# IPA ARCHITECTURE

Version 0.1

30/09/2019

**Author:**
Dirk Schnelle-Walka

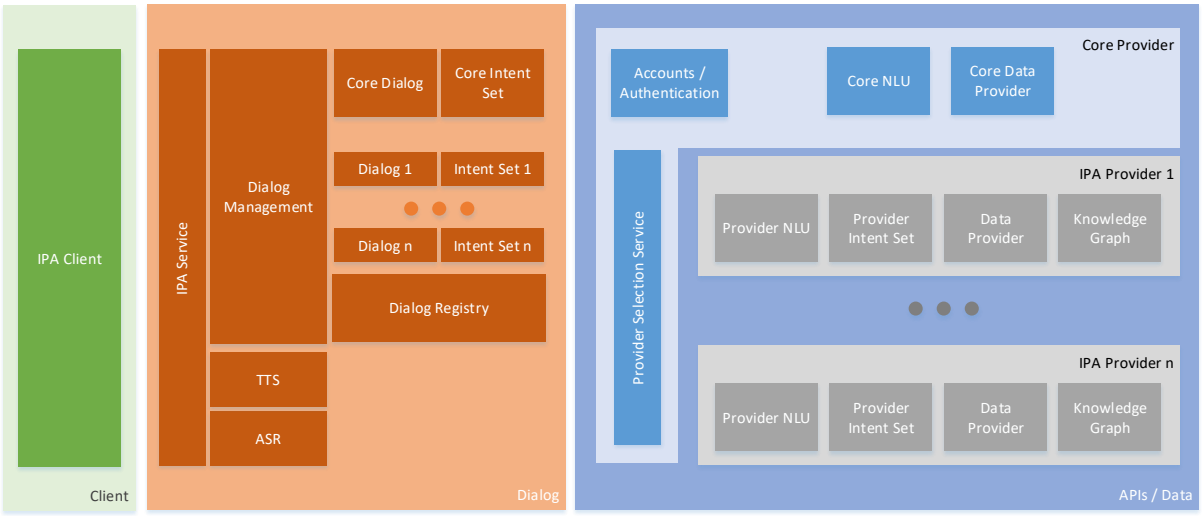
**Last Update:**
2019-09-30

**Status:**
Draft

**Revision History:**

| Ver. | Date | Author | Description of Change |
|---|---|---|---|
| 0.1 | 30/9/2019 | Dirk Schnelle-Walka | Initial version |

# ARCHITECTURE




# CLIENT LAYER

## IPA CLIENT

Client that enables the user to access the IPA via voice. Usually, IPA Clients make use of a microphone to capture the spoken input and a loud speaker to provide responses.

As an extension an IPA Client may also receive commands to be executed locally.

# DIALOG LAYER

## IPA SERVICE

General IPA Service API that mediates between the user and the overall IPA system. The service layer may be omitted in case the IPA Client communicates directly with the Dialog Management.

## DIALOG MANAGEMENT

Component that receives user input as spoken input, updates its internal state, decides upon subsequent steps to continue a dialog and provides output as synthesized or recorded utterances.

For this, it employs several Dialogs that are responsible to handle isolated tasks or intents. The overall set of available dialogs defines the behavior and capabilities of the interaction with the IPA. The Dialog Manager is also responsible for a good user experience across the available Dialogs.

The Dialog Manager determines the Dialog that is best suited to serve the current user input and re-establishes the interaction state for that Dialog. Therefore, it may use the Dialog Registry.

The Dialog Manager also manages the session with a user. Conceptually, multiple sessions can be active in parallel. Dialogs are governed by Sessions, e.g. to free resources of ASR and NLU engines when a session expires. Linguistic phenomena, like anaphoric references and ellipsis are expected to work within a Session. The selected IPA Provider or the Dialog Manager may have leading roles for this task.

The Dialog Manager also features an ASR to convert spoken utterances into text strings and a TTS to convert text strings into audio.

As an extension the Dialog Manager may also return commands to be executed by the IPA Client.

## ASR

The Automated Speech Recognizer (ASR) receives audio files of recorded utterances and generates a recognition hypothesis as text strings. Optionally, the ASR can also generate multiple recognition hypothesis along with a confidence score.

### TTS

The Text-to-Speech (TTS) component receives text strings which it converts into audio data.

### CORE DIALOG

The Core Dialog is able to handle basic functionality via Core Intents to enable interaction with the user at all. This includes among others

- Greetings
- Goodbye
- Exception handling in case a requested service is not available
- Exception handling in case a requested intent cannot be matched to a known Dialog
- Help

Conceptually, the Core Dialog is a special Dialog that is always available.

### CORE INTENT SETS

A Core Intent Set usually identifies tasks to be executed and define the capabilities of the Core Dialog. Conceptually, the Core Intents are always available.

### DIALOG X

The Dialog x are able to handle functionality that can be added to the capabilities of the Dialog Manager through their associated Intent Sets. The Dialogs must server different purpose in a sense that they are unique for a certain task. E.g., only a single flight reservation dialog may exist at a time.

Dialogs receive inputs as intents out of their supported Intent sets along with associated entities and return responses as text strings to be spoken.

Dialogs access the Provider Selection Service to fulfill their task. They maintain state and know which IPA Provider evaluated their request with the help of an identifier.

As an extension Dialogs may also return commands to be executed by the IPA Client.

### INTENT SET X

An Intent Set usually identifies tasks that can be executed within the associated Dialog.

### DIALOG REGISTRY

The Dialog registry manages all available Dialogs with their associated Intent Sets. Dialogs and their Intent Sets can be added or removed as needed.

## APIS/DATA LAYER

### PROVIDER SELECTION SERVICE

A service that provides access to all known IPA Providers. This service also maps the IPA Intent Sets to the Intent Sets in the Dialog layer.

The Provider Selection Service receives input as text strings and returns results as intents from all IPA Providers that are able to reply to the user input along with associated entities.

The Provider Selection Service is stateless and always returns the responses from the IPA Providers along with an identification of the issuing IPA Provider.

### ACCOUNTS/AUTHENTICATION

A registry that knows how to access the known IPA Providers, i.e. which are available and credentials to access them. Storing of credentials must meet security and trust considerations that are expected from such a personalized service. IPA Providers can be added as needed.

### CORE NLU

A component that is able to extract meaning as intents and associated entities from an utterance as text strings.

The Core NLU is able to handle basic functionality via Core Intents to enable interaction with the user at all.

### CORE DATA PROVIDER

A generic data provider to aid the Core NLU determining the intent.

### IPA PROVIDER X

A provider of an IPA service, like

- Google Now
- Alexa
- Microsoft Cortana
- SoundHound
- …

#### PROVIDER NLU

A component that is able to extract meaning as intents and associated entities from an utterance as text strings for IPA Provider X

#### PROVIDER INTENT SET

An intent set that might be returned by Provider NLU to handle the capabilities of Provider X.

#### DATA PROVIDER

A data provider to aid the Provider NLU in determining the intent. This component may also be used to track the IPA Provider's state

#### KNOWLEDGE GRAPH

A knowledge graph to reason about the detected input from the Provider NLU and Data Provider to come up with some more meaningful results.