 INTERAÇÃO POR VOZ NO MUNDO DE JOGO

João Pedro Marques Pereira Ferreira

*Orientador*

Alberto Manuel Brandão Simões

Trabalho de Projeto apresentado(a) ao

Instituto Politécnico do Cávado e do Ave

para obtenção do Grau de Mestre em Mestrado Engenharia em Desenvolvimento de Jogos Digitais:

*Este trabalho não inclui as críticas e sugestões feitas pelo Júri.*

Fevereiro, 2018

 INTERAÇÃO POR VOZ NO MUNDO DE JOGO

João Pedro Marques Pereira Ferreira

*Orientador*

Alberto Manuel Brandão Simões

Trabalho de Projeto apresentado(a) ao

Instituto Politécnico do Cávado e do Ave

para obtenção do Grau de Mestre em Mestrado Engenharia em Desenvolvimento de Jogos Digitais:

*Este trabalho não inclui as críticas e sugestões feitas pelo Júri.*

Fevereiro, 2018

RESUMO

As formas de interação com o mundo dos videojogos poucas alterações sofreram desde a sua massificação. Se analisarmos os diferentes elementos que compõem um videojogo, é possível concluir que áudio, grafismo, inteligência artificial, têm tido uma evolução gradual e substancial. No entanto o campo da interação física, humano jogo virtual, pouca evolução houve. Comando, teclado, rato são os periféricos que mais abundam no mundo dos jogos virtuais e são a norma desde a sua massificação. Com o aumento computacional nas plataformas de videojogos com a soluções de nuvem cada vez mais disponíveis, surgiu a ideia de experimentar e analisar como seria adicionar a voz como forma de interação. Cada vez mais rodeados de dispositivos de interação por voz (Amazon Alexa, Google, Siri, Cortana) é de estranhar que este tipo de interação não tenha sido ainda implementado nos videojogos.

Se olharmos para a evolução gráfica um dos desafios foi aproximar o que vemos no mundo do videojogo com o que vemos no mundo real. Seguindo essa linha de raciocínio e sabendo que a forma mais natural de interagirmos no mundo real é usando a nossa voz.

Neste relatório, vamos analisar, que tentativas já foram feitas neste tema, como construir um protótipo que nos permita interagir com o mundo de jogo usando a voz, desvantagens e limitações, vantagens e oportunidades.

**Palavras-chave:** Interação, voz, videojogo, assistente pessoal, interface, experiência de jogo.

Agradecimentos

Glossário

GUI – Graphic User Interface

AI – Artificial inteligence

Plugins –

Developers -

Cloud -

Websites

Índice

Ídice de Figuras

# Introdução

Nesta secção vamos analisar que conclusões e aprendizagens podemos tirar de soluções atuais semelhantes, a motivação para a criação do projeto, qual a necessidade de incluir voz no mundo dos videojogos e quais as suas vantagens. Vamos apresentar onde vai estar o nosso foco no desenvolvimento, e que partes vão ser deixadas de parte. Por fim apresentamos qual a estrutura deste relatório quais os dados e método que vamos usar para retirar conclusões do projeto.

## Observações da tecnologia

Quando analisamos soluções que envolvem o tema, interação por voz, encontramos vários produtos que usam a voz como a principal forma de interação. Os assistentes pessoais são os que têm mais destaque e que estão atualmente mais massificados. [ASSISTENTES PESSOAIS] A Siri da Apple no ano de [ANALISE DE DATAS E DE ESPECIFICAÇÕES] foi o primeiro a dar o tiro de partida para este tipo de soluções. Seguiram-se Google Amazon Microsoft com soluções semelhantes, e desde então a facilidade de utilização, as possibilidades que oferecem têm subido a cada ano que passa.

Mais especificamente na área dos videojogos, é difícil encontrar nos dias que correm um jogo cuja interação seja feita parcialmente ou totalmente feita por voz [REFERENCIA JOGOS COM VOZ ALISON 2017]. Temos a exceção dos jogos de Karaoke, no entanto quando analisado em detalhe percebemos que a interação usa a voz, no entanto apenas algumas características da mesma, não é analisado o que é dito nem o conteúdo do que está a ser cantado, mas sim, o tom da [DETALHES DA ANÁLISE DA VOZ NOS JOGOS KARAOKE] voz do jogador.

Podemos encontrar pequenas interações em alguns jogos, mas as mesmas, ou não têm um destaque grande no jogo que possa ser considerado uma verdadeira interação, ou são usadas apenas algumas características da voz como o tom e o volume.

## Propósito e a solução para um não problema

Seguindo a análise anterior, percebemos que os assistentes pessoais são bons na ajuda de tarefas mais rotineiras, sem grande complexidade e normalmente repetitivas. Todos estes adjetivos podem ser atribuídos a grande parte dos videojogos. Consultar uma lista, abrir um mapa, marcar um destino, escolher uma tarefa, equipar um item, são tarefas que normalmente conseguimos executar usando uma GUI, diegética, não diegética ou espacial. Dado o contexto deste relatório, a questão natural que surge é: “Porque não usar a voz?”.

Consideramos que a falta da voz na interação num videojogo um não problema, porque como referido acima, a verdade é que sempre conseguimos resolver as tarefas que estes mundos digitais propõem de uma forma normal dado o historial que os videojogos têm no nosso quotidiano bem como outras interações digitais a que somos expostos fora dos jogos, como por exemplo websites e aplicações de telemóvel.

No entanto entendemos que integrar voz como forma complementar de interação no mundo de jogo, pode trazer interações bastante interessantes ao mesmo tempo que acelera a execução de tarefas rotineiras e nos mantêm no mundo virtual sem grande disrupção.

## Espectro do projeto

O desafio que lançamos na criação deste projeto é: Como podemos integrar voz num mundo virtual que nos permita interagir com o mesmo. Executar tarefas que sem a voz requereriam uma GUI. Retirar todas as GUI que não sejam informativas de forma a quebrar o mínimo possível a experiência e envolvência no mundo virtual.

Posto de outra forma, o nosso objetivo inicial é criar um assistente virtual que nos ajude guie e entretenha na nossa viagem pelo mundo do jogo. Consultar missões, pedir direções, consultar inventário, pedir contexto histórico do mundo em que nos encontramos, são alguns exemplos tangíveis que queremos alcançar com a realização deste projeto.

Numa fase posterior após criação do projeto, passamos à análise de como a interação criada, ajudou, prejudicou. alterou a forma como o jogador experiência o jogo. Realizar testes com utilizadores e medir a sua performance e qual a análise que fazem a este tipo de interação.

## Estrutura do relatório

Este relatório está divido em quatro secções. A primeira sendo a introdução que termina neste capítulo, que nos contextualiza e prepara para os temas a seguir abordados.

A segunda baseia-se num estudo prévio das ferramentas, interações e soluções semelhantes existentes à data da realização do relatório. Nesta secção vamos estudar que videojogos já tentaram este tipo de interação e qual o resultado. Que outras soluções fora do mundo do jogo existem com que frequência são usadas e quais as suas maiores vantagens e desvantagens. Nesta secção vamos analisar também em que tecnologias se baseiam a interação por voz, desde NLP passando pela IA, computação na nuvem e terminando nos ChatBots.

A terceira abordamos e apresentamos o plano de desenvolvimento do projeto. As ferramentas escolhidas, a arquitetura tecnológica em que se baseia a solução, que desafios práticos vamos implementar para que sejam resolvidos usando a voz, como vamos criar um ambiente que nos possibilite testar toda a solução.

A quarta e última secção é onde falamos das conclusões. Dos testes que conduzimos com utilizadores, que problemas identificamos na utilização da solução, quais as oportunidades que aparecem neste tipo de solução. Terminamos com um apanhado daquilo que foi proposto e o que foi conseguido e quais as dificuldades de implementação desta solução num videojogo.

# Estado de arte

## IA

## Interface utilizador

## Voz no mundo de hoje

## Reconhecimento Voz

### Big players

### Major problems

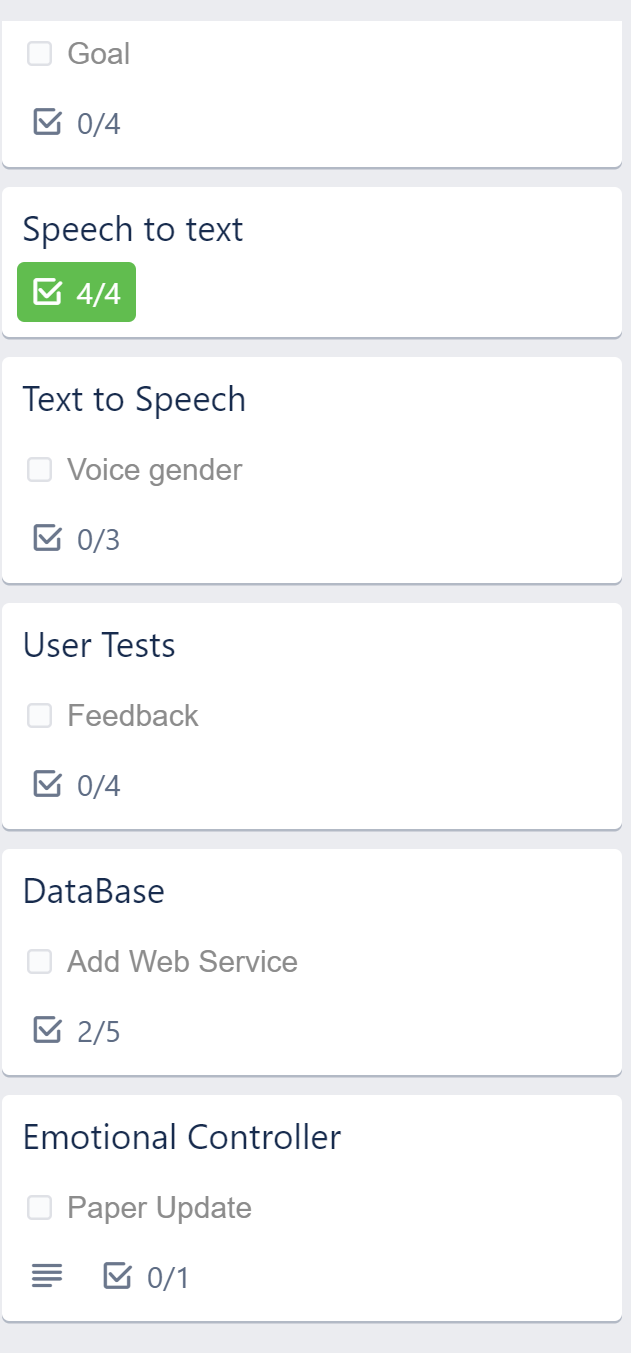
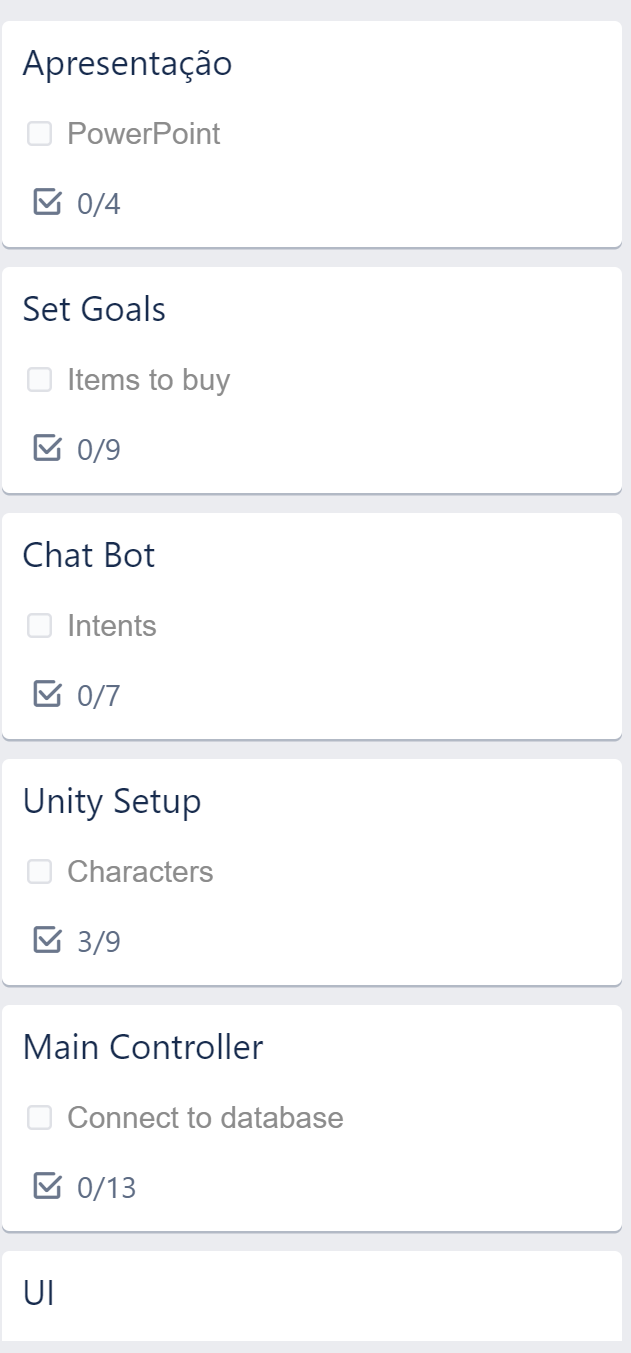
## Chat-bot

# Plano de desenvolvimento

Neste capítulo detalhamos o desenvolvimento do projeto. A distribuição das tarefas, as ferramentas utilizadas, arquitetura da solução, o mundo de jogo e no final um capítulo com algumas observações sobre o resultado final do desenvolvimento e quais seriam os possíveis próximos passos.

## Planificação das tarefas





### Diálogos

Neste tópico apresentamos o objetivo que o jogador quer ver realizado e expomos uma possível conversa entre o jogador e o assistente virtual, uma versão mais simples e outra mais complexa com algumas condicionantes.

**Diálogo A – Versão simples**

Objetivo: O jogador precisa de comprar um item

**J**(Jogador) – “IPCA, quero comprar um item”

**AV** (Assistente Virtual) - “Ok, aqui tens os itens disponíveis” (Aparece UI)

**J** – “Comprar item X”

**AV**– “Ok, vou adicionar ao teu inventário” (UI Selecionada / Item adicionado)

**Diálogo A – Versão complexa**

Objetivo: O jogador precisa de comprar um item

Condicionante: O jogador não tem dinheiro para comprar o item.

**J** – “Quero comprar o item mais caro que existe.”

**AV** – “Não podes, não tens dinheiro suficiente”

**J** – “O que posso fazer para ganhar dinheiro?”

**AV** – “Podes explorar o mundo para ganhar dinheiro, queres que marque algum objetivo no mapa?”

**J** – “Sim, o mais perto”

**AV** – “Ok” (Aparece localizador no mapa)

**Diálogo B – Versão simples**

Objetivo: O jogador quer saber quais as missões que tem para fazer.

**J** – “IPCA, O que é que posso fazer neste jogo?”

**AV** – “Tens vários objetivos, podes começar pelo X, ou explorar o mundo à procura de Y, Z”

**J** – “Ok, direciona-me para o objetivo X”

**AV** – “Vou marcar essa localização”

**Diálogo B – Versão simples**

Objetivo: O jogador quer ir para um determinado objetivo.

Condicionante: O jogador só pode ir para um determinado objetivo, se acertar a uma pergunta do assistente.

**J** – “IPCA, marca no mapa o objetivo X”

**AV** – “Tens que me dizer primeiro: Eu, IPCA, tenho mais ou menos que 15 anos?”

**J** – “Menos”

**AV** – “Errado, só podes ir ao objetivo Y”

OU

**J** – “Mais”

**AV** – “Certo, vou marcar o objetivo escondido” (Aparece localizador no mapa)

## Ferramentas

As ferramentas utilizadas na realização deste projeto são as seguintes

Unity

“Unity is a cross-platform game engine developed by Unity Technologies, first announced and released in June 2005” [*https://en.wikipedia.org/wiki/Unity\_(game\_engine)]* Unity é uma plataforma de desenvolvimento para soluções em tempo real, mais conhecido como um motor de jogo, Unity permite realizar projetos de arquitetura, automação, filmes, entre outros.

A escolha recaiu sobre esta plataforma inicialmente por ser um motor de jogo em que já tínhamos alguma experiência de desenvolvimento sendo reduzida a curva de aprendizagem. Outra das razões é o facto da flexibilidade de integração de outras ferramentas, Unity disponibiliza uma loja virtual, com inúmeros plugins que aceleram e facilitam o desenvolvimento. A última razão é o tamanho da comunidade à volta da plataforma são vários os fóruns, canais de Youtube, livros que partilham conhecimento sobre a plataforma.

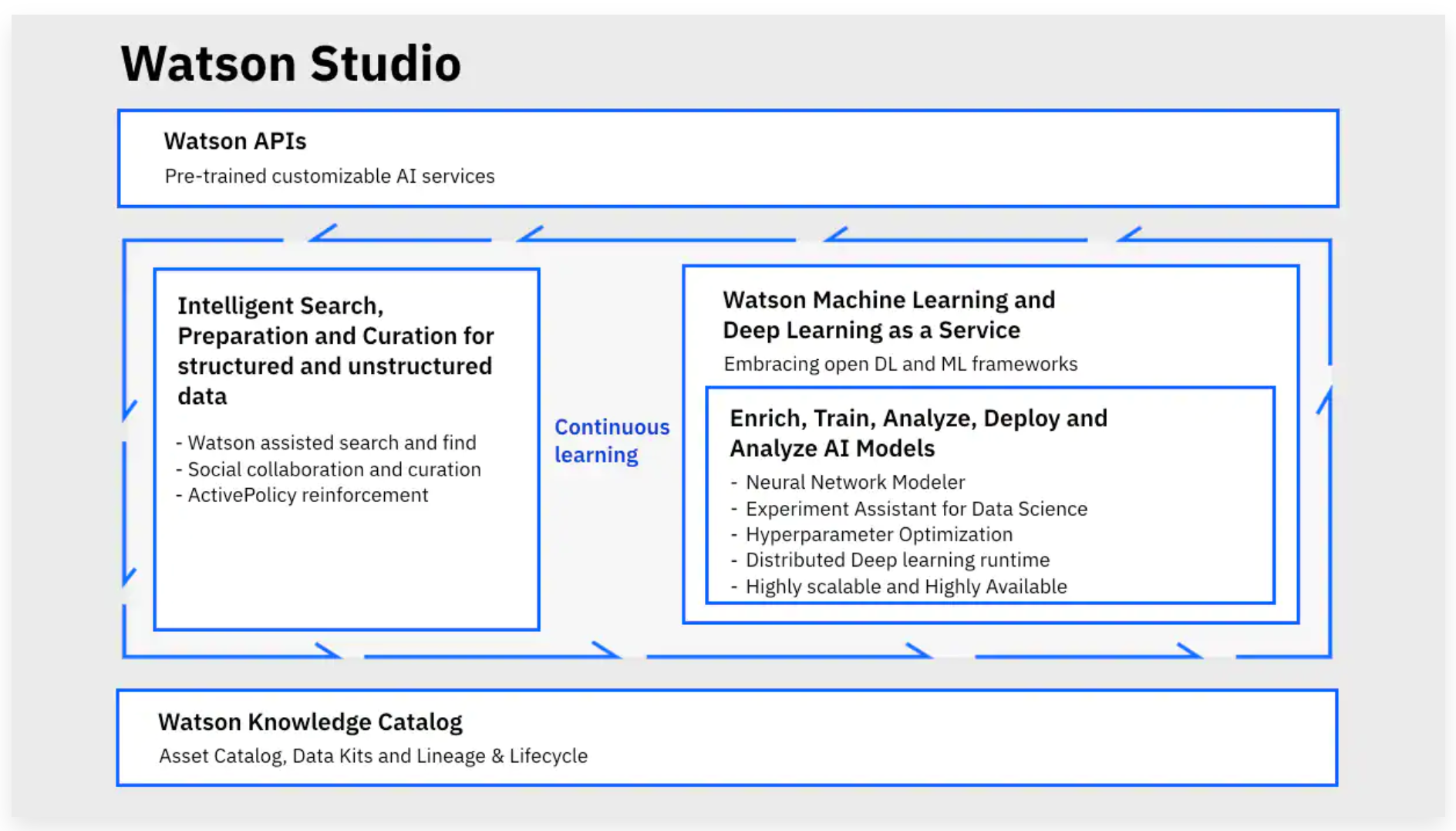
IBM WATSON

“IBM Watson Studio is an integrated environment designed to make it easy to develop, train, manage models, and deploy AI-powered applications and is a SaaS solution delivered on the IBM Cloud.” [*https://www.ibm.com/cloud/blog/announcements/watson-studio-announcement* ]

Watson é uma Plataforma com várias soluções de serviços cognitivos. Serviços que com o uso de AI e com o serviço de Cloud, colocam na mão de developers ferramentas potentes que de outra forma apenas estariam ao alcance de grandes empresas. “Watson is IBM’s suite of enterprise-ready AI services, applications, and tooling.”

A escolha desta plataforma não foi tão fácil como a anterior. Há mais alternativas, e são precisas vários serviços para a realização do projeto. Para o nosso projeto eram necessárias soluções para: Conversão de Voz para Texto; Conversão de Texto para Voz; ChatBot. Com o IBM Watson, conseguimos numa só plataforma agregar esses três serviços.

Outro dado importante prende-se com o facto de a plataforma disponibilizar para os três serviços um plano gratuito que apesar de algumas limitações, serve o propósito do nosso projeto.



## Arquitetura

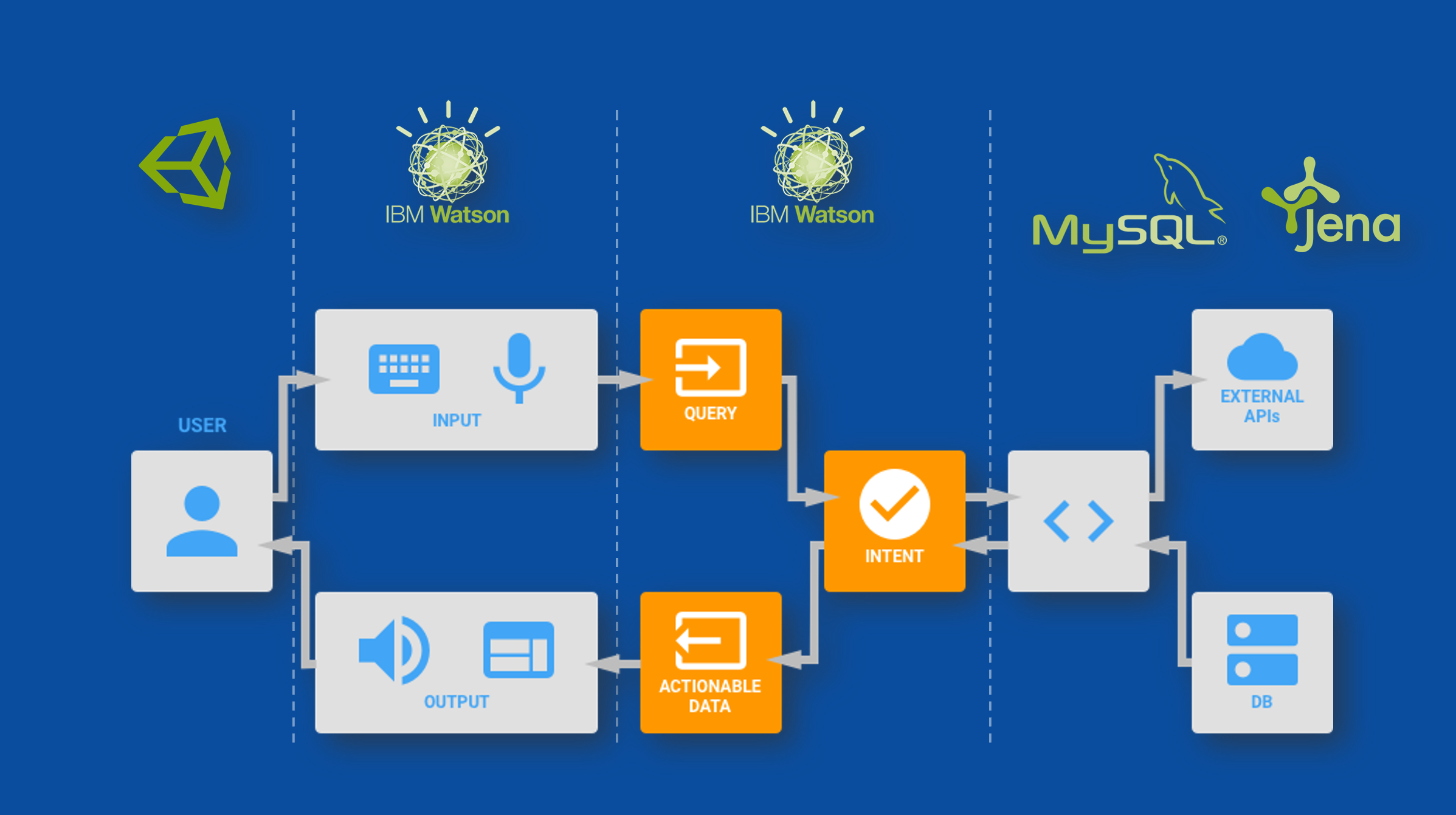
Na imagem abaixo podemos ter uma melhor perspetiva da arquitetura implementada no nosso projeto. Podemos organizar e segmentar o projeto em 3 fases.

A primeira é o input/output, resumindo é a forma como o utilizador dá e recebe informação ao mundo de jogo.

A segunda é a manipulação da informação. Nesta fase acontece o tratamento da informação que provém e que vai para o utilizador. Neste passo é convertida voz para texto e texto para voz, e aqui podemos trabalhar e brincar com entoação, carisma e género da voz que é apresentada ao utilizador.

A terceira é onde se trabalha a lógica e direção e soluções do diálogo que o assistente virtual tem com o jogador.

Com menos destaque na imagem vemos uma fase que podíamos ter incluído, a utilização de um sistema de base de dados. Dado que estamos a desenvolver num ambiente controlado e sabendo em antemão os dados que vamos tratar, decidimos não incluir no projeto esta fase.



## Mundo do jogo

## Trabalho futuro

# Análise de Resultados

## Solução final

## Tempos execução

## Facilidade utilização

## Problemas

## Questionario

## Utilizadores

## Report

## Observação

## Revisão da solução

### Projeto final vs ideia original

### Feedback de utilizadores

# Conclusão

# Bibliografia

SNIPETS DE TESES / SITES / ETC

Word Play: A History of Voice Interaction in Digital Games

Fraser Allison1, Marcus Carter2 and Martin Gibbs1

50 Years of progress in speech and speaker recognition research. ECTI

Transactions on Computer and Information Technology

# Tópicos RETIRADOS DE OUTRAS TESES

ALISON 2017

- Tema e razão de ser do projeto

Aumento de poder computacional

Voz cada vez mais presente

Cloud

PCS com mais hardware/poder computacional

- Forma de investigação

A base do que existe (Sites/teses)

O que está a ser feito/melhorado

O projeto com o questionário

Jogos

- In Verbis Vitrus

- Yoake no Mariko

- Odama

Consola

* FAMICOM Nintendo um comando com microfone

Interações

- Berrar para matar inimigos (legendo f Zelda

- Soprar para mover uma roleta (Kamen Rider Club)

- Negociar com um vendedor (Kid Icarus)

Jogos e interações

- Nntendo DS – BrainAge - Color naming

- Nintendo DS - Nintndogs – Comands to teach a virtual dog

- TomClancy’s End War – Permitia ao utilizador dar comandos a tropas sob a forma de – “quem/o quê/onde”

West vs East

* Nos USA e Europa a ideia era que o controlo fosse uma alternativa a um comando manual
* No Japão a ideia era que houvesse mais diálogo com personagens virtuais

PARTS

Not all the Kinect games use spoken commands in this way, however. Some focus

on creating the illusion that the characters in the game world are simply able to hear

the sound of the player. The stealth horror game Alien: Isolation includes an optional

“noise detection” mode, in which any moderately loud noise detected by the Kinect

microphone alerts the game’s fearsome antagonist (the titular alien) to the location

of the player–character. Similarly, the stealth action game Splinter Cell: Blacklist

allows the player to call out to in-game enemies, prompting them to investigate.3

This feature was well received by players and reviewers alike for supporting a sense

of connection with the game world (Carter, Allison, Downs, & Gibbs, 2015, p. 267).

Voice commands that allowed

players to role-play momentarily as their character were praised, whereas voice

commands that called attention to the player’s position outside the game world were

criticized as feeling “unnatural” and “embarrassing”—an effect that Carter, Allison,

Downs, and Gibbs (2015) termed “identity dissonance” (p. 268).

EXEMPLOS

Skyrim Elder Scrol V

<https://www.youtube.com/watch?v=dpkEoP92QQY>

These have appeared for both PCs (Bot Colony, There

Came an Echo, In Verbis Virtus, Plan Be) and mobile phones (The Howler, Ah!

Bird, Mayday! Deep Space), t

QUOTES

This is the rarely explored power of voice: speaking to a character with your own actual

voice connects you to them, forges a bond. (Shimomura, 2015)

argument made by Nass and Brave (2005) that all speech interaction

automatically evokes an imagined social relationship with the technology