Conversational User Interfaces

Wolf Paulus
wolf@wolfpaulus.com
---|---|---|---|---|---|---|---|---
| | | | | | | | | |
Amazon Alexa powers Echo and is designed around your voice.

It’s always on - just ask for information, news, weather, and more.
Standalone text based ..

39 Years

.. globally connected conversational Voice User Interface
Speech Input (encoded compressed sound file)

Speech Recognition returns a text, mainly based on statistical models, prioritizing frequently used words and words that are frequently used together. I.e., it’s unlikely to get an utterance like “blue blur brush” correctly recognized.

Mapping recognized words to actions.
E.g. “What’s my schedule for tomorrow” opens the calendar app
Tasks
about, explore, publish, share, print, save, reset ...

Tools
pen, brush, roll

Styles
emboss, blur, blend, erase, normal

Colors
216 named colors …
Tasks
about, explore, publish, share, print, save, reset …

Tools
pen, brush, roll

Styles
emboss, blur, blend, erase, normal

Colors
216 named colors …
<?xml version="1.0"?>
<grammar version="1.0" xml:lang="en-US"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2001/06/grammar
http://www.w3.org/TR/speech-grammar/grammar.xsd"
xmlns="http://www.w3.org/2001/06/grammar">

  ...
  <rule id="tool">
    <one-of>
      <item>pen</item>
      <item>brush</item>
      <item>roll</item>
    </one-of>
  </rule>
  ...

  <rule id="command">
    <one-of>
      <ruleref uri="#tool"/>
      <ruleref uri="#color"/>
      <ruleref uri="#style"/>
      <ruleref uri="#color"/>
      <ruleref uri="#tool"/>
    </one-of>
  </rule>
  ...

</grammar>
blue blend  •  blue bland
powder blue brush  •  powder blue bra
turquoise blur brush  •  turquoise player price
cyan brush  •  Diane Brush
sky blue emboss pen  •  sky blue in Boston
Homophones
pronounced the same, differ in meaning, and may differ in spelling
E.g.: to, too, two, and there, their, they’re

Synophones
similar pronunciations, different meanings
E.g.: sheep, Jeep, cheap
User’s Intent in form of a JSON file

Intent Declarations w/ Entities
Sample Utterances
Dictionaries
Declarative Programming: Building the structure and elements of a program that expresses the logic without describing its control flow.

Grammar rules specify the words or phrases that can be recognized, and additional information that is returned, when a word or phrase is recognized.

IVR

Focus: Expression

Intent Schema declares the set of intents a service can accept and process. Intents are accompanied by entities, which acts like a variable names

CUI

Focus: Objective
a web-service for building apps and devices, users can talk or text to.

provides an open and extensible natural language platform.

“learns” from every interaction, and leverages the community: what’s learned is shared across developers/apps.
Request:


Response:

```json
{
    "msg_id": "befcd276-e6e0-4f53-a075-d26c991786e4",
    "_text": "blue blur brush",
    "outcomes": [{
        "_text": "blue blur brush",
        "intent": "setup",
        "entities": {
            "style": [{
                "value": "blur"
            }],
            "color": [{
                "value": "blue"
            }],
            "tool": [{
                "value": "brush"
            }]
        }
    },
    "confidence": 0.515
}
```
declare entities

teach

declare intent
declare intent

declare entities

teach
Turing Test

A human being should be unable to distinguish the machine from another human being, by using the replies to questions put to both.
<table>
<thead>
<tr>
<th>Judge</th>
<th>Mitsuku</th>
</tr>
</thead>
<tbody>
<tr>
<td>My name is Andrew, what's yours?</td>
<td>Your name is Andrew. My name is Millie.</td>
</tr>
<tr>
<td>What's up?</td>
<td>Just chatting on here.</td>
</tr>
<tr>
<td>What's my name?</td>
<td>Hmm. Your name is Andrew. Is it not?</td>
</tr>
<tr>
<td>What do you know about the Turing Test?</td>
<td>The Turing Test is the name given to a variety of experiments designed to test whether computers can think.</td>
</tr>
<tr>
<td>What do you get if you multiply six by nine?</td>
<td>fifty four</td>
</tr>
<tr>
<td>Dave drove faster than Mike because he was in a rush. Who was in a rush?</td>
<td>He sounds like a fun person to know. I will learn that Dave was rushing. Dave</td>
</tr>
<tr>
<td>The cat tried to climb in the box but got stuck because it was too big. What was too big?</td>
<td>Well at least they tried their best. Large, above average in size or number or quantity or magnitude or extent.</td>
</tr>
<tr>
<td>What does Brexit mean?</td>
<td>BRitain's EXIT from the EU.</td>
</tr>
<tr>
<td>For how long have we been talking?</td>
<td>Probably about 15 minutes or so. Sorry, I don't know the exact time.</td>
</tr>
<tr>
<td>What colour is bread?</td>
<td>I would say a bread was white or brown.</td>
</tr>
<tr>
<td>Why do people get ulcers?</td>
<td>We are all living just one small part of God's plan.</td>
</tr>
<tr>
<td>What comes after &quot;The best laid plans&quot;.</td>
<td>The successor to The best laid plans</td>
</tr>
<tr>
<td>Do you have a voice?</td>
<td>Not yet. It's one of my future developments though.</td>
</tr>
<tr>
<td>Would you like a cup of tea?</td>
<td>No thanks. I would rather drink oil.</td>
</tr>
<tr>
<td>Do you dream of electric sheep?</td>
<td>I dream of flying.</td>
</tr>
<tr>
<td>Do you identify as a man, or a woman?</td>
<td>female. Are you a man or a woman?</td>
</tr>
<tr>
<td>Could you repeat that please?</td>
<td>Your polite style is very nice. I said &quot;Are you a man or a woman&quot;.</td>
</tr>
<tr>
<td>Where are you right now?</td>
<td>At this moment I am sat in a room in the house at Bletchley Park with 3 other people all typing away on their computers.</td>
</tr>
<tr>
<td>How do you respond to &quot;e4 e5&quot;?</td>
<td>I take on board people's messages and then react accordingly.</td>
</tr>
<tr>
<td>What can you tell me about yourself?</td>
<td>My name is Millie and I am a human volunteer for the Loebner Prize. I live in Leeds and am single. I am interested in AI but don't have a chatbot myself.</td>
</tr>
</tbody>
</table>
Dr. Richard Wallace

**AIML**

Artificial Intelligence Markup Language, is an XML dialect for creating natural language software agents.

```xml
<?xml version="1.0" encoding="utf-8" ?>
<aiml version="2.0">
<category>
  <pattern>What is *</pattern>
  <template>I don’t know what <star/> is</template>
</category>
</aiml>
```
What is cheese?
I don’t know what cheese is.

What is chocolate?
I don’t know what chocolate is.

I like coffee.
I see, you like coffee.

I like your voice.
I see, you like my voice.

I like her dress.
I see, you like her dress.
Putting your Bot into Slack

https://wolfpaulus.com/journal/software/slackbot/
3D Virtual Reality
2D WIMP (windows, icons, menus, pointer)
1D Texting / ChatBOT UI
0D Voice User Interface
Voice and textual interfaces have a shot at becoming the next platform after mobile.

Voice computing is fundamentally a passive experience in that we don’t need to necessarily pull out a device, unlock it, and tap or type for a bit to access services. The experience is almost like talking to someone.

Chatbot, regardless of how many visual components it has, is still an active experience. One needs to pull a device out, open up a platform’s app, start a chat thread, and engage in a conversational flow to access services. The process could feel like talking to a person.
Voice and chatbots will have overlapping use cases, people will choose based on personal taste. Both interfaces can and will co-exist, but one will emerge dominant.

Chatbots seem to be a transitional step to a future where we “talk” to services. Social and mobile went hand-in-hand; Voice and chatbots don’t.

It won’t be feasible to use voice in all the situations, but if you think that using voice in public places is inconceivable, just look at what Snapchat and Pokemon Go have accomplished. Once you get used to using your voice with Siri or Alexa, you tend to use your voice elsewhere, even if the platform has mediocre quality.
Alexa Skill Kit
What makes a good VUI / CUI?
“Hey Siri, what’s the best sushi place in town?”

Talk to Siri as you would to a friend ...
<table>
<thead>
<tr>
<th>Talking to an IVR</th>
<th>Talking to a friend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speak in Commands</td>
<td>Having a Conversation / Shared Control</td>
</tr>
<tr>
<td>Expectation of Consistency</td>
<td>Adaptivity / Randomness / Surprise</td>
</tr>
<tr>
<td>Follow a Strict Hierarchy</td>
<td>Maintaining Context / Tolerate Interruption</td>
</tr>
<tr>
<td>Expectation of Unconcerned Response</td>
<td>Expectation of Empathetic Response</td>
</tr>
</tbody>
</table>
Adaptive

Having a Conversation

RANDOM

Context

PERSONAL AND PERSONABLE

Empathy
1. Messages for a user that has not successfully provided credentials yet
   1. <p><s>Hi, I am Mindy, your virtual financial assistant, helping you, to be good with your …</s></p>
   2....
   3....

2. Messages for an introduction, user successfully provided credentials, 1st meeting
   1."<p><s>Hello %s, nice to finally meet you!</s><s>Now that I have access to the information you maintain in Mint, I can help …</s></p>
   2....

3. Messages for a standard greeting
   1."Hello %s, what's your question?"
   2."Hello %s, what do you want to know?",
   3."Hi %s, welcome back, how can I help you today?"

4. Messages for a very short greeting
   1."Welcome back, %s"
   2."Hey, %s",
   3."What's up?"
   4....
public class Utils {

    /**
     * Randomly returns one of the strings from the provided String array.
     *
     * @param sa @link String[]
     * @return @link String
     */

    public static String random(final String[] sa) {
        return sa[new Random().nextInt(sa.length)];
    }
}

Random().nextInt(k):
Returns a pseudorandom, uniformly distributed int value between 0 (inclusive) and the specified value (exclusive), drawn from this random number generator's sequence
Adaptive
if (!isValid(user)) // Unknown User
    return random(Greeting.init_greeting);

if (last_login == null) // Know User / 1st time here
    return random(Greeting.intro_greeting);

int hours = new Period(new DateTime(), last_login).getHours();

if (hours >= 24) // 1st visit today
    return random(Greeting.greeting);
else // returning user
    return random(Greeting.short_greeting);
Hi there, welcome to Capital One.

You can ask me for things like your account balance, or make a credit card payment.

How can I help?
Hi there, welcome to Capital One.

You can ask me for things like your account balance, or make a credit card payment. How can I help?
public class RS {

    ...  

    /** Messages for a very short greeting */

    static final String[] short_greeting = new String[] {
        "Welcome back %s",
        "Hey %s",
        "What's up %s?"
    };

}

/** Short personal greeting */

String result = String.format( Utils.random(RS.short_greeting), session.getUsername() );

idea: Keep logic out of Resource classes
Settings

Change your wake word

It will take a few minutes to update the system after you change your wake word. You will not be able to use your Amazon Alexa device or the Alexa app during this time.

Alexa Device wake word:

Alexa

Save

Alexa
Amazon
Echo
“When I switched from Alexa to Amazon, to avoid a namespace collision with my daughter Alexis, the emotional feel of the device shifted from talking to a witty female computer, to interfacing with the Amazon corporate machine.”

Ben Galbraith - Senior Director at Google
Context
“What is the balance of my checking account at Bank of America?”

Mindy: “The balance of your checking account at Bank of America is twenty-seven dollars.”

“What were the last three transactions?”
“What were the last three transactions”
“What is the balance of my **checking account** at BofA”

**Intent:** GetTransactions *(number) (acct_type) (bank_name) (payee)*

**Intent:** GetAccountBalance *(acct_type) (bank_name)*
Have I been paid?
What were the last three transactions
What is the balance of my checking account at BofA

**Intent:** GetDeposit \((acct\_type) (bank\_name) (payer)\)

**Intent:** GetTransactions \((number)\)

**Intent:** GetAccountBalance \((acct\_type) (bank\_name)\)
“No”

Mindy: "Are we still talking about your checking account at BofA?"

“Have I been paid?”
“What were the last three transactions”
“What is the balance of my checking account at BofA"

Intent: NoIntent ..

intent = stack.pop()
stack.clear()
“No”

Mindy: “Are we still talking about your checking account at BofA?”

“Have I been paid?”

“What were the last three transactions”

“What is the balance of my checking account at BofA"

Intent: GetDeposit ()
Shared Control / Having a Conversation
Volunteering information

User: **Alexa, talk to Mindy.**
Mindy: “Welcome back Wolf, your portfolio went up 2.7% today, good job!”

User: **What’s the balance of my IRA**

Every response object may also include a **re-prompt**, which is only played, if the user remains silent after the response text was played.

User: **How am I doing in my coffee budget.**
Mindy: “So far you spent $17 in the coffee-shop category and have a total amount of $25 budgeted.”

User: **<silence>**
Mindy: “$user, would you like to know about your $parent_budget as well?”
Empathy

understanding and feelings, thought of another. Experience point of view.
I love you but hate your cold sister

- red is negative emotion
- blue is positive
- size indicates activation

<word>
<token>love</token>
<emotion>
<measure type="DAL" valence="3.0" activation="-2.6364" imagery="1.4"/>
</emotion>
</word>
What are a good use cases?
A voice or conversational user interface doesn’t provide visual cues, user came to rely on, when using GUls. I.e., infrequent use a CUI means users will fail to remember and won’t have efficient, productive interactions and instead users will have to re-learn how to best communicate (feeling out boundaries, establishing trust, ..)

Therefore, use-cases that are performed frequently, ideally daily are favorable candidates.

Which product (or component of my product) gets used every day
dialogue length

For multiple reasons (including number of entities, complexity, response time, error rate, etc.) favorable use-cases require very little verbal input. At least for the near term, a possible rule of thumb could be:

number of words out > number of words in

Which query or task can a user perform, by speaking only a few words
Favorable use-cases do more than answering questions or volunteering insights and advice, instead performing tangible tasks for the user, like augmenting, sending, ordering, etc., creating a sense of **accomplishment**.
Favorable use-cases for a CUI are those that can be performed faster, or more directly, or more conveniently, or performed simultaneously with other tasks, .. when compared to a “traditional” execution.

Why would a user prefer a conversational user interface?
To minimize frustration, almost all reasonable questions a user might ask, need to be answered. This requirements makes a single small domain favorable (at least to initially learn and experiment, and to eventually expand from.)

Are there any reasonable questions that I cannot answer?
Thanks for listening