USC Institute for Creative Technologies

University of Southern California

Toward Fluid Conversational Interaction in Spoken Dialogue Systems

David DeVault

University of Southern California
Adjunct Research Assistant Professor

Ementive Systems, LLC
Founder

2016-11-05

The work depicted here was sponsored by NSF and the U.S. Army. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

12 Years of Spoken Dialogue Systems Research

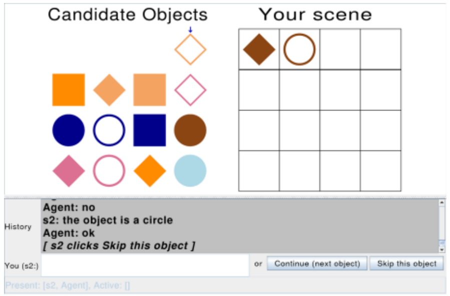
Conflict Resolution Agent
(Gratch et al., 2016;
DeVault et al., 2015;
Gratch et al., 2015)

Eve Agent
(Paetzel et al., 2014, 2015;
Manuvinakurike et al., 2015-2016)

SimSensei Kiosk
(DeVault et al., 2014;
DeVault et al., 2013)

SASO4 scenario
(Plüss et al., 2011; DeVault & Traum, 2013; Traum et al., 2012)

SASO-EN scenario
(Traum et al., 2008;
DeVault et al., 2009-2011)



Agent: no
s2: the object is a circle
Agent: ok
[s2 clicks Skip this object]

COREF
(DeVault & Stone
2005-2009)

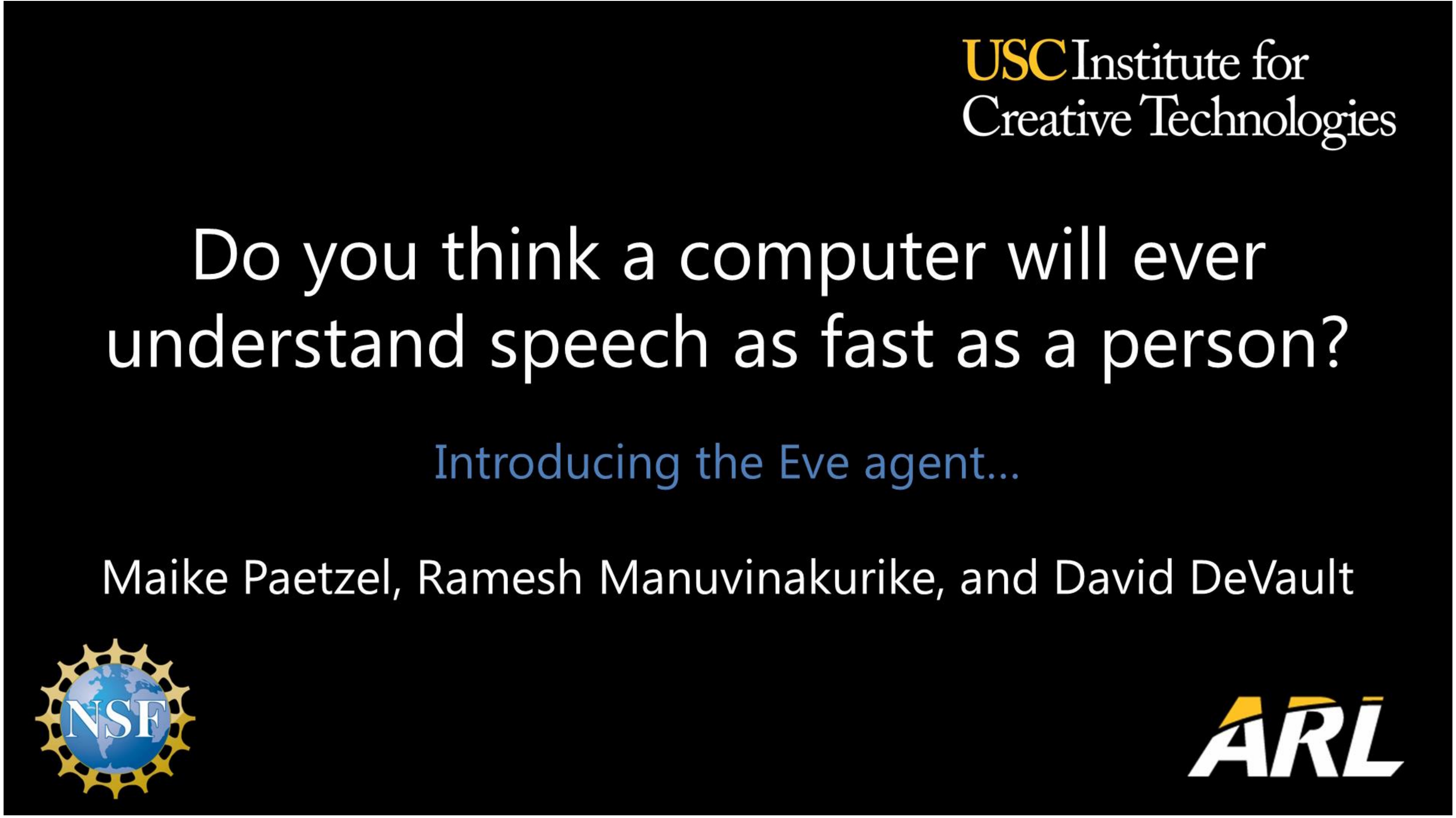
Major “Uphill Battles” for Spoken Dialogue Systems

- ~~Automatic speech recognition~~
- Broad coverage semantics
- Multi-domain / multi-application dialogue policies
- **Fluid conversational interaction**
  - **Turn-taking / mixed-initiative**
  - **Incremental (word-by-word) speech processing**
  - **Dialogue modeling**

What isn’t “fluid” about talking to current SDSs?

- Nearly all SDSs use simplistic turn-taking protocols
  - “Ping-pong” assumption (one DA per turn, no overlapping speech)
  - All user-initiative / all system-initiative
- Users can’t tell if systems are understanding them
  - High response latency, no backchannels (“uh huh”, nods)
- Users don’t know when they can speak or what they can say
  - Single questions or commands: okay
  - Anything else: completely unpredictable
  - Interaction easy to derail
  - Every single utterance is a heavy-weight decision for users

Example 1: The Eve Agent



USC Institute for Creative Technologies

Do you think a computer will ever understand speech as fast as a person?

Introducing the Eve agent...

Maike Paetzel, Ramesh Manuvinakurike, and David DeVault

NSF

ARL

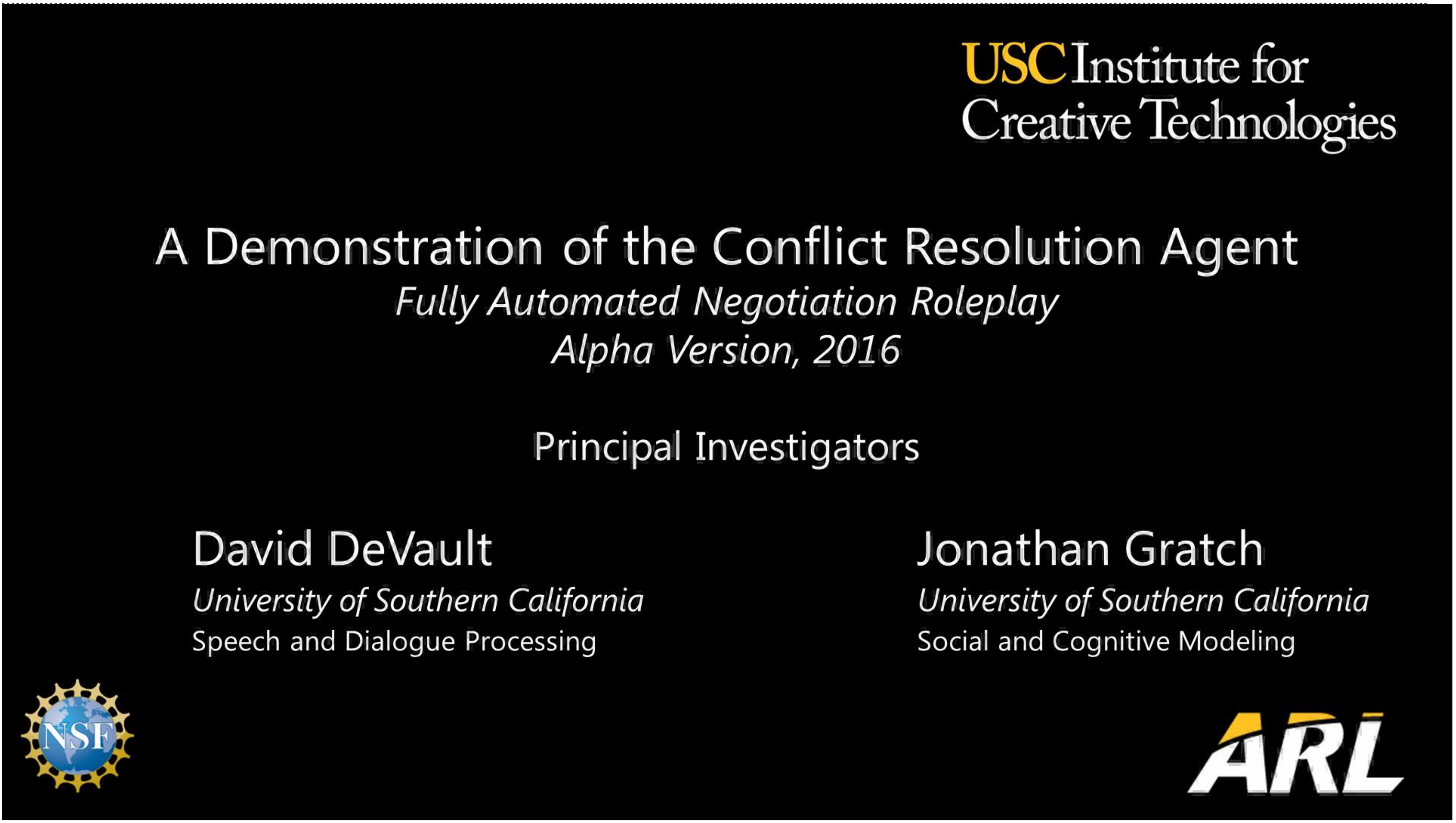
(Paetzel et al., 2014, 2015; Manuvinakurike et al., 2015-2016)

What’s interesting about Eve?

- Users strongly prefer this agent to versions with higher response latency
  - Perceptions of efficiency, understanding, naturalness

    (Paetzel, Manuvinakurike, and DeVault, SigDial 2015; Best Paper Award)
- *In small domains we can use modest amounts of data to build systems that understand user speech very well and very quickly*
- But what about domains where richer models of understanding and turn-taking are needed?

Example 2: The Conflict Resolution Agent



USC Institute for Creative Technologies

A Demonstration of the Conflict Resolution Agent

Fully Automated Negotiation Roleplay

Alpha Version, 2016

Principal Investigators

David DeVault
University of Southern California
Speech and Dialogue Processing

Jonathan Gratch
University of Southern California
Social and Cognitive Modeling

(Gratch et al., 2016; DeVault et al., 2015; Gratch et al., 2015)

What’s interesting about this?

- Support for a wide range of utterance types
- Mixed-initiative
- Fairly fast-paced interaction

How do we make progress?

- Stop making simplistic assumptions about turn-taking and the structure of individual turns
- Use better models of time in interaction
- Develop more extensive, more general, more data-driven dialogue models
- More and bigger human-human conversation data sets

Thank you!

devault@usc.edu