Optimal data detection of data transmitted over a linear channel can always be implemented through the Viterbi algorithm (VA). However, in many cases of interest the memory of the channel prohibits direct application of the VA. A popular and conceptually simple method in this case, studied since the early 1970s, is to first filter the received signal in order to shorten the memory of the channel, and then to apply a VA that operates with the shorter memory. We shall refer to this general concept as a channel shortening (CS) receiver. Although studied for almost four decades, an information theoretic understanding of what such a simple receiver solution is actually doing is not available. In this paper, we show that an optimized CS receiver has a direct correspondance to the chain rule of mutual information. Furthermore, we show that the tools for analyzing the ensuing achievable rates from an optimized CS receiver are precisely the same as those used for analyzing the achievable rates of a minimum mean-square-error (MMSE) receiver.