The Speech Recognition Virtual Kitchen: Launch Party

Andrew Plummer\textsuperscript{1}, Eric Riebling\textsuperscript{2}, Anuj Kumar\textsuperscript{2}
Florian Metze\textsuperscript{2}, Eric Fosler-Lussier\textsuperscript{1}, and Rebecca Bates\textsuperscript{3}

\textsuperscript{1}Dept. of Computer Science and Engineering, The Ohio State University, Columbus, OH; U.S.A.
\textsuperscript{2}Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA; U.S.A.
\textsuperscript{3}Dept. of Integrated Engineering, Minnesota State University, Mankato, MN; U.S.A.
{plummer.321,fosler-lussier.1}@osu.edu, \{erik,fmetze,anujk\}@cs.cmu.edu, bates@mnsu.edu

Abstract

We present updates to the Speech Recognition Virtual Kitchen (SRVK) environment, a repository of pre-configured Virtual Machines (VMs) containing tools and experiments in the speech and language field. SRVK promotes community sharing of research techniques, fosters innovative experimentation, and provides solid reference systems as a tool for education, research, and evaluation. VMs provide a consistent environment for experimentation, without requiring tedious installation of many individual tools, a web-based community platform complements the VMs, allowing users to jointly explore, learn and collaborate using VMs. In this Show&Tell demo, we present the infrastructure to the speech community, along with several example VMs and a set of online error analysis tools. We solicit feedback from the community, in order to further guide development of the kitchen, which we hope to grow into a widely used community resource.

Index Terms: speech recognition, virtualization, educational tools, research infrastructure

1. Introduction

The depth and breadth of disciplines related to Automatic Speech Recognition (ASR) research and education has long reached a point where serious attention to community organization and infrastructure is of critical importance to its continued development and growth, and potential cross-disciplinary expansion. The following facets face significant challenges:

Basic ASR research and education: Speech recognizers incorporate knowledge from linguistics, phonetics, acoustics, signal processing, statistical modeling, graph theory, and artificial intelligence; expecting students to become experts in all of these areas, before attempting to work on speech recognition systems, is unrealistic.

Advanced ASR Research: Building and maintaining a state-of-the-art ASR system has moved beyond the ability of a single developer; it is difficult for all but the largest of university labs to build or maintain an end-to-end system, and adapt it to new conditions as required.

Cross-disciplinary ASR research and education: The challenges above pose a high bar for developing new research groups, making it difficult for institutions without active ASR researchers to integrate ASR projects into their educational curricula or field research projects which include ASR.

With the Speech Recognition Virtual Kitchen model, we extend the model of lab-internal knowledge transfer and infrastructure sharing to a community-wide effort through the use

2. The Speech Recognition Virtual Kitchen

The Speech Recognition Virtual Kitchen [1] provides the organization and infrastructure for meeting the numerous challenges facing the ASR research community in two key ways. The first is through the establishment of a set of VMs with associated repositories that facilitate the exchange of VMs among members of the ASR community and other interested parties. The second is through a web-based community platform which complements the repositories, and allows physically disconnected users to jointly explore VMs, learn from each other, and collaborate. Using open-source resources, such as Debian Linux derivatives as a platform, Kaldi [2] as a recognizer, and the TED-LIUM corpus [3] as example data, for example, allows us to create an infrastructure that can be freely shared.

The Kitchen repository contains ‘Kitchen’ VMs and software packages prepared by the community and/or the maintainers of the SRVK that are known to work, and for which doc-
A Virtual Worlds VM [1] with a basic speech recognition
A Kaldi [2] Live Decode VM for recognizing English speech,
ing a succession of classes and labs at Carnegie Mellon and
demonstrate a number of VMs including VMs developed dur-
Provider repository
which any new installation (tools, data, etc.) can be verified.
using VirtualBox (https://www.virtualbox.org). The VM may
be fully configured with software and experiments that are used
illustrate a pedagogical concept or advanced research tech-
section presents our current inventory.
Alternatively, the VM may be “bare” to reduce the size of
initial download, and later customized by installing addi-
tional software, data, or tools supporting a certain set of ex-
periments, a tutorial, reference log-files, etc., from either the
Kitchen repository or third party servers.
Having VMs originate from the same configuration elimi-
ates compatibility issues – once an installation script has been
found to work on one VM, it will work on all VMs. Online
discussion forums are also provided for each VM so users of
a specific VM can share their experience.
The user can vary experimental setups by changing the pro-
vided scripts or writing new code, and compare the results to the
provided log-files and baseline results. Moreover, by connect-
ing local resources such as host file systems or microphones,
Kitchen experiments can be adapted to run with new data. Thus,
the Kitchen provides a locally available reference setup, against
which any new installation (tools, data, etc.) can be verified.
When users create useful new VMs (the ultimate goal of
SVRK), or make significant modifications to existing ones, they
may share them with the ASR community through the SRVK
Provider repository. A script is available on the SRVK web-
site that makes it easy to package and upload any scripts, data,
experiments, or virtual machines that users have created. The
Provider repository and upload script together provide a simple
way for users to contribute and share their own recipes and ap-
pliances to the SRVK without having to ship entire VMs. Other
users may obtain Provider VMs as they would ‘Kitchen’ repos-
itory VMs and packages. As a ‘Provider’ VM matures, the
Kitchen organizers will be able to support the community, and
turn it into an “official” Kitchen VM with documentation, etc.
The SRVK provides for broader community interaction via
forums on its web-site, allowing users who downloaded simi-
lar VMs from the SRVK to connect with each other, and dis-
cuss their research, effectively porting a lab-based model of
knowledge transmission to a global scale. Additionally, the
SRVK provides access to a number of online error analysis tools
that facilitate the comparison of experimental results, and could
even support a permanent evaluation mode with eternal “high-
scores” on standardized test sets.

3. Show & Tell at INTERSPEECH 2014
At INTERSPEECH 2014, we will demonstrate for the first time the
SVRK web-site (http://speechkitchen.org/), and solicit community feedback on the interface, the desired
functions, teaching needs, and other desiderata. We will also
demonstrate a number of VMs including VMs developed dur-
ing a succession of classes and labs at Carnegie Mellon and
Ohio State University. Demo VMs will include:
• A Kaldi [2] Live Decode VM for recognizing English speech,
both from pre-recorded input files and from a microphone.
This VM illustrates how SRVK may enable easy distribution
of educational materials, including VMs that contain speech
recognition experiments, data and tool-kits.
• A Virtual Worlds VM [1] with a basic speech recognition
system interfacing with Second Life. This VM exemplifies the
pedagogical and cross-disciplinary thrust of SRVK: students
implemented the VM, improved the dialog capabilities of the
system, added a face detector and emotion recognition, and
performed experiments on a POMDP – all without the need for
a dedicated machine to host the environment.
• A teaching VM with classroom exercises used in teaching a
Speech and Language Technology class. This VM includes
in-class team-based tutorials for building different compo-
nents of simplified speech recognition systems.

In addition to VMs, we will present the current web-based
data and error analysis tools. These could develop into an eval-
uation server, which will accept results on standardized test sets,
and returns results in scientifically meaningful and graphically
pleasant, easy-to-work-with formats. Users can dissect the out-
put of the system that they built, compare it to the results that
other users got on the same dataset, and see if their particular
system is sensitive to noise, makes mistakes for fast speakers,
suffers from out-of-vocabulary words, etc.

4. Outlook
The SRVK is an ongoing effort, funded by an NSF CRI (Com-
munity Research Infrastructure) grant. Several important issues
still need to be addressed. Intellectual property issues often
stand in the way of widespread sharing, since different tool-kits
and data may need to be licensed. We will present to the com-
community several methods of distribution that preserve intellectual
property rights, while a consistent environment will allow end
users to install open or closed-sourced systems and data with
consistent results. VMs also lend themselves to facilitate re-
search in the cloud, as a means of collaboration between groups,
or to easily provide computing resources for workshops – the
SRVK would like to facilitate this.
We will also collect ideas for other scenarios in which this
infrastructure will be useful, including fields that are data in-
tensive (synthesis, dialog systems, NLP, computer vision, data
mining). This may be mutually beneficial, as incubating ASR
in other fields by providing an easy-to-use, non-trivial research
environment will boost the relevance of speech and language
technologies across disciplines. We are interested in accumulat-
ing systems for virtualization from within the ASR community,
as well as those from adjacent fields.

5. Acknowledgments
The authors would like to thank Jesse Seguin, Malav Bhavsar,
and Siddharth Kothari for their contributions to this work.
SVRK is supported by the United States’ National Sci-
cence Foundation under grants CNS-1305365, 1305319, and
1305215.

6. References
kitchen: An initial prototype,” in Proc. INTERSPEECH. Port-
[2] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek,
N. Goel, M. Hannemann, P. Motlíček, Y. Qian, P. Schwarz,
J. Silovský, G. Stemmer, and K. Veselý, “The Kaldi speech recog-
nition toolkit,” in Proc. ASRU. Big Island, HI; USA: IEEE, Dec.
2011.
matic speech recognition dedicated corpus,” in Proc. Eighth In-
ternational Conference on Language Resources and Evaluation
lium.univ-lemans.fr/en/content/ted-lium-corpus.