JOHNS HOPKINS
UNIVERSITY

human language technology
center of excellence

HLTCOE
ASR Research Toolkit

The **Human Language Technology Center of Excellence** at **Johns Hopkins University** is developing a toolkit for advanced **Automatic Speech Recognition** (ASR) research. We believe the time is right for a next-generation toolkit that can take advantage of modern developments in high-level languages and libraries, as well as in software development tools and practices. The toolkit uses the Python language and uses the high-performance Numpy package for efficient, large-scale numerical work. It includes comprehensive HTML tutorials and reference material built from extensive documentation and examples closely linked to the source code. The examples in this documentation are also used as part of a large, automated test suite.

Dataflow Architecture

Researchers, scientists, and engineers habitually use flow diagrams to describe their experimental or production configurations. In this regard, tools for ASR research must move large amounts of data between functional modules. But there is often a mismatch between the ease of sketching the data flow and processing logic in a diagram and the effort required to actually implement a new experiment. Traditionally, each novel experimental configuration involves writing customized code and scripts in multiple computer languages so as to get the various data in the correct form and bring them together where the algorithmic logic operates on them. Having to undertake a custom coding and scripting effort for each new line of research is a barrier to rapid progress.

The HLTCOE toolkit eliminates this mismatch with a dataflow architecture in which the graphical connections of a configuration are realized directly by the toolkit. That is, using the toolkit, the researcher builds a graphical structure of connected processing elements. The toolkit handles the details of moving the data between the researcher's elements. This allows researchers to focus more time on creating processing elements with novel algorithms, and spend less time managing data movement. The dataflow architecture also supports building systems that stream data though a processing network continuously and indefinitely from heterogeneous, asynchronous sources. Furthermore, any experimental processing network can be paused at any time, and the live objects, models, and processing elements can be examined and adjusted directly in the Python interpreter.

ASR Research Functionality

The toolkit already includes components and modules for audio sources, signal processing, basic parametric modeling, lexicons, context-free grammars, HMM decoding, graphical and lattice operations, and novel statistical validation tools, as well as numerous low-level utilities. Many components can display themselves visually, e.g., HMM topologies and dataflow networks display themselves as directed graphs. The toolkit includes simple and powerful components for seamlessly vectorizing and distributing almost any time-consuming user function onto a computational cluster. The toolkit is readily extensible, and many other modules are being developed. We are also providing hooks that support integration with other toolkits and external components.

Status and Licensing

We are currently preparing the initial internal release of the toolkit. Our intent is to distribute the toolkit under an open license and to make the toolkit freely available to other ASR and machine-learning researchers.

Interested? Please contact **Ken Basye** kbasye1@jhu.edu or **Hugh Secker-Walker** hsw@jhu.edu for additional information or to be added to our mailing list for upcoming announcements.