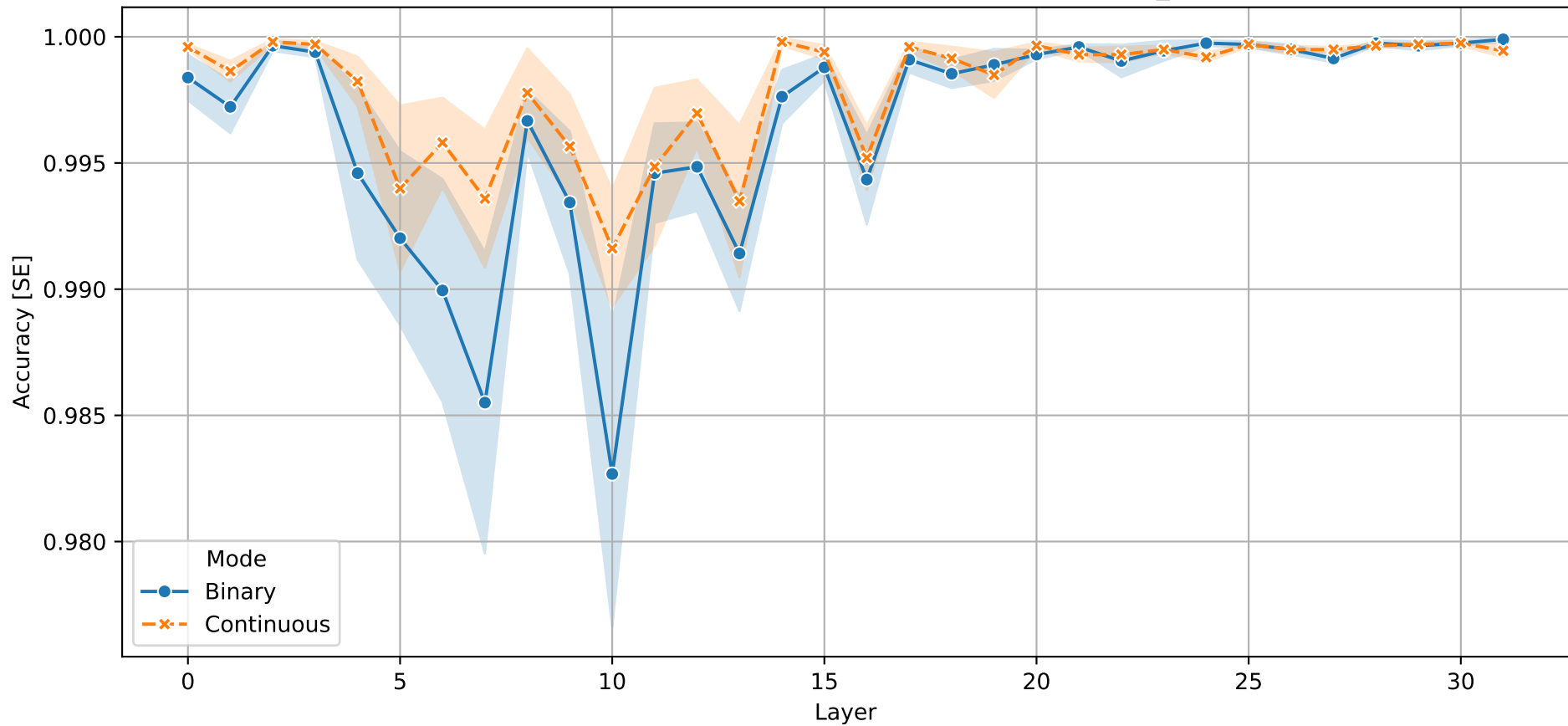| Method	Metric	Binary	Continuous
Full Layer	f1_best_layer	31.0	14.0
Full Layer	f1_max	1.0	1.0
Full Layer	f1_mean	0.9966	0.998
Full Layer	f1_std	0.0072	0.0043
Single Neuron	f1_best_layer	0.0	2.0
Single Neuron	f1_max	1.0	1.0
Single Neuron	f1_mean	0.823	0.862
Single Neuron	f1_std	0.0293	0.0633
Top-K Neurons	f1_best_layer	31.0	24.0
Top-K Neurons	f1_max	1.0	1.0
Top-K Neurons	f1_mean	0.9389	0.9793
Top-K Neurons	f1_std	0.0672	0.0328

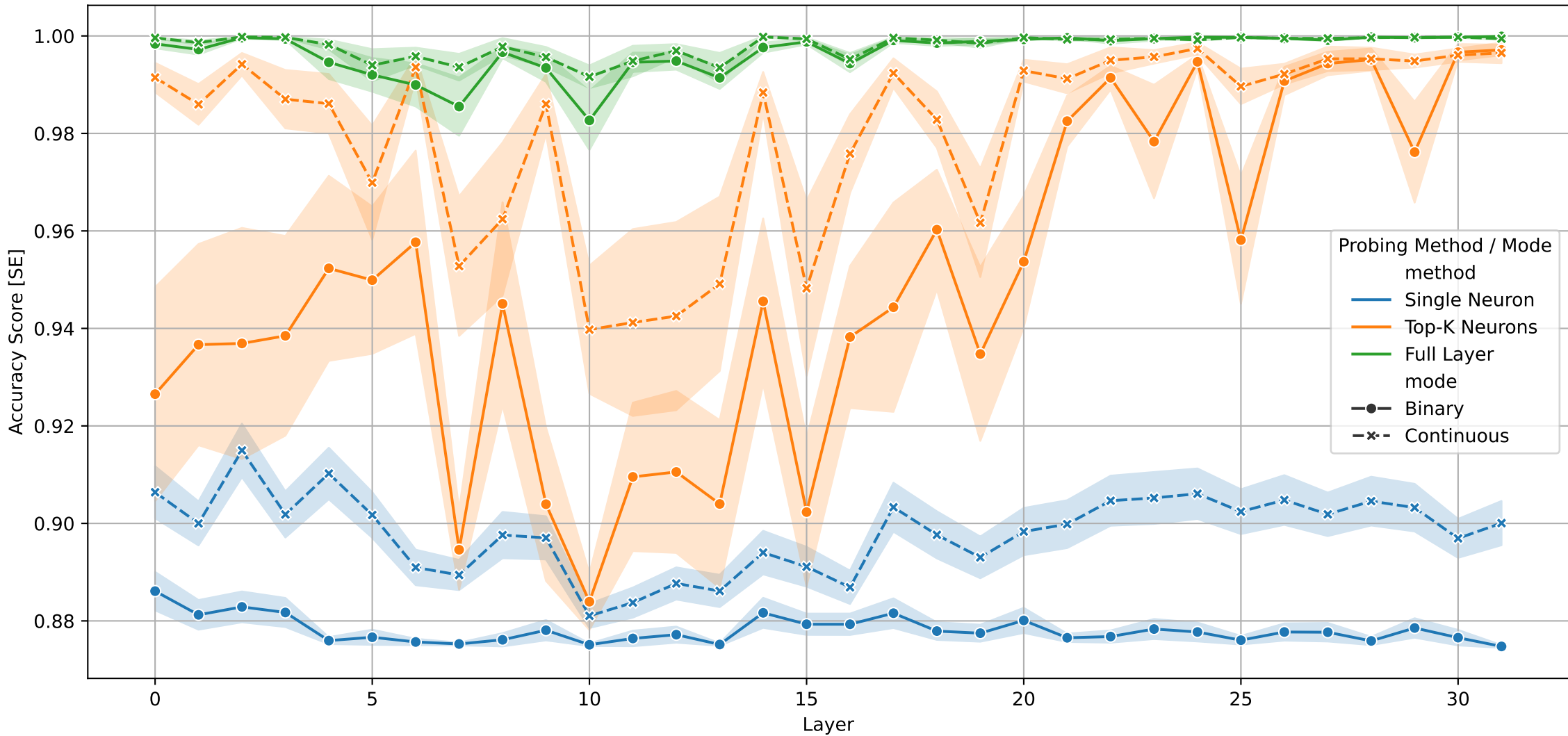Accuracy per Layer – Single Neuron Probing for europarl_lang

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	31.0	2.0
Full Layer	accuracy_max	1.0	1.0
Full Layer	accuracy_mean	0.9966	0.998
Full Layer	accuracy_std	0.0071	0.0043
Single Neuron	accuracy_best_layer	0.0	2.0
Single Neuron	accuracy_max	1.0	1.0
Single Neuron	accuracy_mean	0.878	0.8982
Single Neuron	accuracy_std	0.0177	0.0408
Top-K Neurons	accuracy_best_layer	31.0	24.0
Top-K Neurons	accuracy_max	1.0	1.0
Top-K Neurons	accuracy_mean	0.9495	0.9801
Top-K Neurons	accuracy_std	0.0508	0.0305