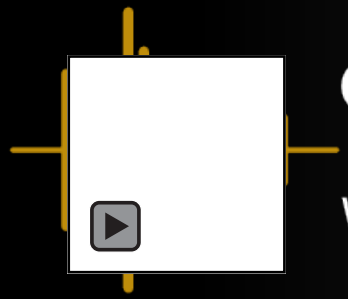


OPEN VOICE NETWORK

Voice worthy of user trust

INTEROPERABILITY PROTOCOLS FOR CONVERSATIONAL ASSISTANTS

Presented by the Open Voice Network Architecture Work Group

March 22, 2023

Moderator: Deborah Dahl, Conversational Technologies

Speakers:

David Attwater, TalkMap

Emmett Coin, ejTalk

OPEN VOICE NETWORK
Voice worthy of user trust

This meeting will

be recorded

end on time

THE LINUX FOUNDATION

be conducted in accord with the antitrust guidelines of the Linux Foundation

What You'll Receive Today

Today, we unveil

Approach
Definition
Proposed Specifications
for
Delegation of Dialog Events

A key building block for agent-to-agent interoperability

- About the Open Voice Network
- The OVON's interoperability vision & work
- Demonstration of interoperability between two conversational agents
- Dialog Event Specifications
- How you can get involved
- Discussion and comments
  - *Please add your comments and questions to the chat*

Introducing the Open Voice Network

A non-profit, vendor-neutral community of the Linux Foundation

The Enterprise Users of Conversational AI
And those who build and develop for them

That is facing these issues

Serving this constituency

- GEN AI & LLMS NEW COMPLEXIT
- CLAIMING OF DATA
- BIG TECH WALLED GARDE
- ONE PLATFORM ONE ASS'T CAN'T DO IT ALL
- ABSENCE OF STANDARDS
- CONNECT OUTSIDE THE FIREWALL?
- DEEPFAKES AUTHENTICATION

The Open Voice Network is working toward

Voice that works like the web

Through this

Technical Standards

Voice worthy of user trust

Behavioral & Compliance Guidelines

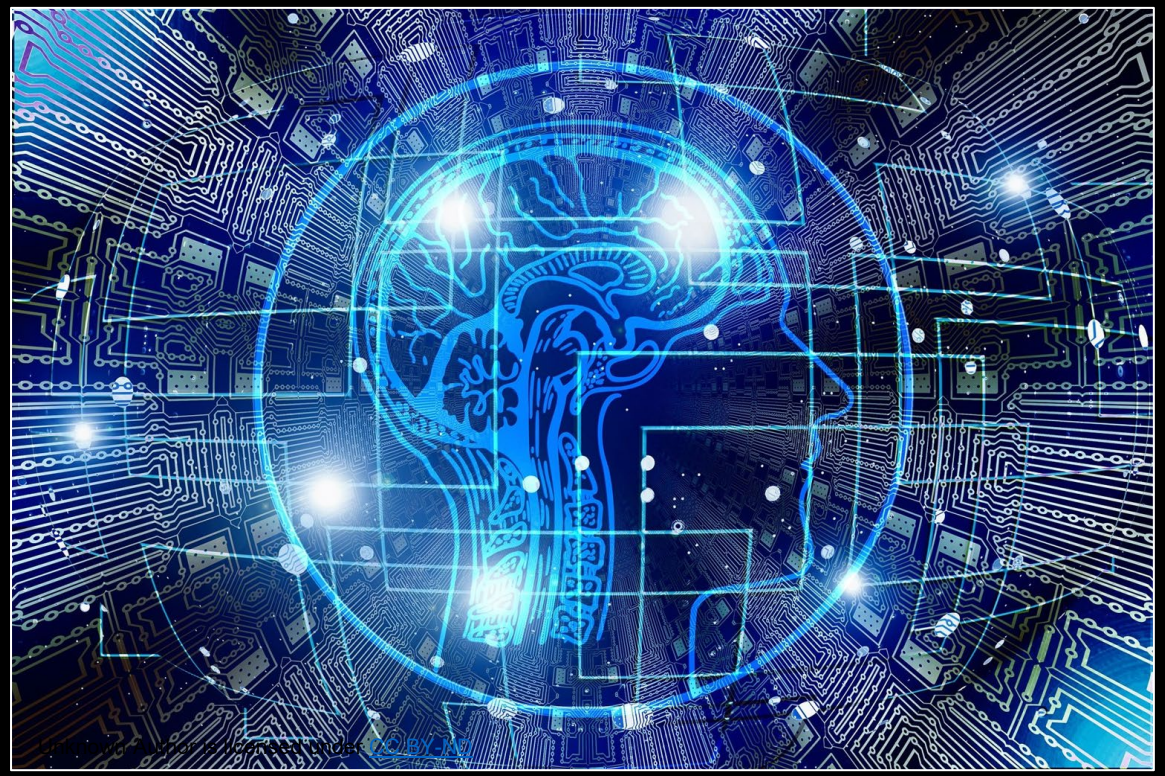
Trust Mark Initiative

2023.03.14

Why should conversational assistants be able to interoperate?

- Users are rapidly adopting conversational assistants which they use to conduct transactions, obtain information, or to control devices
- No single assistant can do everything that a user might need
- When an assistant reaches the limits of its expertise, the user can only try to find another assistant and start over
- Interoperable assistants collaborate so that one agent can pass on a task to a second agent
- This requires that assistants share a common protocol for sending information about tasks – a standard

Why do we need interoperability in the age of Large Language Models (LLMs)?



- LLMs will inspire entrepreneurs to create many more conversational assistants, which will make interoperability even more valuable
- LLMs do not know information about you
- This information is password protected or behind paywalls
- LLMs are deployed in the cloud, but not all applications are suitable for cloud deployment
- There are millions of existing conversational assistants – it will be a long time, if ever, before they're all converted to use LLMs

OVON Conversational Assistants Standards: Where We Are

- Published description of interaction patterns
- Initial demonstration
- Draft specifications for dialog events

Interaction Patterns: The Cornerstones of Interoperability

OPEN VOICE NETWORK

- How can assistants cooperate to satisfy users' goals?

One agent consults behind the scenes with another – Mediation

One agent speaks for another – Channeling

One agent hands off the conversation to another – Delegation

Find out more:
https://openvoicenetwork.org/publications/

Today’s Topics

- Demo of delegation interaction mode
- Dialog events specification

OVON Assistant Interoperability Demo

Emmett Coin, ejTalk

Overview

- Playground
  - Experiments with Methods and Formats
- Real Interactions
  - Multiple Assistants and Servers
  - Prototype Event Exchanges
- Model of Operation
  - WWW for Assistants
  - Lots of Sites/Pages

You will See:

- Server Window, Client Window, OVON Events
- Simple Delegation Scenario
  - No Play-by-Play, Fun, Just to “get the feel”
- Practical (boring?) Scenario
  - Walk-thru, Color commentary
- In Depth View and Discussion
  - Sequence Diagram, Architecture

OVON Dialog Events Specification

David Attwater, TalkMap

What is a dialog event?

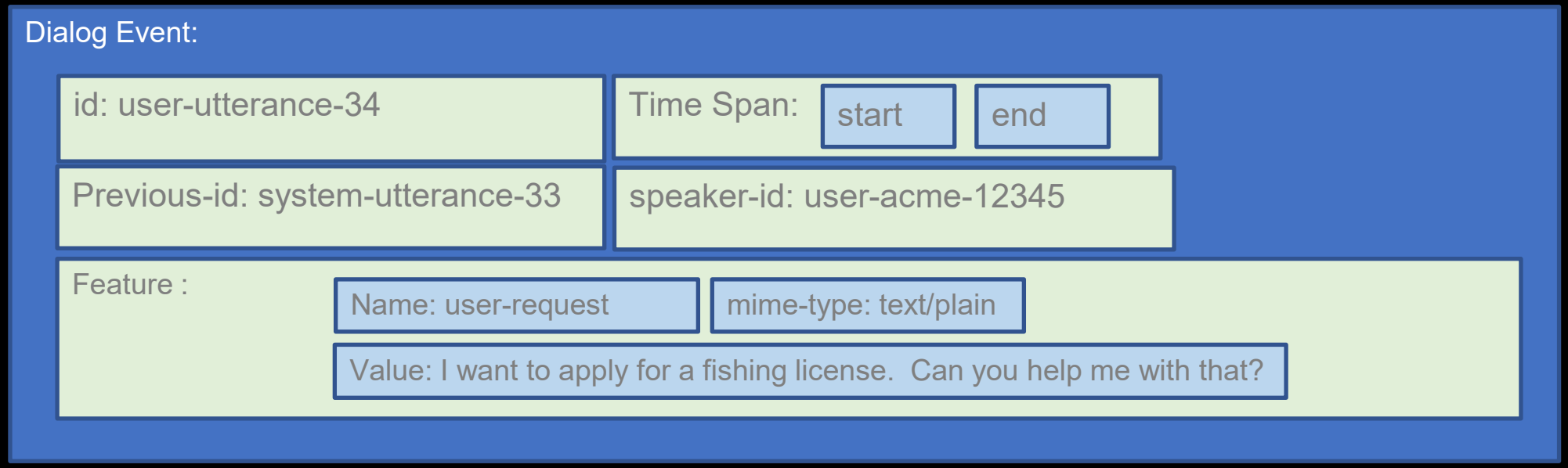
OPEN VOICE NETWORK

- A standard extensible data structure
- Represents an event in a dialog
- Belongs to a single speaker
- Spans a period of time
- Contains one or more ‘features’ related to the event
- Each features has a defined type (e.g.)
  - Audio
  - Text
  - Semantics
  - Pronunciation
  - Gesture
  - Pen input etc.
- Some of these will be OVON standards
- Custom Types are also allowed

Dialog Event

An extensible, standardized way to representing user or system actions in a dialog

A Simple Dialog Event



Example: Delegation

OPEN VOICE NETWORK

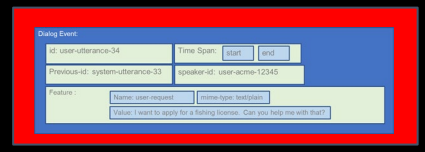
User

Client

Agent A

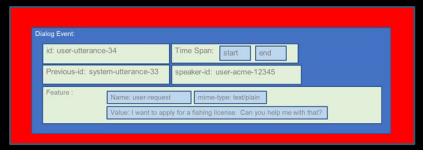
Agent B

I want to apply for a fishing license..



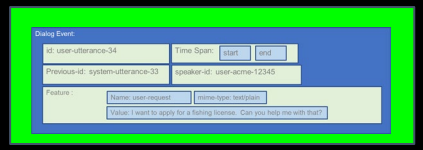
I want to apply for a fishing license..

I'll pass you to the licensing agency.



I'll pass you to the licensing agency.

This is the licensing agency agent. In which district …



This is the licensing agency agent. In which district …

Where could they be used?

- Dialog Events will be embedded in higher-level messages
- Passing Dialog Events to and from agents
  - User Requests to agents
  - System Responses from agents
  - Agents speaking to each other using language
- Keeping Dialog History
  - Passing History between agents (e.g. for Context or Generative AI)
  - Archiving Conversation Transcripts (including human-human transcripts)
- Interfaces to Speech-To-Text API services
- Interfaces to NL Interpretation API services

Events are extensible

Dialog Event:

id: user-utterance-34

Time Span: start end

Previous-id: system-utterance-33

speaker-id: user-acme-12345

Feature:

Feature: audio

mime-type: audio/mp4

Value Ref: http://localhost/zyx1245.mp4

Feature :

Feature: user-request

mime-type: text/plain

Value: “I want to apply for a fishing license”

Feature :

Feature: user-meaning

mime-type: ovon-semantic-1.0

Confidence: 0.99

Value:

Intent: ApplyLicense

Slots: License_Type: fishing Etc.

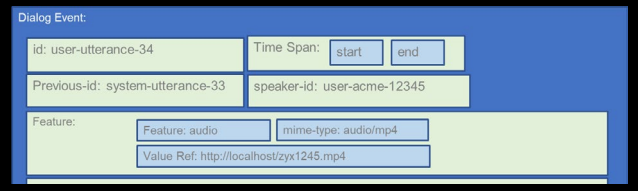
Feature :

type: proprietary-model

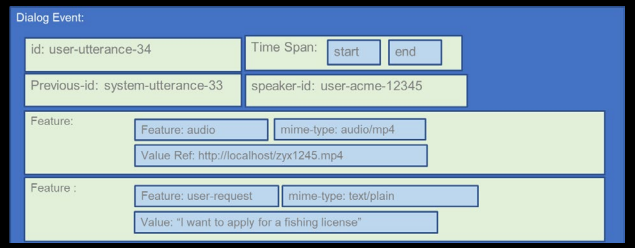
Value: proprietary content

OPEN VOICE NETWORK

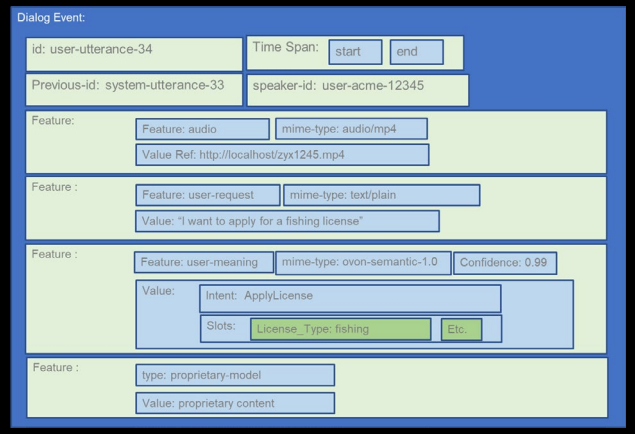
Events are Extensible



Audio from end-point



Text added by speech-to-text



NL Interpretation added by NL Model

What will it look like?

Dialog History

No limit to the number of speakers

Previous utterance from same speaker


```
{
    "history": [
        "{'speaker-id': 'b5y09lky5KU5', 'id': 'user-utterance-30', 'previous-id': 'user-utterance-28', 'span': {"start_time": "2022-12-20 …
        "{'speaker-id': 'b5y09lky5KU5', 'id': 'user-utterance-30', 'previous-id': 'user-utterance-28', 'span': {"start_time": "2022-12-20 …
        "{'speaker-id': 'b5y09lky5KU5', 'id': 'user-utterance-30', 'previous-id': 'user-utterance-28', 'span': {"start_time": "2022-12-20 …
        "{'speaker-id': 'b5y09lky5KU5', 'id': 'user-utterance-30', 'previous-id': 'user-utterance-28', 'span': {"start_time": "2022-12-20 …

        "{'speaker-id': 'b5y09lky5KU5', 'id': 'user-utterance-30', 'previous-id': 'user-utterance-28', 'span': {"start_time": "2022-12-20 …
    ]
}
```


No fundamental difference between machine and users

Utterances from different speakers can overlap

This format can also be used to represent transcripts of human-human dialog, for example conversations between agent and customer in a contact center

Dialog Event Summary

A standardized container that can carry standard and non-standard data

Embedded within higher level specifications

Simple things are simple

Hard things are possible

Agents do not need to know anything about each other

- Standardized features allow interworking

But agents can also have tightly coupled interactions

- Proprietary features allow richer custom interactions

Requirements Specification Published by the Dialog Events Working Group:
https://docs.google.com/document/d/18K17gb_sdj8dr3KB72hc0AVpVF9nRdGQWdjbe5AUjaw

Next Interoperability Goals

- Write specifications for high priority capabilities
  - Dialog event management
  - Inter-agent protocols
  - Data sharing
  - Data packets
  - Context and history
- Demonstrate specifications and use cases with increasingly capable proof of concept demos -- You've just seen Demo 1
- Recruit external partners to collaborate on testing our specifications with their use cases

OPEN VOICE NETWORK

Discussion

please put your questions in the chat

Following up

- JOIN THE DEVELOPMENT DISCUSSION. During our weekly session (Tuesdays, 11 a.m. Eastern) or asynchronously via Slack. This is worthy of the best brains in the industry. We would love to discuss your participation with you, and at your convenience.
- JOIN THE OVON DEMONSTRATION PROGRAM. This is the first of many proposed specifications. We need you to take them, use them, test them. You have our e-mail addresses here on this slide – please reach out to us via email or LinkedIn for a private conversation.
- The recording of this webinar will be available shortly on the OVON website.
- Additional questions can be sent to
  - hello@openvoicenetwork.org
    Or via the form at https://openvoicenetwork.org/contact/
- You will receive a questionnaire from Open Voice following this webinar. Please fill it out at your earliest convenience.