Wearable cognitive assistance applications can provide guidance for many facets of a user’s daily life. This thesis targets the enabling of a new genre of such applications that require both heavy computation and very low response time on inputs from mobile devices. The core contribution of this thesis is the design, implementation, and evaluation of Gabriel, an application platform that simplifies the creation of and experimentation with this new genre of applications. An essential capability of this platform is to use cloudlets for computation offloading to achieve low latency. By implementing several prototype applications on top of Gabriel, the thesis evaluates the system performance of Gabriel under various system conditions. It also shows how Gabriel is capable of exploiting coarse grain parallelism on cloudlets to improve system performance, and conserving energy on mobile devices based on user context.