Unconventional Approaches to Gathering and Sharing Resources for Spoken Dialog Research

Maxine Eskenazi, Sungjin Lee, Tiancheng Zhao, Ting Yao Hu, Alan W Black

Language Technologies Institute – Carnegie Mellon University

5000 Forbes Ave Pittsburgh PA 15213 USA

E-mail: max,awb @cs.cmu.edu, tianchez, tingyaoh@andrew.cmu.edu, junion@yahoo-inc.com

Abstract

The DialRC and DialPort projects have employed unconventional approaches to data gathering and resource sharing. The projects started sharing by distributing the speech, transcription and logfile data gathered by the Let's Go system. That system has responded to over 220,000 calls from real users of the Allegheny County Port Authority. The Let's Go platform proved to be a very successful way to run studies, with a dataflow of about 1300 dialogs per month. Thus, DialRC built a research platform that was used by other researchers, enabling then to run studies with the Let's Go real users. Challenges were also run on this platform. Finally DialPort follows in the footsteps of DialRC by creating a spoken dialog portal with real users that other dialog systems can be connected to. This paper examines the impact that these activities have had on the spoken dialog research community.

Keywords: spoken dialog, portal, data sharing

1. Introduction

Over the past ten years, the Dialog Research Center has taken unconventional approaches to gathering and sharing resources. The Center has focused on providing the means for researchers from other sites to share data and run studies. Data gathering is a novel approach with speech from real spoken dialog system users being logged and distributed. The novel approach to running studies centers around opening up access to the Let's Go platform by both distributing its software and inviting researchers from outside Carnegie Mellon to run studies on it.

Data gathering and sharing began with the Let's Go project (Raux et al 2006). Working with the Port Authority of Allegheny County (PAT), the team created a telephone-based spoken dialog system that answers the phone for PAT callers in the evenings and on weekends. It gave bus schedule information when humans were not working. The system went live on March 5, 2005 and is still functioning. It has been "live" every day except one (machine room flooding) and has expanded to 24/7 availability. In the fall and winter of 2010, the coverage was expanded from 10 bus routes in the East End of the city to 60 routes that pass through the East End during some part of their trip. In the summer of 2014, it became directly accessible via a phone number that has been advertised on the buses. The system has logged over 220,000 dialogs over its 11 years of existence.

The dialog with Let's Go is relatively simple, but has been found over the years to still be complex enough to study interesting research issues. The users provide a time, a departure stop, a destination stop and, optionally, a bus route and the system provides the times and the number of the appropriate bus. The system deals with a wide variance on how bus stops are described by users and it must respond to real callers and the consequences of real telephone background noise from crying babies, loud TVs, and traffic noise.

The team that built Let's Go carried out many studies on it. It was later made available to others in the community. In this way, Lets Go has provided: real user data; the system software; and a platform on which to run studies.

2. The Dialog Research Center Activities

The Dialog Research Center (DialRC) was formed with a grant from the National Science Foundation. The goal was to make the Let's Go products widely available to the research community. The following sections give more detail on those products

2.1 Speech and logfile data from the live Let's Go system

The data consists of speech files, both of the whole dialog and also of each user turn. They are accompanied by the system logfiles for the corresponding dialogs. An interface relates the two, where a summarized logfile is viewed and each

corresponding turn can be heard. The data is available in two forms. There is a small test set that is directly available on the web. The over-700GB dataset is available by exchange of hard drives. The latter has so far been distributed to 17 groups throughout the world. Some of the data has been labeled. About 6 months of calls have been labeled by an expert. One year of calls has been labeled through crowdsourcing (Parent and Eskenazi, 2010). Other data was labeled for use in the Challenges described below.

2.2 MyBus software

A simplified version of the Let's Go system was created to be used to teach students about spoken dialog systems. MyBus uses that Olympus spoken dialog architecture (Bohus et al 2007) and was first distributed at a tutorial presented at HLT 2008 (Raux et al 2008). The software is downloadable, has a wiki for discussion and members of the DialRC group field questions from its users. MyBus could also be integrated in a course on spoken dialog. Since the software provides a good introduction to Let's Go, some researchers have used it to prepare for the creation of a full blown system that they later ran for a study on Let's Go live.

2.3 The research platform

DialRC made the Let's Go system available to other researchers. They prepared their version of the system and the DialRC team tested its robustness. When the system passed, being at least as robust as the live system at the time, it was allowed to "go live". The platform was used for tests of real vs paid users (Ai et al 2007) and of lexical entrainment (Stoyanchev 2009, Stoyanchev and Stent 2009), for example. It took time for researchers to accept the vision of a commonly-shared platform. When it was finally accepted as a new paradigm, it became the source of the Spoken Dialog Challenge (Black and Eskenazi 2009).

2.4 The Spoken Dialog Challenge

The Spoken Dialog Challenge (SDC) (Black et al 2010, Black et al 2011) was designed to bring together spoken dialog system researchers on a common task. Since comparisons of dialog and evaluation techniques are hard to carry out between different systems, different domains and different user populations, the goal was to offer one domain and user population that allows more common bases of comparison. The goal was also to provide large quantities of real user data as the basis of comparison. The Challenge was not seen as a competition, but rather a comparison of methodologies. Thus each group built a Let's Go system of its own and ran it live on the Let's Go phone number.

The Spoken Dialog Challenge 2010 was divided into three stages: development, control testing, and live testing. For development, the full source code for the system was released as well as the data (the text logs as a download, and the text plus audio as a disk mailing service). 10 groups received the data. Participating groups could use their own dialog architecture if they desired.

Even groups who didn't build on the existing Let's Go source code found the Let's Go language models, grammars, etc. very useful. In the end, four groups produced working systems for the control tests (two universities in the UK, one industry research lab in the US, and CMU's base system). The control tests used spoken dialog experts to call each system with a given scenario. Although the completion rate was higher than for live tests, most of the callers were not from Pittsburgh, and many were non-native speakers (or spoke non-US dialects of English).

The initial results of the systems (on the control tests) were presented at a well-attended special session at SLT2010 in Berkeley, CA in December 2012, while the final live test results (after hand-labeling all of the dialogs) were presented at SIGDIAL2011 in Portland, OR in June 2011. Although WER rate generally correlates with task completion, there were different system orderings for task completion depending on control or live tests. This again highlights the observation that optimization for lab test subjects may not reflect the outcome with real users.

The second Challenge, SDC2011, had a total of 4 participating systems, though these came from only two institutions. This allowed closer comparisons of specific system differences, but did not offer the breadth of systems that participated in the first year.

The clear theme of the SDC2011 participating systems was **dialog state-based** techniques. Although some general changes were necessary in the default system from the first year (due to schedule changes, and increased route coverage), the default system was fundamentally the same as SDC2010 so some cross-year comparison was possible. Since both teams had had experience in producing robust systems, control tests were not carried out. The live tests took place from December 2011 to February 2012. The two groups (four systems) taking part in SDC2011 submitted result papers to SLT2012 (Miami, FL, December 2012).

The rise of the interest in dialog state during this

Challenge gave rise to a new type of Challenge, the Dialog State Challenge.

2.5   The Dialog State Challenge

The first and second Dialog State Tracking Challenges (DSTC) (http://research.microsoft.com/en-us/events/dstc/) is a follow on from Spoken Dialog Challenges 2010 and 2011. A number of researchers in the domain wanted a means to have better comparisons and to accurately estimating a user's goal in a spoken dialog system. Having a common task and a common large dataset answered this need. The results of the Challenge were presented at SIGDIAL 2013. They used the Let's Go data and DialRC provided support for the data and its annotation.

2.6   The impact of DialRC on the spoken dialog research community

Since the goal of DialRC was to serve the community, its success can be measured by how much its products were used by the spoken dialog community.

In order to determine impact, a targeted search of the literature in spoken dialog was conducted. This reflects both how aware the community is of the DialRC approaches (gaining mention in a paper) and whether they have actually used the products (paper results being based on their use). As mentioned above, those products are:

- the distributed speech, labels and system log data,
- the MyBus/Let's Go system,
- studies run on the Let's Go platform,
- and participation in the Spoken Dialog Challenge and/or the Dialog State Challenge.

The assessment below, carried out in 2013, (maring the end of the DialRC funding) refers to papers found using keywords such as "Let's Go", "DialRC", names (authors of the Let's Go papers), and "dialog challenge". For publications between 2008 and 2012, a total of 216 references (non-CMU publications) were found. Figure 1 shows the total number of publications by year. There is a steady increase over the years.

Figure 1 indicates that there is significant awareness of the DialRC products within the research community. To determine whether the products were being adopted and actually *used* for publications, the papers were read by the DialRC team. Figure 2 breaks the data in Figure 1 down into two parts: the references that simply *mention* DialRC products and those that actually *use* them. We see that mention of

DialRC started out strong in 2008 and increased from 2009 to 2012. In 2008 there were few authors who actually *used* DialRC products, but this increased in the following years. It is interesting to note that some of the authors, who only mentioned the products one year, went on after that to actually use them. In 2012 the number of product users seemed to have leveled somewhat, while the total number of publications increased.

More detailed examination of the publications from 2009 and 2012 reveals a wide range of topics. The dialog research community has varying and changing interests (e.g. from simulated users to multiparty dialogs). The references to DialRC covered *eleven different topics*, as seen in Figure 3. Although a large portion (116) were about system architecture, we note that discourse (19), user behavior (20) and evaluation (32) were also well-represented.
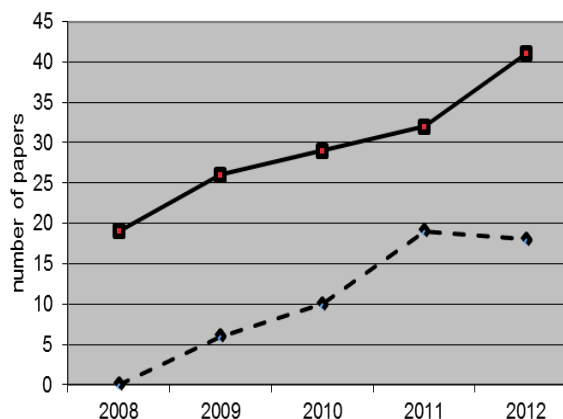

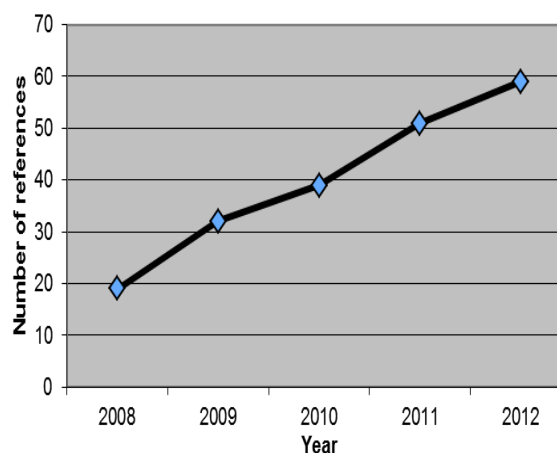Figure 1. References by year of appearance

Figure 2. Total references (dashed line) that *use* the DialRC products and total references that *mention* (solid line) the products by year.
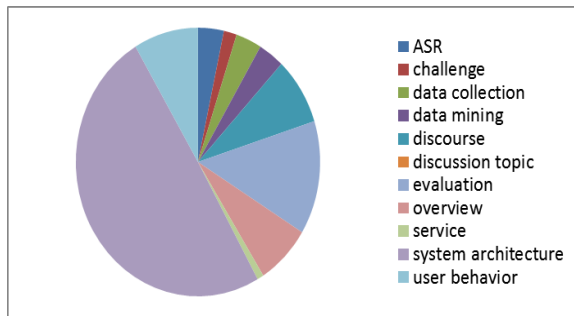
Figure 3. DialRC topics in spoken dialog research –
from the 2009-2012 literature

In the long term, DialRC use should result in substantial contributions, such as journal papers, theses and book chapters. Although we see that conference (112), symposium (5) and workshop (44) papers have indeed been the most prevalent (two of the conference papers are main keynotes at Interspeech, Steve Young 2010, Julia Hirschberg, 2011), we also note a reasonable number of references in journal papers (40) and book chapters (12). And, interestingly, there are many theses (18). Nine of the theses actually used the system or the data as an integral part of the thesis work.
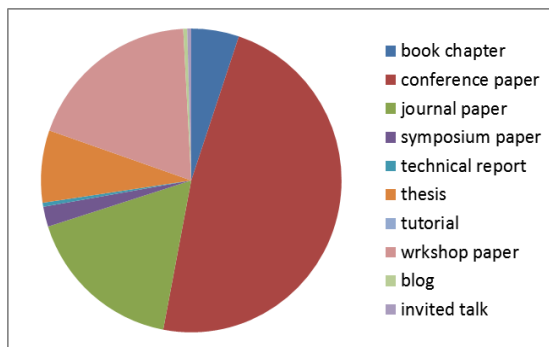

Figure 4. DialRC and types of publications

2.7   DialRC products beyond 2012

Even after the end of DialRC funding, researchers have continued to use its products. For the period from 2013-2015, Figure 5 shows the persistence of the influence of DialRC. The results shown above for 2012 are included in this Figure for reference. There are a total of 92 papers that mention DialRC products over this three year period. There are 31 that actually use the products. We see that there is a gradual decline in the number of papers mentioning the products and a decline and then steady actual use of the products. This is partially due to the Dialog State Challenges and to the distribution of a Let's Go user simulation. It can also be attributed to a database that was built in Germany using the Let's Go material.
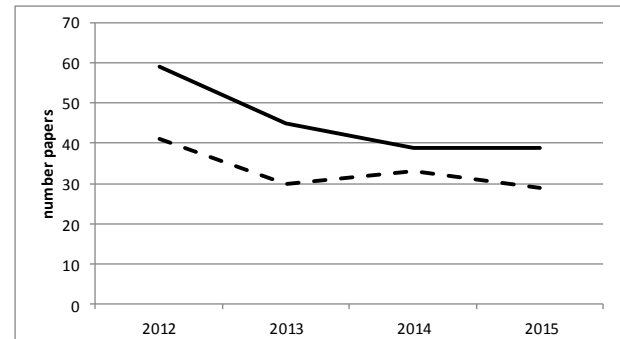

Figure 5. Use of the products after the end of the DialRC grant– solid line is all papers mentioning the DialRC products, dashed line is papers using the products.

3.   A Spoken Dialog Portal

We see from the large number of publications that the products distributed by DialRC fulfilled a need in the spoken dialog research community. But the community and its research needs evolve. In the years since DialRC and the study above, there have been more spoken dialog systems, spurred on by the advent of SIRI and other personal assistants.

While it is relatively easy for industrial systems to get real users, academia has more difficulties. And the academic systems are very diverse, going from web-based to phone-based to app-based to robot-based. Given the DialRC team's long experience with real users, it was natural to evolve from getting users for one system to getting users for the research community. The concept of a portal began to take shape. The idea was to have a single user-directed interface that would link all possible academic systems. In this way, the cost of attracting users could diminish and the added-value to the user of finding many different types of assistants would be attractive. The DialPort project was born and Skylar came into existence.
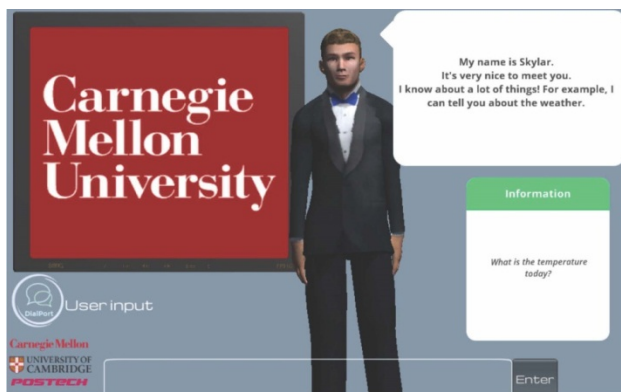
Figure 6. Skylar, the butler, is a portal where real users can speak to academic dialog systems

Skylar (Figure 6.) is the avatar of the portal. The user meets Skylar upon entering the portal and it is Skylar's job to determine what the user wants and to convince them to try all of the connected spoken dialog systems. Since keeping a real user engaged involves providing a large quantity of varied information, Skylar knows about the weather and hotels and restaurants. It is also endowed with a chatbot. But every few turns, it tells a user that it has a friend who can give them information about, for example, hotels in San Francisco (the Cambridge University system). It encourages the user to ask about that information. When the user does ask about another system, another avatar appears (chosen by the system developers amongst the characters available in the Unity software) to talk to the user. The transfer appears seamless to the user and the control of the dialog goes back to Skylar when the other dialog is finished.

Characterized as a butler, Skylar's movements are coded to resemble those of television butlers so that its positions, for example, when it is listening, are easy to interpret by any user. At present Skylar is linked to the Cambridge system. In the summer of 2016, it will be linked to Let's Go. In the fall of the same year, it will be linked to two more systems. Just as the first connection was an interesting challenge, the Cambridge system and the Let's Go system are phone-based. Later connections involving apps and/or robots should provide further interesting challenges.

4. Conclusion

The DialRC products and the DialPort activities are creating novel research opportunities. The platform has given the DialRC team a large real user base and has afforded many studies that could not have been carried out without this quantity of data. Annotating the data gave the team experience in crowdsourcing. Finally the portal is giving the team experience in interfacing systems of very different natures.

6. Bibliographical References

Ai, H.; Raux, A.; Bohus, D.; Exkenazi, M.; and Litman, D. 2007. Comparing Spoken Dialog Corpora Collected with Recruited Subjects versus Real Users. *In Proc. 8th SIGDial Workshop on Discourse and Dialogue,* Antwerp, Belgium.

Black, AW, Burger, S, Conkie, A, Hastie, H, Keizer, S, Lemon, O, Merigaud, N, Parent, G, Schubiner, G, Thomson, B, Williams, JD, Yu, K, Young, S, and Eskenazi, M, 2011, *Spoken Dialog Challenge 2010: Comparison of Live and Control Test Results*, Proc. SIGDIAL 2011, Portland.

Black, A., Burger, S., Langner, B., Parent, G., Eskenazi, M. (2010). Spoken dialog challenge 2010. Proceedings of SLT 2010, Berkeley, CA,

Black, A.W. and Eskenazi, M. (2009). The spoken dialogue challenge. Proc SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL '09), London, UK. pp. 337-340.

D. Bohus, A. Raux, T. Harris, M. Eskenazi, and A. Rudnicky, 2007, Olympus: an open-source framework for conversational spoken language interface research, HLT-NAACL 2007 workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technology, Rochester, NY, USA.

Eskenazi, M., Black, A. W., Raux, A., Langner, B. (2008). Let's Go Lab: a platform for evaluation of spoken dialog systems with real world users. Proc Interspeech 2008, Brisbane, Australia, p. 219.

Parent, G. and Eskenazi, M. (2010). Toward better crowdsourced transcription: Transcription of a year of the Let's Go bus information system data. Proceedings of SLT 2010, Berkeley, California.

Raux, A., Bohus, D., Langner, B., Black, A., Eskenazi, M., 2006, Doing Research on a Deployed Spoken Dialogue System: One Year of Let's Go! Experience, Proc. Interspeech 2006, Pittsburgh.

Raux, A., Langner, B., Black, A. and Eskenazi, M. (2008) Building practical spoken dialog systems.

Proceedings of ACL/HLT 2008 Tutorial, Columbus, Ohio, United States.

Stoyanchev, S. (2009). Impact of responsive and directive adaptation on local dialog processing. PHD Thesis, State University of New York at Stony Brook, NY, United States.

Stoyanchev, S. and Stent, A. (2009). Lexical and syntactic priming and their impact in deployed spoken dialog systems, In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers (NAACL-Short '09). Boulder, CO, United States, pp.189-192.

7. Language Resource References

The Let's Go data is distributed by the DialRC. https://dialrc.org/.