In GRUs the current input does not directly modify the value of the hidden state h\_t. This is done through modifications of its two gates: the reset (z) gate, update gate(z) and its intermediate or candidate memory unit h\_tilde. Once calculated, the current hidden state h\_t is then updated by z and h\_tilte. The main differences between GRUs and LSTMs are the number of gates and maintenance of cell states. Unlike the GRUs, LSTMs have 3 gates (input, forget, output) and maintains an internal memory cell state, which makes it more flexible, but less efficient memory and time wise. However, since both of these networks are great at addressing the vanishing gradient problem required for efficiently track long term dependencies. Choosing between them are usually done using a rule of thumb. As such, it is recommended that you first train a LSTM, since it has more parameters and is a bit more flexible, followed by a GRU, and if there are no sizable differences between the performance of the two, then use the GRU.