UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

LENNON SALES FURTADO

Uma Proposta de Interação por Voz em Aplicações de Visualização da Informação

Belém - PA
2016

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

LENNON SALES FURTADO

Uma Proposta de Interação por Voz em Aplicações de Visualização da Informação

Dissertação apresentada a Banca examinadora do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Pará, como requisito para a obtenção do título de Mestre em Ciência da Computação na área de concentração em Gestão da Informação.

Orientador: Prof. Dr. Bianchi Serique Meiguins.

Coorientador: Prof. Dr. Nelson Cruz Sampaio Neto.

Belém - PA
2016

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Furtado, Lennon Sales, 1990-
Uma proposta de interação por voz em aplicações de visualização da informação / Lennon Sales Furtado. - 2016.

Orientador: Bianchi Serique Meiguins;
Coorientador: Nelson Cruz Sampaio Neto.
Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Ciência da Computação, Belém, 2016.

1. Computação gráfica. 2. Serviços de informação. 3. Tecnologia de informação. I. Título.

CDD 23. ed. 006.6

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

LENNON SALES FURTADO

UMA PROPOSTA DE INTERAÇÃO POR VOZ EM APLICAÇÕES DE VISUALIZAÇÃO DA INFORMAÇÃO

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Pará como requisito para obtenção do título de Mestre em Ciência da Computação, defendida e aprovada em 25/02/2016, pela banca examinadora constituída pelos seguintes membros:

Prof. Dr. Bianchi Serique Meiguins
Orientador – PPGCC/UFPA

Prof. Dr. Nelson Cruz Sampaio Neto
Co-Orientador – PPGCC/UFPA

Profa. Dra. Marcelle Pereira Mota
Membro Externo – FACOMP/UFPA

Prof. Dr. Elói Luiz Favero
Membro Interno – PPGCC/UFPA

Visto:
Prof. Dr. Jefferson Magalhães de Morais
Coordenador do PPGCC/UFPA

Prof. Dr. Jefferson Magalhães de Morais
Coordenador do PPGCC
Mat.: SIAPE: 2378314

AGRADECIMENTOS

Agradeço a Deus por ter me concedido persistência, força, direção e equilíbrio para enfrentar e superar os desafios de um novo tempo.

Ao meu orientador, Professor Doutor Bianchi Serique Meiguins, que não se limitou a um excelente despertador de potencial (capacidade de desenvolvimento humano), muito embora somente isso constituísse motivos suficientes de agradecimentos, todavia, indo além, desempenhou o papel de quem não somente motiva ou encoraja, mas que verdadeiramente acredita, sem desistir.

A meu coorientador Professor Doutor Nelson Cruz Sampaio Neto, primeiro pela suas excelentes aulas o qual tive o privelégio de assistir. E segundo pela paciência em me orientar.

A meus pais (João A. C. Furtado e Marlise M. C. S. Furtado), que são meus incentivadores e patrocinadores dos meus sonhos.

Aos meus irmãos (Hugo S. Furtado e Tássia S. Furtado) e melhores amigos pelo auxílio prestrado.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

“Cada um de nós é um herói em potencial. Mas, muitos só descobrem isso debaixo dos escombros de um terremoto" (Antônio Francisco).

Sumário

LISTA DE ABREVIATURAS E SIGLAS

ASR	*Automatic Speech Recognition* (Reconhecimento automático de voz).
TTS	*Text-to-Speech* (Texto para voz)
VI	Visualização de informação
LAPS	Laboratório de processamento de sinais
UFPA	Universidade Federal do Pará
CV	Comandos de Voz
3D	Três Dimensões
InfoVis	*Information Visualization*
API	*Application Programming Interface*
LN	Linguagem Natural
NUI	*Natural User Interface*

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

Muitos estudos apontam a importância da interação na área de Visualização de Informação (*InfoVis*) para o sucesso de uma boa visualização de dados. Este interesse é potencializado pelo crescimento e complexidade dos dados armazenados eletronicamente nas várias áreas do conhecimento. As pesquisas de interação em *InfoVis* têm fomentado o uso de interfaces não convencionais, além dos tradicionais teclado e mouse, tais como: comandos de voz, rastreamento de olhos, comandos por gestos, interfaces cérebro-computador, entre outras. O presente trabalho teve o objetivo de apresentar aspectos de concepção, desenvolvimento e avaliação de uma interface por comandos de voz para interação em aplicações de *InfoVis* 3D. Sendo que como cenário de uso foi utilizada a técnica de *InfoVis* de Dispersão de Dados em 3D. Ademais, para o reconhecimento de voz é empregado a API Coruja, que se utiliza do motor de reconhecimento Julius, para suporte ao Português-Brasileiro. Por fim, serão apresentados testes de usabilidade para a avaliação da interface e interação proposta. Os testes utilizam a abordagem de tarefas de usuários para avaliar as sub-tarefas de *InfoVis* na aplicação, tais como: configuração, filtro, seleção, entre outras. Para realizar a avaliação da carga de trabalho subjetiva do usuário foi utilizada a metodologia *NASA Task Load-Index*, a qual identifica a carga de trabalho total do usuário nas diferentes tarefas realizadas. Tais testes apontaram eficácia da interface apresentada, tendo uma diferença de tempo de 24% em relação ao mouse e 26% de carga de trabalho a mais do que no mouse.

Palavras-Chave: Visualização de Informação, IVOrpheus, Interação por voz, Dispersão de Pontos em Três Dimensões.

ABSTRACT

Several studies point out the importance of interaction in Information Visualization (InfoVis) field to the success of good data visualization. And this interest is compounded by the growth and complexity of data electronically stored in various areas of knowledge. The researches in interaction applied in InfoVis have encouraged the use of non-conventional interfaces, besides the traditional keyboard and mouse, such as voice commands, eye tracking, gesture controls, brain-computer interfaces, among others. This work aims to present aspects of design, development and evaluation of an interface for voice commands to interact in InfoVis 3D applications. The InfoVis technique used as usage scenario is the 3D Scatterplot. In addition to speech recognition is used Coruja Software, which uses the recognition engine Julius, to support Brazilian Portuguese. Finally, usability tests are presented for evaluation of the interface and interaction proposed. The tests used the approach of user tasks to evaluate InfoVis sub-tasks in the application such as: configuration, filter selection, among others. And to perform subjective workload assessments was used NASA Task Load-Index methodology, which identifies the user's total workload in the different tasks performed. These tests showed the interface efficacy, having a time difference relative to the mouse of 24%, and a workload 26% above the mouse's workload.

Keywords: Information Visualization, IVOrpheus, Voice Interaction, Scatterplot 3D.

CAPÍTULO 1- INTRODUÇÃO

De acordo com Card, Mackinlay, & Shneiderman (1999), a Visualização de Informação (*InfoVis*) é o uso de um ambiente computacional interativo que possibilita representação visual de dados abstratos para amplificar a cognição do usuário sobre um conjunto de dados e seus relacionamentos. A grande quantidade de dados eletrônicos que é armazenada diariamente pelos vários sistemas computacionais tem evidenciado a área de *InfoVis* como uma das áreas que pode auxiliar nessa problemática de análise de grande quantidade de dados.

A massiva quantidade de dados e a dimensionalidade das bases destes têm apresentado desafios, tais como: necessidade de mais espaço visual nos dispositivos, a complexidade dos dados, novas técnicas de *InfoVis*, novas formas de interação, entre outras. Os ambientes 3D se apresentam como alternativa aos ambientes 2D por apresentarem uma dimensão a mais para representação visual dos dados.

Contudo, uma questão constantemente relacionada aos ambientes 3D é a dificuldades de interação em tais ambientes. Uma primeira hipótese para o entendimento desta dificuldade apresentada pelo usuário, está em usar um meio de interação 2D - como mouse e teclado - para interagir com o ambiente 3D. Segundo Jankowski (2013) outra hipótese está na qualidade da interface de interação que não orienta ou guia o usuário adequadamente na realização de suas tarefas.

Este trabalho tem o objetivo de apresentar aspectos de concepção, desenvolvimento e avaliação de uma interface por comandos de voz para interação em aplicações de *InfoVis* 3D. Sendo que como cenário de uso foi utilizada a técnica de *InfoVis* de Dispersão de Dados em 3D. E para o reconhecimento de voz é utilizado a API Coruja, que se utiliza do motor de reconhecimento Julius, para suporte ao Português-Brasileiro.

Como diretrizes básicas para a concepção da aplicação, tem-se a adaptabilidade da interface à base de dados de forma automática, a geração dinâmica das gramáticas utilizadas pelo sistema de reconhecimento de voz e construções de visualização de dados pelos usuários. Por conseguinte, serão apresentados testes de usabilidade para a avaliação da interface proposta.

1.1 Motivação

Ambientes tridimensionais apresentam um desafio para interação, pois os meios de entrada padrão, no caso, mouse e teclado, são basicamente meios de interação bidimensionais. Com isso, técnicas de visualização de informação que exploram os aspectos das três dimensões, carecem de modos de navegação e configuração eficientes.

O estudo ora apresentado propõe um meio de interação por comandos de voz baseado em interface natural, que permite ao usuário uma interação intuitiva com a visualização 3D. Também permite aumentar a audiência de pessoas que possam interagir com uma visualização tridimensional.

Sendo que o sistema aborda a mesma preocupação apontada por Vannevar Bush (1945), que visa otimizar o processo de interação-humano-computador, deixando a "conversa" entre os mesmos mais "amigável", ou seja, com o foco na interação intuitiva, haja vista que quanto mais natural for a interação, maior a gama de pessoas alcançadas. Assim, atendendo, tanto a leigos, pesquisadores e estudantes, permitindo a aplicação deste método em vários cenários, como: escolas, lugares públicos, reuniões, apresentações e ambientes de pesquisa.

1.2 Justificativa e Contribuição à Área

Segundo a IBM (2015), o mercado precisa de uma população de 4.4 milhões de "*Data Scientists*". O que enfatiza o problema da massiva e crescente quantidade de dados e da dificuldade de analisar e visualizar toda sua magnitude.

Tendo o foco em tal problemática, a presente pesquisa propõe a metodologia de interação que explora aspectos da comunicação natural, fazendo uso de comandos de voz para navegar e tornar as tarefas dos usuários em visualização de informação mais intuitivas e, principalmente, com a intenção de que esta abordagem permita ao usuário interagir de maneira eficiente ou eficaz com a visualização de informação proposta.

Para tal objetivo é proposta uma interface que contempla a interação via comandos de voz em um ambiente tridimensional de visualização de informação, utilizando o motor Coruja para o reconhecimento de voz. Assim, sendo uma contribuição deste trabalho, esta ferramenta, IVOrpheus, que tem suporte a interação por voz em português brasileiro, pois, durante o levantamento bibliográfico não foi

encontrada ferramenta de visualização de informação que dê suporte a esta língua, que faça uso de software livre para o reconhecimento de voz e utilize a técnica de dispersão de pontos em 3D/2D.

Neste sentido, tendo como as principais contribuições os resultados obtidos nos testes com os usuários e as dificuldades encontradas no desenvolvimento do projeto, tendo em vista de que estes servirão como base para sustentar trabalhos futuros na área de interface natural aplicado a visualização de informação. Dessa forma, trabalhos que visarão uma abordagem que possibilite uma otimização na utilização da voz, como meio de interação principal ou secundário em ambientes de visualização de informação.

1.3 Objetivos

Objetivo geral deste trabalho é de apresentar aspectos de concepção, desenvolvimento e avaliação de uma interface por comandos de voz para interagir com aplicações de *InfoVis* 3D. E como meio de validação da interface proposta, são realizadas duas avaliações, uma qualitativa, através da aplicação do questionário que visa identificar o nível de dificuldade subjetiva do usuário sobre a execução das tarefas e sub-tarefas, e outra quantitativa através da medição do tempo do usuário na execução das tarefas. Para assim, poder acompanhar a eficiência ou a eficácia da utilização da interface de voz em uma visualização de informação em 3D.

Os objetivos específicos deste trabalho são:

- Analisar o estado da arte na utilização de interfaces por comandos de voz em ambientes de visualização de informação;
- Identificar qual abordagem na utilização de interação por voz em ambientes de visualização de informação em 3D, que permite ao usuário maior eficiência na interação com a mesma;
- Identificar as limitações no reconhecimento de voz e indicar possíveis soluções;
- Identificar os comandos por voz que melhor se aplicam para a visualização de informação proposta e organiza-los em níveis segundo seu escopo;
- Apresentar proposta de comandos de interação por voz que satisfaçam as tarefas e sub-tarefas de visualização de informação;
- Apresentar uma proposta de interação por voz que obedeça o mantra da visualização de informação proposto por Shneiderman (1996).
- Avaliar a interação orientada ao meio de comunicação natural.

1.4 Estrutura da Dissertação

Neste capítulo introdutório foram apresentados os problemas motivadores, justificativas, contribuições para à área e os objetivos desta dissertação de mestrado.

O capítulo 2 apresenta a fundamentação teórica para melhor compreensão dos assuntos abordados por toda a dissertação.

O capítulo 3 apresenta a revisão bibliográfica, para análise dos resultados dos trabalhos examinados, verificando suas abordagens enquanto a utilização da voz como ferramentas (motores de reconhecimento/ sintetizadores de voz), níveis de interação e entre outras características.

O capítulo 4 apresenta a aplicação IVOrpheus, que é uma ferramenta de visualização por dispersão de pontos em três dimensões com entrada por comandos de voz e mouse. Também, apresenta-se os aspectos de implementação e arquitetura da aplicação.

O capítulo 5 apresenta a descrição dos testes realizados com os usuários. Sendo utilizado como medida qualitativa o NASA–TLX e o questionário para identificar o nível de dificuldade subjetiva durante a execução das tarefas e sub-tarefas. E como medida quantitativa o tempo dos usuários.

O capítulo 6 apresenta os desafios encontrados, trabalhos futuros, as considerações finais e publicações aceitas.

CAPÍTULO 2 – FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão mostrados conceitos das áreas de interação homem-computador e de visualização de informação, que serão importantes para o entendimento da dissertação.

2.1 Visão Geral da Interação por Voz

O reconhecimento de voz tem o objetivo de reduzir o gap na interação-humano-computador aproximando-se da interação-humano-humano, a fim de criar uma interface de interação por voz é necessário conhecer o público alvo, as tarefas que desejam ser realizadas e quais os cenários se aplicaram esta interface (Karat, Vergo, Nahamoo, 2003).

A importância de conhecer o público alvo está em conhecer certas características que devem ser atentadas no desenvolvimento de uma interface por voz, como, por exemplo, se o público alvo fala nativamente a língua abordada. Ou acerca da idade do público alvo, pois as pessoas de pouca idade tendem a ter uma voz mais aguda ou, quanto a pessoas de mais idade, que por qualquer questão médica tenha dificuldade na fala. Ou quanto ao nível educacional dos indivíduos, tendo em vista que estudos mostram que se obtém um melhor desempenho que tem no mínimo uma escolaridade fundamental (Te`eni, Carey e Zhang, 2005).

Existem basicamente quatro tipos de tarefas que podem ser realizadas por voz. Sendo elas, composição, transcrição, transação e colaboração.

Composição: tem como objetivo criar um documento de texto utilizando a voz, a exemplo, escrever um e-mail por voz. Transcrição: tem como objetivo converter uma conversa humana em um documento de texto. Transação: tem como objetivo realizar uma transação através de comandos de voz, a exemplo, uma busca na internet. Colaboração: tem como objetivo reconhecer de forma síncrona (realizado simultaneamente por duas ou mais pessoas) ou de forma assíncrona (realizado em tempos diferentes por duas ou mais pessoas) os diferentes comandos de voz para interagir com um sistema.

É altamente relevante conhecer o ambiente que o sistema de reconhecimento vai

ser aplicado. Seja este um lugar com pouco ruído, como um escritório silencioso ou para aplicação em ambientes com altos níveis de ruídos, como para um espaço militar.

Tendo como público alvo pessoas que falam nativamente o idioma português brasileiro e com um nível de escolaridade fundamental adiante, e fazendo o uso da tarefa de transação em ambientes com pouco ruído, a ferramenta proposta neste trabalho tem sua maior característica possibilitar a interação "*hands-free*" em um ambiente de visualização de informação em três dimensões, alcançado uma gama maior de pessoas.

2.2 Níveis de Interação por Voz em Ambientes de Infovis

Os trabalhos abaixo tem por objetivo de apresentar a aplicação de interfaces naturais, utilizando a tecnologia de reconhecimento automático de voz (ASR) e de texto para fala (TTS), demonstrando três diferentes níveis de aplicação de tais meios de entrada/saída em visualização de informação. Os artigos estão baseados nos três níveis respectivamente.

No trabalho, ***"****CineCubes: Aiding data workers gain insights from OLAP queries****"***, houve a utilização da tecnologia de sintetização de voz, no caso, TTS, para apresentar a resposta desejada, convertendo a resposta gerada em texto para áudio. Sendo que o usuário entrava com uma query na base de dados, e posteriormente era realizada uma visualização narrativa (*slide show*), onde o TTS apresentava os dados ao usuário, assim ampliando sua experiência (Nível 0).

No artigo, *"A Study of Manual Gesture-Based Selection for the PEMMI Multimodal Transport Management Interfac****e"***, o qual se utiliza de uma visualização com interação por gestos e voz voltada para controladores de trafego. Na ferramenta, o emprego da voz é voltado para manipulação da visualização e para a seleção de dados (Nível 1). A exemplo de comandos de voz, temos "*Show map of specified suburb", "Select single landmark at specified point" e "Select group of landmarks within specified area".*

Enquanto, que no trabalho ***"****Articulate: A Semi-automated Model for Translating Natural Language Queries into Meaningful Visualizations****",*** os autores propõem um processo para traduzir a entrada verbal do usuário em um comando de LN (Linguagem Natural) ao invés de um comando baseado em gramática, usando um tradutor de LN

com algoritmos de aprendizado de máquina. O sistema *Articulate* consegue reconhecer uma busca em LN do usuário e aplicar regras pré-determinadas de acordo com a categoria léxica de cada palavra. Depois da busca ser devidamente processada, o resultado é mostrado em um grafo contendo os dados desejados em uma técnica de visualização 2D, decidida pelo sistema com base nos tipos de dados e de busca (Nível 2).

Na Figura 2.1, são sintetizados os diferentes níveis de utilização de voz em visualização de informação.

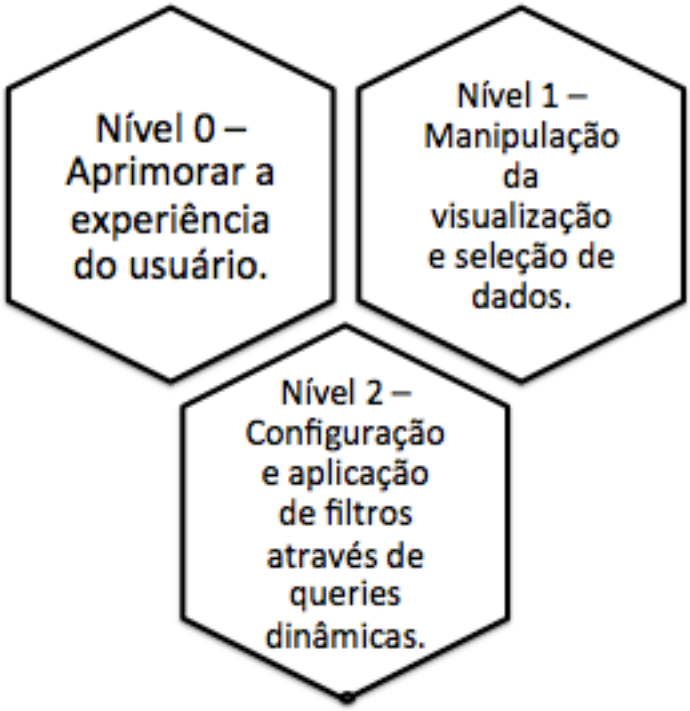

Figura 2.1. Níveis de interação por voz.

2.3 Visão Geral da Visualização da Informação

Segundo Chen (2002), a visualização da informação estuda a representação de dados não baseados em aspectos físicos, como, mapas ou corpo humano, ou seja, estuda a representação visual e interativa de dados abstratos visando ampliar a compreensão dos mesmos. A visualização de informação é uma interface entre o humano e a informação que por definição não é exclusiva de meios computacionais. Apesar de que, nas últimas décadas o suporte computacional tem sido um fator importante para o

avanço deste campo.

A visualização de informação é uma alternativa para a informação por dados verbais ou textuais. Tendo a capacidade de transmitir mais informação que a informação textual (Ward, Grinstein, Keim, 2010). Isto acontece, pois o humano intrinsicamente tem a capacidade de encontrar padrões e distingui-los. Uma vez que, o meio de entrada de informação de maior largura de banda no ser humano é a visão.

Por este e outros fatores, houve um crescimento no uso de visualização tanto da informação quanto cientifica nas últimas décadas. Podendo ser notado que a visualização está presente em uma miríade de lugares no cotidiano das pessoas. Como, mapa de trens e metrôs, gráfico da bolsa de valores, placas de trânsitos indicando curvas, gráficos com a previsão do tempo, gráficos nos jornais representando o número de pessoas contaminadas com alguma enfermidade e assim por diante.

Entre as técnicas mais conhecidas de visualização de informação está a de dispersão de pontos (Scatterplot) que é uma das mais antigas e utilizadas (Ward, Grinstein, Keim, 2010). Em função disto, foi escolhida neste trabalho a técnica de dispersão de pontos em três dimensões, visto que o público alvo do mesmo são pessoas que falam nativamente o idioma português brasileiro e com ensino fundamental como nível de escolaridade básico. Isso implica que tais pessoas não necessariamente são da área de visualização de informação, assim a utilização de uma técnica amplamente conhecida se faz necessário.

2.4 Pipeline de Visualização da Informação

O Pipeline é uma sequência de estágios que podem ser estudados independentemente em termos de algoritmos, estrutura de dados e sistema de coordenadas (Ward, Grinstein, Keim, 2010). Na Figura 2.2 é apresentado o pipeline de visualização de informação, sendo os estágios, informação, tabela de dados, estrutura visual e a visualização. A seguir será descrito cada estágio:

- Modelagem dos dados: os dados a serem visualizados devem ser estruturados visando facilitar o processo de gerar a visualização. O nome, o tipo, e o valor de cada atributo devem estar disponíveis assegurando um rápido acesso e permitir modificações;

- Seleção de dados: envolve identificar o subconjunto de dados que será visualizado;
- Dados mapeados visualmente: o núcleo do pipeline de visualização da informação está em mapear os dados para sua posição, cor, forma e tamanho;
- Manipulação da Visão: o usuário deve especificar vários atributos da visualização que são relativamente independentes da base, como alterar o mapa de cores entre outros atributos;
- A geração da visualização: a projeção dos objetos da visualização variam de acordo com mapeamento utilizado. Além de apresentar os dados, algumas visualizações também incluem informação suplementar para facilitar a interpretação, tais como, eixos, estados e anotações.

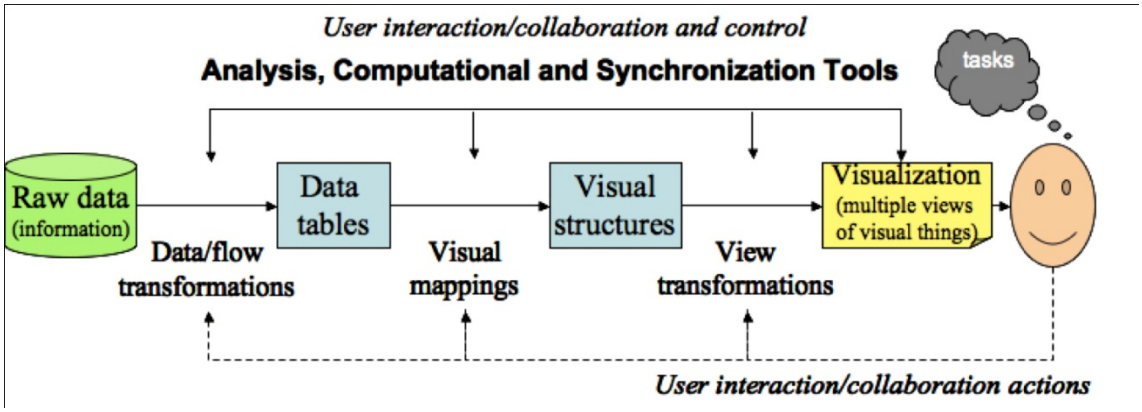

Figura 2.2. Um dos pipelines de visualização de informação. (Ward, Grinstein, Keim, 2010)

CAPÍTULO 3- REVISÃO BIBLIOGRÁFICA

Este capítulo tem o objetivo de revisar trabalhos e analisar seus resultados, metodologias e ferramentas, adotando um meio sistemático para tal, com o fim de se ter um panorama da aplicação de interação por voz nas áreas de visualização de informação e visualização científica.

3.1 Revisão Sistemática

Para a revisão foi utilizado uma forma simplificada e adaptada do guia de Kitchenham (2004), onde a Figura 3.1 apresenta a visão geral do protocolo de revisão.

Figura 3.1 - Protocolo da revisão sistemática adaptado (Kitchenham, 2004).

3.2 Atividade: Planejamento

Com o objetivo de estreitar a relação entre homem e máquina, utiliza-se meios de entradas baseados no meio de comunicação natural, a exemplo, interação por gestos, por rastreio dos olhos, entre outros. Porém, entre todos os meios de interações naturais, nenhum é tão desenvolvido quanto o meio de comunicação por voz. Devido isto, é notório o crescente uso de sistemas reconhecedores de voz em uma miríade de áreas. Desta forma, reunir e revisar trabalhos na área servem de arcabouço para a prova da relevância e ratificam a abordagem deste trabalho. Para isso, foi empregado um protocolo que descreve as atividades realizadas na revisão sistemática.

3.2.1 Desenvolvimento do Protocolo

As atividades executadas no desenvolvimento do protocolo de revisão sistemática estão descritas a seguir:

1. **Questões de Pesquisa**: Estas questões deverão guiar a leitura dos artigos e deverão ser respondidas ao final da revisão.

Questão principal: Qual o estado da arte da aplicação de interação por voz em uma ferramenta de visualização de informação?

Questão Secundária 1 (QS1): Quais as técnicas de visualização de informação e ferramentas de reconhecimento de voz são utilizadas nos experimentos?

Questão Secundária 2 (QS2): Quais as intenções primárias do uso da interação por voz nestes trabalhos (Comandos utilizados)?

Questão Secundária 3 (QS3): Como são medidos quantitativamente e qualitativamente os testes de cada trabalho?

Questão Secundária 4 (QS4): Como está sendo aplicada a entrada por voz em ambientes de visualização em três dimensões?

2. **Mecanismo de Pesquisa** – As palavras usadas na busca de trabalhos foram: "*Information Visualization*","*Voice Interaction*, ASR(*Automatic Speech Recognition*)". Para melhorar os resultados da busca, a lógica de combinação entre as palavras foi: (("*Voice Interaction*" *OR* " *Automatic Speech Recognition*") *AND* ("*Information visualization*"). A busca procurou qualquer parte do trabalho que contivesse as palavras usando as ferramentas dos indexadores de trabalhos científicos on-line: Science Direct[1], ACM Digital Library[2] e IEEExplore[3].

3. **Critérios de Seleção:** Além das palavras utilizadas na busca, foram aplicados filtros no tipo de conteúdo e ano da publicação para otimizar a busca. Assim, encontrando trabalhos mais relevantes, novos e alinhados com o tema. Sendo que

[1] Science Direct - http://www.sciencedirect.com/

[2] ACM Digital Library - http://dl.acm.org/

[3] IEEEXplore - http://ieeexplore.ieee.org

somente trabalhos publicados em periódicos ou conferências dentro dos anos de 2000 até 2015, foram incluídos para análise. Posterior a esta busca foi realizada outra, porém sem a aplicação de filtros para evitar exclusões equivocadas.

4. **Critérios de Extração dos Dados:** Após a busca pelos indexadores, uma verificação rápida de cada trabalho é efetuada com a leitura do título e resumo de cada um. A partir daí, os trabalhos restantes são catalogados com a extração dos dados principais, a saber, título, ano da publicação, país onde foi desenvolvida a pesquisa, nível de interação, técnica de visualização, dimensionalidade da visualização, método de reconhecimento (Gramáticas ou diálogo), ferramentas de reconhecimento de voz, avaliação, comandos de voz, componentes da interface multimodal, utilização de ASR ou TTS, idioma reconhecido e quantidade de usuários nos testes.

3.3 Condução

A condução da revisão teve duas coletas, sendo feita a primeira coleta de informações em junho de 2015 e uma coleta para atualização dos trabalhos em julho de 2015.

3.3.1 Selecionar Estudos Primários

Foram encontrados no *Science Direct* 40 artigos, sendo que após a leitura do título e resumo de cada trabalho foram selecionados 19 artigos para uma pesquisa mais aprofundada. O mesmo sucedeu com a ACM e a IEEE que apresentaram respectivamente 18 e 780 artigos que continham em seus metadados as palavras "*Information Visualization*" + "*Automatic Speech Recognition*" ou "*Voice Interaction*". Sendo que após a leitura do título e resumo foram retirados para analise completa 15 artigos da ACM e 42 artigos da IEEE, com intuito de responder a questão primária e as questões secundárias.

Em seguida foi aplicada uma análise mais detalhada dos artigos, onde foram lidos o título, o resumo, a introdução, o núcleo do trabalho, e as considerações finais dos autores. Sendo que após a análise, foram retirados os falsos positivos assim restando apenas 3 artigos da *Science Direct*, 4 Artigos da ACM e 10 artigos da IEEE. Totalizando 17 artigos que estavam alinhados com o tema desta pesquisa.

3.3.2 Extração dos Dados

A extração dos dados se deu pela leitura dos trabalhos selecionados e preenchimento de um formulário para posterior apresentação. Os dados que foram preenchidos neste formulário foram: ano, país onde foi desenvolvida a pesquisa, nível de interação, técnica de visualização, dimensionalidade da visualização, método de reconhecimento (Gramáticas ou dialogo), ferramentas de reconhecimento de voz, avaliação, comandos de voz, componentes da interface multimodal, utilização de ASR ou TTS, idioma reconhecido, quantidade de usuários nos testes.

3.3.3 Síntese dos Dados

Essa etapa consiste em organizar os dados extraídos para apresentação dos gráficos que serviram como panorama geral e base para futuras análises.

3.4 Resultados

Essa seção responde as questões principais e secundárias e discute os resultados identificando avanços, lacunas e desafios em geral.

3.4.1 Panorama dos Estudos Analisados

A leitura dos 17 artigos selecionados possibilitou uma análise panorâmica dos estudos que abordam a utilização da voz como meio de interação em ambientes de visualização de informação. O gráfico da Figura 3.2 apresenta o número de publicações por ano, entre 2000 até 2015. Apenas os anos de 2006, 2009, 2011 e 2012, não tiveram publicações.

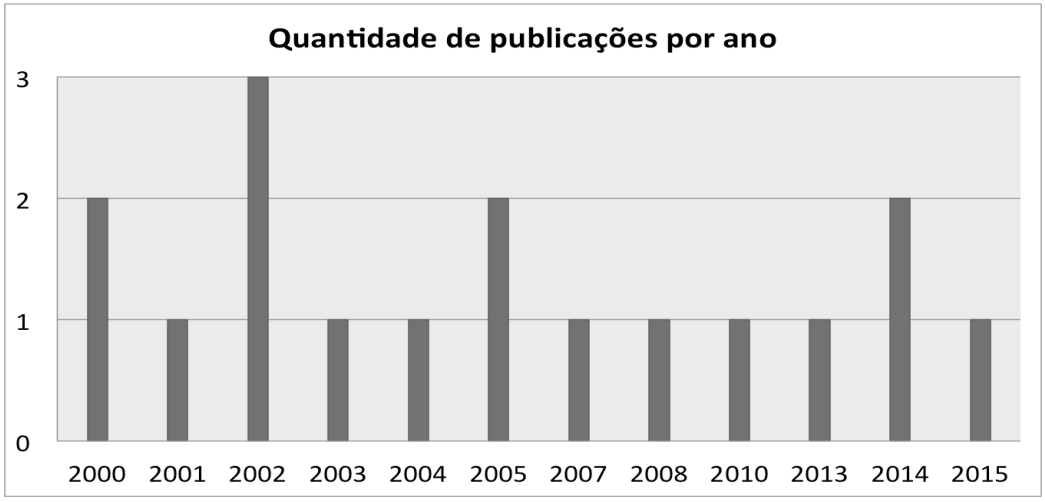

Figura 3.2 – Quantidade de publicações por ano.

Observando a Figura 3.2, é possível acompanhar que no início da década de 2000 houve um grande interesse em explorar um meio de interação baseado em linguagens naturais (Wegman, 2000), (Bohus et al, 2002), (Chen et al, 2005) e nos demais outros anos. Houve trabalhos que apresentaram não somente entrada por voz, mas miríade de outros meios de entrada, como gestos mais voz (Lubos et al, 2014). Porém, mesmo assim, é notória a escassez de trabalhos relacionados a área de interação por voz em ambientes de visualização.

Na Figura 3.3, é possível conferir os países onde foram realizadas as pesquisas. Onde os EUA com 11 publicações se destaca por possuí o maior número de trabalhos publicados na área.

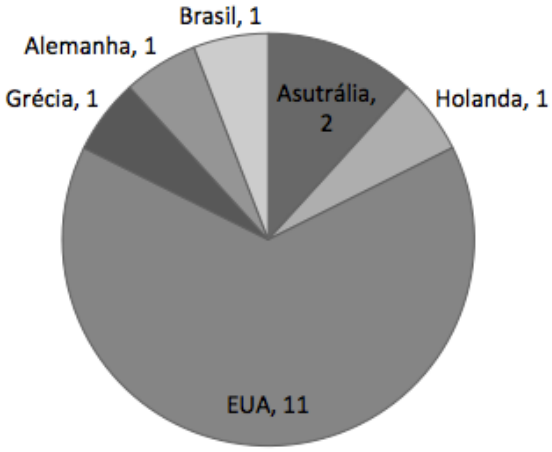

Figura 3.3 - Publicações por país relacionadas a Interação por voz em ambiente de visualização de informação entre 2000 e 2015

Durante a leitura dos artigos foi possível notar que a grande maioria das ferramentas descritas nas publicações utilizavam-se de sintetizadores de voz ou reconhecedores de voz da língua inglesa, como pode ser acompanhado na Figura 3.4. Tendo apenas dois trabalhos que reconhecessem outro idioma, sendo Português Brasileiro (Krammes et al. 2014) e o espanhol (Saverio et al. 2004).

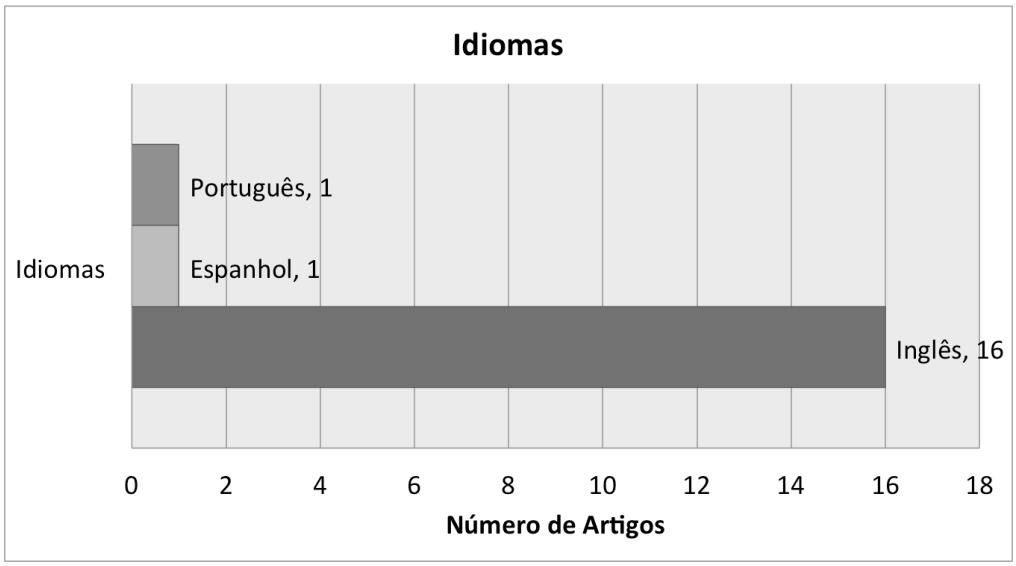

Figura 3.4 - Publicações por idiomas reconhecidos.

A maioria dos trabalhos revisados reconhecia apenas uma língua, a exceção de (Saverio et al. 2004), que reconhecia inglês e espanhol conjuntamente. Além do reconhecimento de voz, várias publicações abordaram a utilização da voz como saída (Heeren et al. 2007) e (Gkesoulis et al. 2015), enquanto outros utilizaram tanto o reconhecimento de voz (ASR) quanto a voz sintetizada (TTS), tendo, respectivamente, as funcionalidades de entrada e saída (Bohus et al. 2002). A porcentagem do uso destas tecnologias nos trabalhos analisados pode ser acompanhada na Figura 3.5.

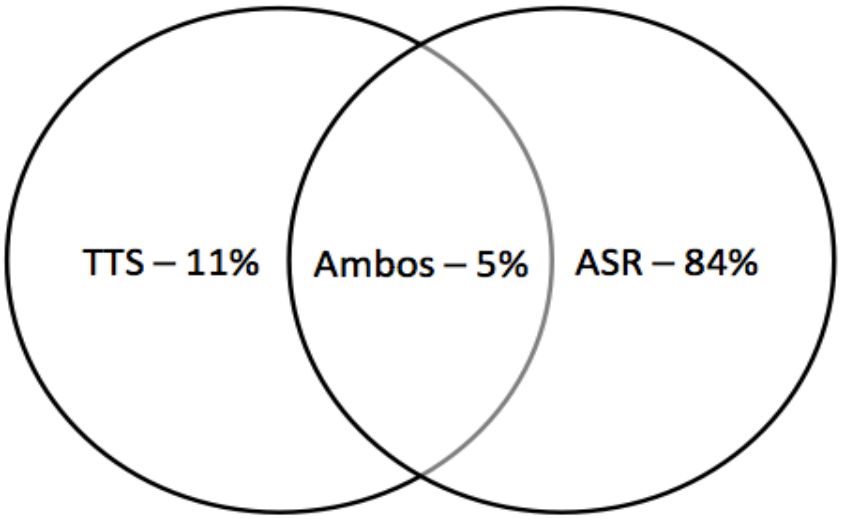

Figura 3.5 – Utilização de ASR e TTS nas publicações analisadas.

Na Figura 3.5, pode ser visto que a utilização de voz como saída tem uma

proporção inferior a utilização da mesma como entrada. Como aponta (Gkesoulis et al. 2015), a utilização de saída por voz sintetizada (TTS) pode incomodar alguns usuários, sendo que neste mesmo artigo alguns usuários durante os testes de usabilidade optaram por remover esta funcionalidade.

Como mostra a Figura 3.5, o reconhecimento de voz (ASR), foi amplamente utilizado nas publicações, sendo que este possui duas técnicas majoritárias para o efetivo reconhecimento das palavras ditas pelo o usuário, sendo a técnica de utilização de Gramáticas a mais comumente empregada e a segunda técnica a de reconhecimento de Diálogo com uma parcela menor de utilização. As Gramáticas tem maior eficiência no reconhecimento de voz, devido limitar a quantidade de palavras reconhecidas. Pois, apenas as palavras presentes nas gramáticas serão “escutadas” pelo motor de reconhecimento de voz.

Em contrapartida, o meio de reconhecimento por Ditado (Diálogo) é mais próximo do modo como o ser humano interage por linguagem natural. Pois, o reconhecedor faz uso de um extenso conjunto de palavras, que as vezes podem passar da casa dos milhares, e de modelos probabilísticos para prever as próximas palavras a serem ditas. Todo este aparato que suporta o reconhecimento de Ditado (Diálogo), possui um alto consumo dos recursos computacionais e por possuir uma quantidade maior de palavras, aumenta a probabilidade do reconhecedor gerar uma hipótese errada.

Por isso, é possível observar na Figura 3.6 que sua utilização de Ditado (Diálogo) é inferior a utilização por gramáticas, tendo apenas uma participação de 20% nos trabalhos analisados.

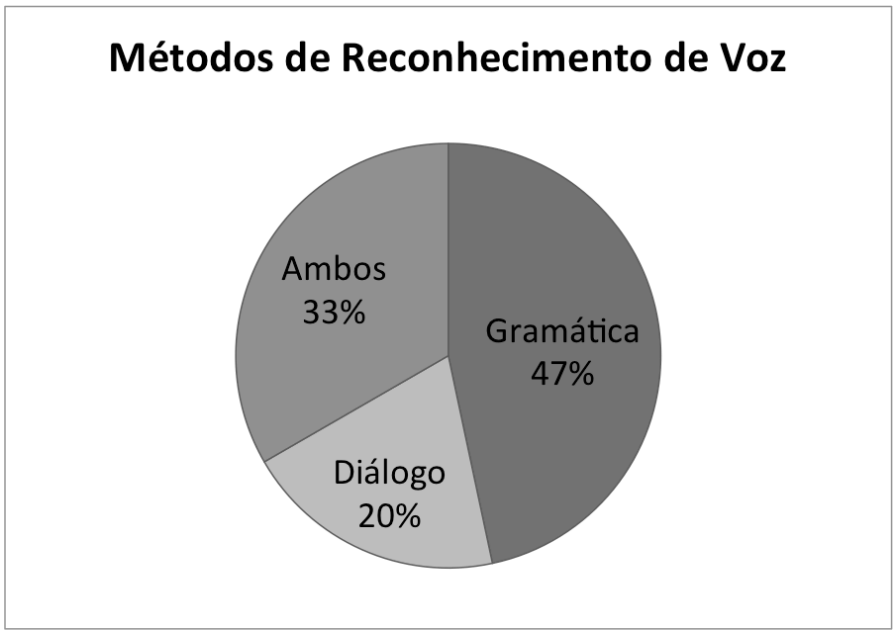

Figura 3.6 – Métodos de Reconhecimento de Voz.

A combinação dos meios de reconhecimento de voz pode ser acompanhada no trabalho (Myers et al. 2002), onde o autor utiliza o meio por gramáticas para interação com a aplicação e o meio por Diálogo para reconhecer e gerar uma visualização do que os usuários conversão enquanto não estão interagindo com o sistema. E para tal funcionalidade, o autor faz uso de dois motores de reconhecimento, sendo o XCalibur (Finke et al. 1999) e Janos (Finke et al. 1997).

Como o sistema é voltado para fins militares, é interessante fazer uso desta abordagem, pois, as conversas feitas durante a interação com o sistema servem como anotações do que os oficiais estavam pensando naquela determinada situação. É importante frisar que o custo computacional é bem elevado seguindo essa abordagem, por isso, no trabalho referenciado é utilizado vários servidores para tal.

Foram escassos os trabalhos que utilizaram apenas a voz como meio de entrada, pois a maioria fez uso de uma abordagem multimodal, ou seja, possui mais de um meio de entrada. Apesar de todos os trabalhos possuírem os meios de entrada padrão (teclado e mouse), eles serão desconsiderados com o fim de ater o foco nos meios de entrada não convencionais.

Sabendo disso, a voz estava geralmente acompanhada de entrada por gestos (Corradini et al. 2002) com a justificativa de que esta é a forma que mais se aproxima

do meio de comunicação entre humanos. Não obstante as publicações analisadas relataram uma rica combinação de entrada por voz com outros meios, como, o *touch* (Krammer et al. 2014), todos estes meios podem ser observados na Figura 3.7.

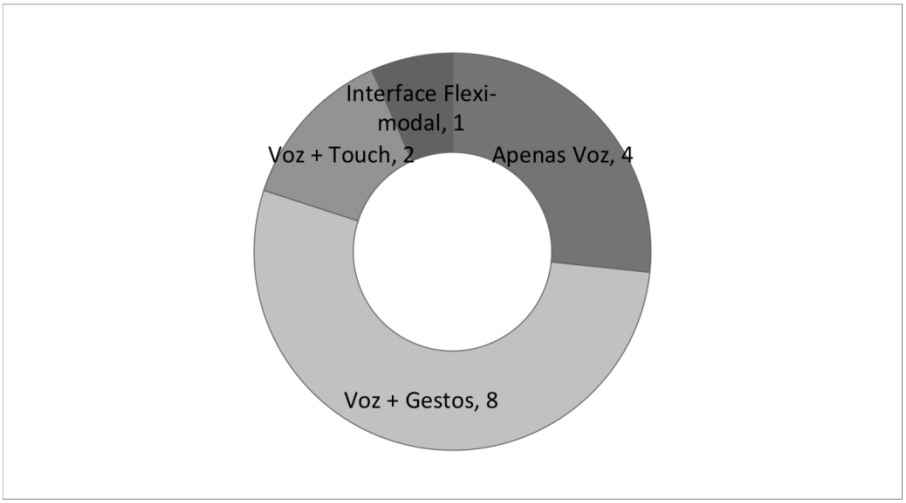

Figura 3.7 – Meios de entrada.

Os trabalhos que utilizaram apenas voz como meio de entrada tiveram em sua maioria a utilização do meio de reconhecimento por Diálogo, usando a abordagem onde o usuário realizava as perguntas através de *queries* para a aplicação e a mesma gerava uma visualização resposta (Sun et al. 2010). Estes artigos justificavam que o usuário sabia a pergunta que queria fazer para a base de dados, porém, não necessariamente sabia como transformar esta dúvida em comandos de interface (Cox et al. 2001).

Na imagem acima é possível notar que um dos trabalhos (Myers et al. 2002), fez uso de um conjunto de entradas, no caso, uma interface *flexi-modal*. Essa interface utiliza os seguintes meios de entrada: combinando voz, gestos, escrita a mão, rastreamento de olhos, interação por laser, manipulação direta por *touch*, e a manipulação por diversos aparelhos portáteis.

Além das interfaces multimodais, foi notório durante a análise dos trabalhos, a presença de técnicas tanto de visualização de informação quanto de visualização científica, sabendo que ambas áreas foram consideradas nesta revisão bibliográfica. A Figura 3.8, revela a proporção da atuação destas áreas nas publicações, objetivando apresentar um panorama geral da presença da visualização científica e da visualização de informação nos trabalhos revisados.

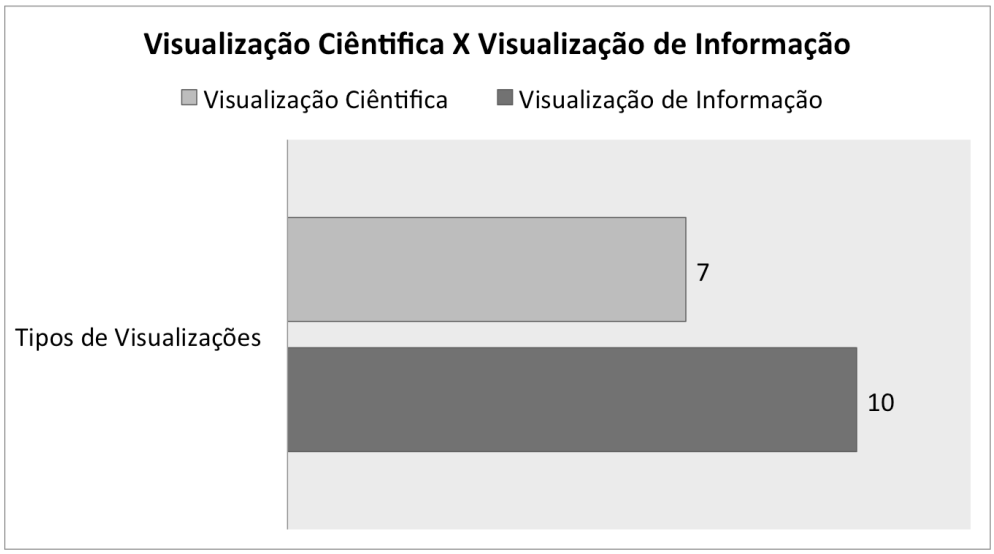

Figura 3.8 – Panorama geral da presença de visualização cientifica e de informação presente nos trabalhos analisados.

Durante a análise das publicações foi averiguado a quantidade de usuários envolvidos nos testes das ferramentas desenvolvidas. Este fator foi observado com o intuito de investigar o padrão e encontrar uma média de usuários suficiente para realizar os testes e comprovar sua usabilidade. Na Figura 3.9 é representada a quantidade de usuários por anos e a média de usuários envolvendo todos os trabalhos que forneceram estes dados.

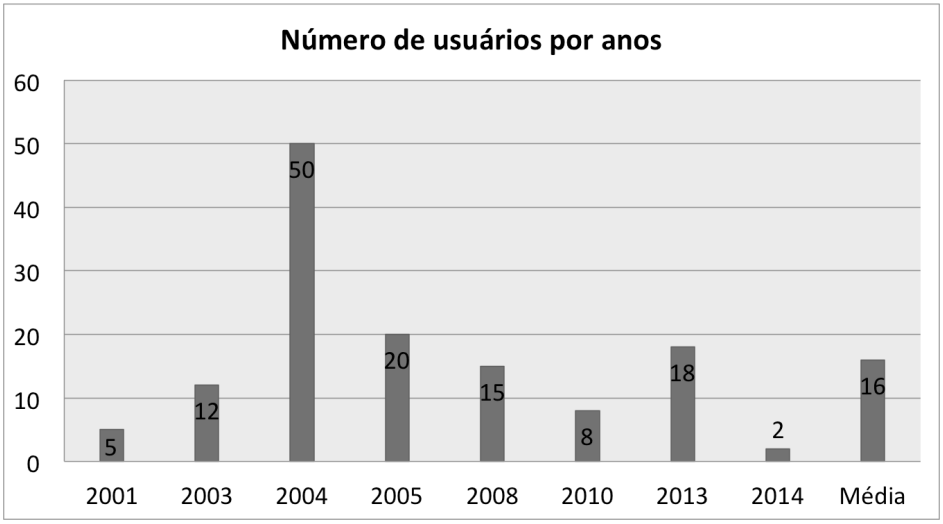

Figura 3.9 Quantidade de usuários por anos e média de usuários envolvidos nos testes de usabilidade.

Os dois extremos aconteceram nos anos de 2004 e 2014 respectivamente, onde em 2004 no artigo (Saverio et al. 2004), foram realizados testes com 43 alunos e 7

professores da faculdade de *Virginia Tech*. Em contrapartida, em 2014, foram utilizados apenas 2 usuários para avaliar a ferramenta "*Touching the Cloud*" (Lubos et al. 2014).

Durante a análise dos artigos, foi constante a presença de testes que mensuravam medidas quantitativas, como tempo para realizar alguma tarefa (Chen et al. 2005) e (Sun et al. 2010), como também medidas qualitativas que mensuram a impressão subjetiva do usuário sobre a ferramenta (Miller et al. 2008), (Gkesoulis et al. 2015) e (Sabir et al. 2013).

3.4.2 Discussão Sobre a Pergunta Principal

A pergunta principal "Qual o estado da arte da aplicação de interação por voz em uma ferramenta de visualização de informação?", foi utilizada como referencial para este levantamento e revisão bibliográfica. Nos 15 anos de publicações analisadas, destaca-se que os artigos se encontram em três níveis distintos de interação por voz.

- **Nível 0:** aprimorar a experiência do usuário.
- **Nível 1:** manipulação da visualização e seleção de dados (Mimetização do Mouse).
- **Nível 2:** Aplicação de configurações e filtros através de queries dinâmicas.

A Figura 3.10, apresenta a proporção destes níveis nos trabalhos analisados.

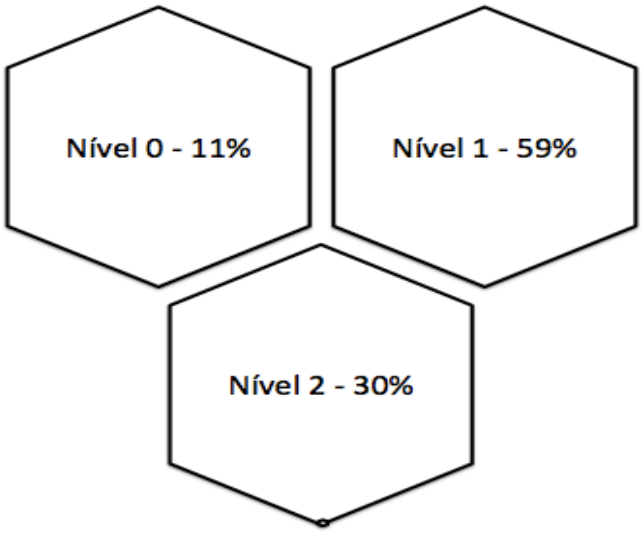

Figura 3.10 Proporção dos níveis de interações nos trabalhos analisados.

Alguns trabalhos do nível 0 (Zero) utilizaram tanto reconhecedores de voz quanto sintetizadores de voz, com o objetivo de aprimorar a experiência do usuário. A

exemplo, a utilização de ASR para reconhecer o discurso presente em um vídeo e assim gerar uma legenda para que o usuário possa acompanhar o vídeo através de áudio e texto (Heeren et al. 2007). Enquanto que a utilização de TTS, se deu na leitura dos dados presentes na ferramenta *CineCubes* (Gkesoulis et al. 2015), que utiliza uma visualização narrativa para apresentar os dados ao usuário.

Ao passo que o nível 1 (Um) teve uma maior parcela de participação nas publicações analisadas, onde a voz assumia o papel de mimetizar o uso do mouse. No caso, a voz era majoritariamente utilizada para manipular e selecionar elementos da visualização. Isto pode ser acompanhado nos trabalhos (Chen et al. 2005), (Miller et al. 2008), (Sabir et al. 2013), (Andries et al. 2000) e (Ali et al. 2005), onde mesmo se tratando de técnicas de visualização diferentes, os mesmos utilizam o reconhecimento de voz desta maneira.

O nível 2 (Dois) de interação por voz é o nível que traz a menor carga cognitiva ao usuário, pois este é o mais próximo da interação entre humano-humano. Apesar de ter uma parcela de participação menor que o nível 1, o nível 2 apresenta maior eficiência na interação por voz e consequentemente uma maior complexidade para os autores desenvolverem este tipo de interação. A utilização deste modo de interagir pode ser acompanhado nos trabalhos (Cox et al. 2001), (Sun et al. 2010) e (Sharma et al. 2003).

O estado da arte pode ser entendido como uma evolução natural na interação por voz entre homem e máquina, onde os níveis descriminados na Figura 3.10 representam esta evolução, sendo o nível zero o de menor expressão por fazer uso da voz apenas como uma saída para o usuário.

E para comentar sobre os níveis 1 e 2, será usado uma analogia. Nesta situação, um cliente chega em uma banca de revistas e deseja comprar uma revista de automóveis especifica. Usando a abordagem baseada no nível 1 de interação por voz, o cliente inicia o diálogo, falando ao vendedor que quer uma revista de automóveis. O vendedor aponta a seção de automóveis, então o cliente pede para que o vendedor pegue a terceira revista da direita para esquerda na quarta linha da terceira coluna. A mesma situação poderia ser resolvida no nível 2 de interação por voz. O cliente chega na banca de revistas e

pede exatamente o que quer, a revista de automóveis especifica e o vendedor entregaria a revista ao cliente.

Esta analogia ilustra a abordagem mais eficiente na utilização da interação por voz, no caso o nível 2 de interação. Mesmo tendo grande complexidade em sua implementação, tem o efeito oposto para o usuário, devido retirar a necessidade de lembrar de ícones e comandos de interface, assim, diminuindo a carga cognitiva do usuário.

3.4.3 Discussão da Pesquisa - Questão Secundária 1 (Qs1): Quais as técnicas de visualização de informação e ferramentas de reconhecimento de voz foram utilizadas nos experimentos?

Para o reconhecimento de voz foram empregadas diferentes tecnologias, como pode ser visto na Figura 3.11. Entre elas, as duas mais utilizadas foram o IBM Via Voice e Microsoft speech API. Para o reconhecimento de voz em português Brasileiro no artigo (Krammes et al. 2014), foi utilizado a *Android* Speech API.

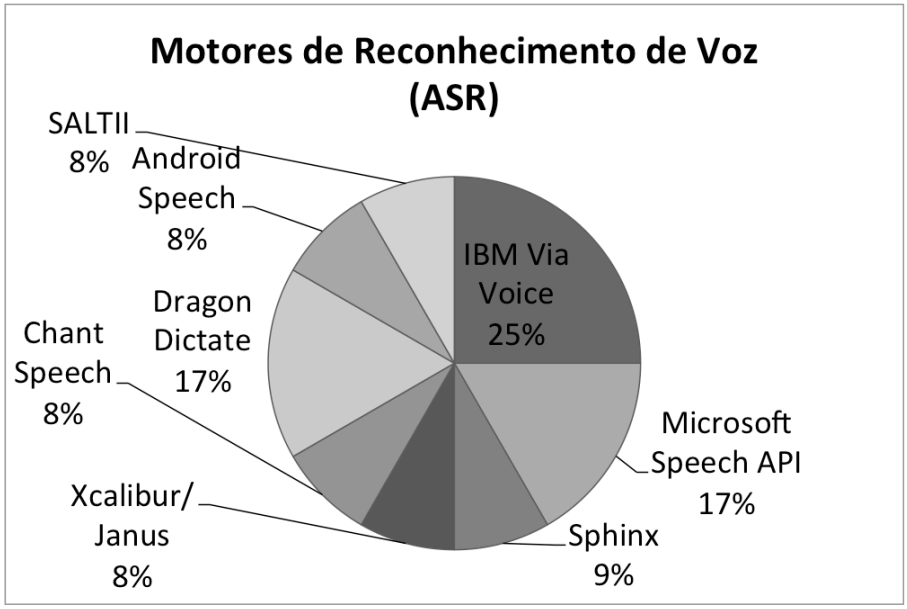

Figura 3.11. Motores de Reconhecimento de Voz(ASR).

Apesar do extensivo uso pela comunidade científica a plataforma IBM Via Voice foi descontinuada. Mesmo assim, é notória a importância e o crescimento de outros motores de reconhecimento de voz.

Nos trabalhos mais recentes a utilização do motor da Microsoft é evidente, tanto quanto a utilização do motor de reconhecimento *Sphinx* que é da *Carnegie Mellon*

University, isto se dá principalmente por ser de acesso público. Enquanto que a utilização de reconhecedor pago pode ser visto no trabalho (Corradini et al. 2002), que fez uso do *Dragon Dictate*. Todavia, ao se tratar dos sintetizadores de voz, no trabalho (Gkesoulis et al. 2015), foi utilizado o TTS de código aberto *MARY* (DFKI, 2013).

Além dos diversos tipos de ferramentas, as publicações tiveram uma diversidade de tipos de visualizações apresentadas em suas aplicações. Entre elas, as mais tradicionais como, gráfico de barras ou por dispersão de pontos tanto em duas quanto em três dimensões. E algumas outras visualizações mais sofisticadas, como, *Computacional Fluid Dynamics*.

A Figura 3.12, apresenta todas as visualizações presentes nos trabalhos analisados, sendo que estão inclusas as visualizações em 2D e 3D.

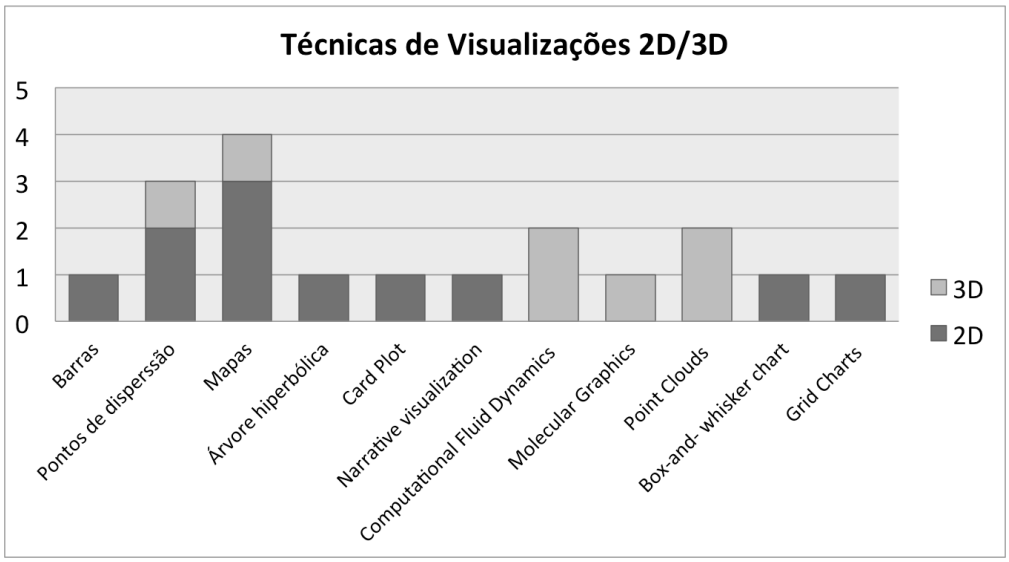

Figura 3.12 – Técnicas de Visualizações 2D/3D.

Ao observar a Figura 3.12, nota-se a presença de uma técnica chamada *Narrative Visualization***.** *E*ssa técnica pode ser vista no trabalho (Gkesoulis et al. 2015).

A abordagem guia o usuário através de um conjunto de slides contendo os dados referentes a query gerado pelo usuário e utiliza de estruturas narrativas, no caso a utilização de um TTS para narrar a história segundo a visão do autor, assim, introduzindo os dados ao usuário. Além desta técnica, houve uma extensiva utilização de visualizações geográfica em 2D/3D, geralmente utilizado para controle de tráfego

(Chen et al. 2005), quanto para situações de controle emergencial (Sharma et al. 2003).

3.4.4 Discussão da Pesquisa - Questão Secundária 2 (QS2): Quais as intenções primárias do uso da interação por voz nestes trabalhos (Comandos utilizados)?

Selecionar e criar anotações foram os comandos mais simples aplicados a interação por voz (Lubos et al. 2014). Dependendo do meio de entrada que somava com a interação por voz, os comandos variavam.

A exemplo, o trabalho (Sabir et al. 2013), onde o utilizava em conjunto com a voz uma interface de gestos, seus comandos eram, *Reset, Hands On, Hands Off, Wake Up, Go to Sleep, Zoom ,Select All ,Select Up, Select Ligand Select Proximity, Select Residue, Select Chain, Next Ligand, Deselect, Rotate, Copy e Paste*. Nesta aplicação a voz tinha o controle de ativar/desativar o reconhecimento de gestos através dos comandos *Hands On*/*Hands Off* e *Wake Up*/*Go to Sleep* para desabilitar/habilitar o reconhecimento de voz e utilizava da lógica de selecionar o atributo desejado mais próximo ao cursor e, após selecionado, poderia navegar de forma hierárquica pelos atributos relacionados através do comando *Select Up*.

Enquanto que na aplicação militar, vista no artigo (Myers et al. 2002), a abordagem de utilização da voz era a de selecionar elementos da visualização mencionado, como o nome da unidade, o tipo da unidade, a afiliação da unidade, ou a combinação destes. E cada unidade selecionada podia-se inferir perguntas como alcance de fogo, ou mobilidade. Segue abaixo uma imagem da aplicação na tela de um dispositivo portátil.

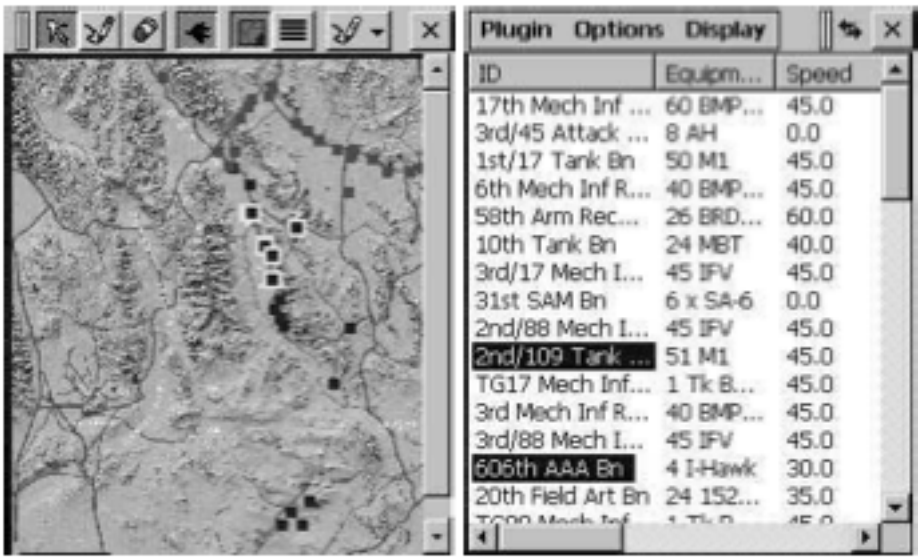

Figura 3.13 – Projeto *flexi-modal and Multi-Machine User Interfaces Battleboard* (Myers et al. 2002).

3.4.5 Discussão da Pesquisa - Questão Secundária 3 (QS3): Como são medidos quantitativamente quanto qualitativamente os testes de cada trabalho?

No trabalho (Saverio et al. 2004), a avaliação se deu com 50 usuários, sendo 9 professores e 41 alunos, onde cada um completava tarefas de navegação no software. Após, era medida a impressão subjetiva de cada um, tendo uma aprovação de 74% pelos usuários. Enquanto no artigo (Miller et al. 2008), foi utilizado o tempo no teste de usabilidade para mensurar quantitativamente as tarefas realizadas por cada usuário. E qualitativamente era aplicado um formulário com a escala Likert para que os usuários inferissem sua impressão subjetiva do software.

A avaliação quantitativa foi basicamente igual em todos as aplicações, onde era passado tarefas e o tempo para a completude delas era medido. Já qualitativamente, a impressão subjetiva do usuário sobre o software era medida através de questionários pós tarefas, onde os mesmos escolhiam uma escala que correspondesse com seu nível de satisfação em relação ao software. Em poucos trabalhos houve uma comparação entre a ferramenta proposta com outras ferramentas. Foi o caso de (Sun et al. 2010), que comparou o desempenho de sua ferramenta passando as mesmas tarefas para diferente usuários. Sendo que, metade destes usuários usavam a aplicação proposta e a outra metade o Excel. Ademais, a ferramenta provou sua eficiência na completude das tarefas.

Apenas um artigo deixou de medir a experiência do usuário para medir o desempenho do software. No trabalho (Gkesoulis et al. 2015), verifica o tempo de

execução dos diferente módulos do software para aferir seu desempenho.

3.4.6 Discussão da Pesquisa - Questão Secundária 4 (QS4): Como está sendo aplicada entrada por voz em ambientes de visualização em três dimensões?

No artigo (Sabir et al. 2013), a interação por voz é utilizada prioritariamente para selecionar ou retirar a seleção dos atributos da visualização e para manipulação da informação de forma geral, com zoom, girar e mover. Enquanto nos trabalhos que utilizam a técnica de visualização nuvem de pontos em 3D (*Point Clouds*) (Lubos et al. 2014) e (Krammes et al. 2014), a mesma é utilizada para criar anotações em uma área selecionada, bastando o usuário falar o comando *annotation* e depois ditar a anotação que será vinculada com a área selecionada. Assim possibilitando o mesmo isolar aquela área ou seleciona-la pelo seu rótulo(anotação).

Enquanto que no artigo (van Dam et al. 2000) que faz uso da técnica de visualização científica 3D *computational fluid dynamics* (CFD), utiliza um conjunto de gramáticas voltando o uso destas para criação de novas camadas na aplicação e a manipulação da visualização, como mover e girar estas camadas.

Na publicação (Sharma et al. 2003), faz uso da voz em um mapa tridimensional através das seguintes funções, solicitar, responder e informar. A interação mais frequentemente utilizada é solicitar informações da visualização (por exemplo, pedir um mapa). Nos casos em que o há ambiguidade no pedido ou o ASR não reconheceu completamente a solicitação, mas a aplicação irá responder com uma pergunta que solicitará ao usuário mais informações para resolver essas ambiguidades. Assim, quando o usuário responder, será permitida a aplicação completar o pedido inicial. A terceira ação que um usuário pode executar é informar, em outras palavras, se comunica a aplicação sobre fatos que são relevantes.

CAPÍTULO 4- APLICAÇÃO

O IVOrpheus é uma ferramenta de visualização de informação que utiliza a técnica de dispersão de pontos em três dimensões, fazendo uso de três dimensões espaciais (X,Y e Z) e de três canais visuais (Cor, Forma e Tamanho) para representar os dados. Além do mais, a ferramenta aceita entrada por voz baseada em gramáticas e reconhece comandos em Português Brasileiro.

Esta ferramenta se enquadra no nível 1 (um) de interação por voz (ver Figura 2.1), pois a mesma utiliza a interação por linguagem natural para manipulação da visualização e seleção dos dados (mimetizando o mouse).

E como visto na seção 3.4.1 Panorama dos estudos analisados, as gramáticas tem uma maior acurácia no reconhecimento das palavras, enquanto comparada com o método de reconhecimento de dialogo. Com intuito de diminuir a quantidade de hipóteses erradas geradas pelo sistema Coruja. Foi adotado o meio de reconhecimento por gramáticas visando melhorar a experiência do usuário em relação ao sistema utilizado.

Logo, como analisado na revisão bibliográfica nas seções 3.4.4 Discussão da pesquisa - Questão Secundária (QS2) e 3.4.6 Discussão da pesquisa - Questão Secundária (QS4). Foi possível notar uma similaridade nos comandos de voz adotados nos artigos avaliados durante a revisão bibliográfica. E tais comandos geralmente abrangem as funcionalidades referentes a manipulação da visualização, seleção de objetos da visualização e transformação da visualização. Baseado nisto, tais funcionalidades foram contempladas na ferramenta IVOrpheus. Através das macro funcionalidades de configurar, filtrar, interagir e detalhes sobre demanda.

No mais, neste capitulo serão mostrados os componentes da aplicação, as funcionalidades e aspectos conceituais da ferramenta e por fim, o gerenciamento das gramáticas utilizadas na ferramenta IVOrpheus.

4.1 Ferramentas Utilizadas

Para o desenvolvimento da aplicação IVOrpheus com a interface de interação natural. Foram utilizados softwares livres, no reconhecimento da voz (Coruja) e para a geração da visualização por pontos de dispersão (JMathplot). Abaixo é apresentado uma

visão geral sobre estas ferramentas e sua utilização na aplicação IVOrpheus.

4.1.1 Coruja

O processamento de voz inclui várias tecnologias, dentre as quais o reconhecimento automático de voz ou ASR (Rabiner, 1993), e a síntese de voz ou TTS (Taylor, 2009), são as mais proeminentes.

Sistemas TTS são módulos de software que convertem textos em linguagem natural em voz sintetizada. O reconhecimento de voz pode ser visto como o processo inverso ao TTS, onde o sinal de voz digitalizado é convertido em texto.

A proposta apresentada neste trabalho utiliza a funcionalidade de interação via reconhecimento automático de voz em Português Brasileiro. Para isso, fez uso do software livre Coruja (Silva et al. 2010), este oferece modelos acústico e linguístico, além de uma interface de programação (API) própria construída para facilitar a tarefa de controlar o motor de reconhecimento de voz Julius. Essa API contém métodos e eventos que permitem ao programador abstrair requisitos de baixo nível da engine.

Visando flexibilidade quanto à plataforma, a versão mais atual do Coruja (Oliveira et al. 2011) oferece suporte a especificação Java Speech API (JSAPI), como mostrado na Figura 4.1, tanto em aplicações com gramática controlada (ou comando-e-controle), como texto-livre (ou ditado).

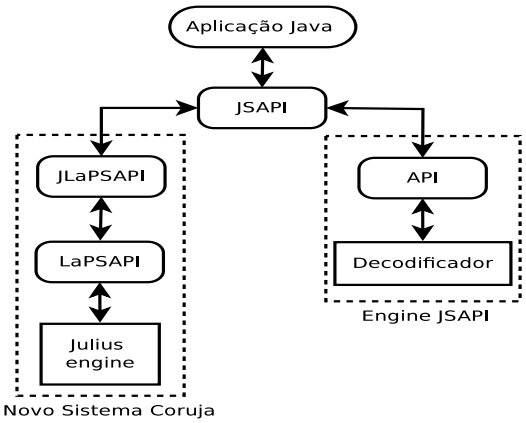

Figura 4.1. Visão Geral da arquitetura do Coruja.

Nessa arquitetura, o acesso ao Coruja é feito através de código especificado pela JSAPI. Com isso, existe a possibilidade de alternar entre o Coruja e qualquer outro "engine" que siga a especificação JSAPI, como o Sphinx4, por exemplo, sem a necessidade de alteração no código Java.

4.1.2 JMathplot

JMathplot (Richet, 2007) é uma API multiplataforma desenvolvida em Java. A mesma foi utilizada no IVOrpheus para o desenvolvimento do módulo de visualização de informação. Sendo que a técnica escolhida para o estudo de caso é a de dispersão de pontos em três e duas dimensões. Porém, a API não se limita a somente esta técnica. Possuindo outras técnicas em três e duas dimensões. Como, nuvem de pontos, histogramas, coordenadas paralelas, gráfico de linha, *Staircaseplot,* entre outras.

O JMathplot foi escolhido devido possuir uma documentação bem escrita e por ser de código aberto. Assim, permitindo modificar e adaptar a técnica de visualização de informação desejada. No caso do IVOrpheus foi modificado os meios de entrada de dados na interação com a técnica de dispersão de pontos em 2D/3D.

4.2 Aspectos Conceituais

IVOrpheus é uma aplicação de visualização de informação (*infoVis*), e atende as diretrizes básicas de uma boa ferramenta de *infoVis* definidas por (Shneiderman, 1996), comumente denominado de o mantra de visualização da informação, sendo este composto por : visão geral dos dados – usuário deve ter uma noção geral dos dados para análise; zoom semântico - focar em um subconjunto dos dados; filtros – diminuir o conjunto de dados de análise; e detalhes sob demanda – apresentar dados que não estão representados visualmente (dados ocultos).

4.2.1 Interface

O IVOrpheus possibilita dois tipos principais de interação: uso do mouse, teclado e por comandos de voz. As interfaces por comando de voz tem se tornado mais populares, muito em função dos dispositivos móveis (SIRI, 2016), (CORTANA, 2016) e (Google Now, 2016), e de maneira geral, pela maior precisão no reconhecimento da fala.

A primeira diretriz para construção da interface do IVOrpheus é que ela fosse única tanto para interação com teclado e mouse, quanto para comandos de voz. Foram considerados cinco pontos principais (Beasley et al. 2001), (Lee et al. 2006):

- Botão equivalente a funcionalidade *Home*: Um comando/botão que retornasse o usuário para um ponto inicial conhecido;

- Comunicação significativa: usuário deve identificar facilmente os comandos disponíveis para interação e seus significado, e ainda é possível obter ajuda sobre os comandos disponíveis na tela;
- Mínima ação do usuário: os comandos devem ser simples em cada tela, um clique do mouse ou um comando de voz. Os dados de entrada devem estar na tela ou permitir poucos dados para entrada pelo usuário;
- Consistência e padronização na forma de interagir e telas: as formas de interação e padronização das telas são mantidas. Por exemplos, os mesmos comandos de cancelar e voltar em todas as telas, mesmos comandos em contextos diferentes para operações semelhantes;
- Reconhecimento de voz independente do orador.

A interface do IVOrpheus está dividida em três áreas principais, como mostra a Figura 4.2: a barra de Opções (1), a área de Visualização (2) e a barra de Menu (3).

Em todas as áreas, cada opção apresentada na tela pode ser realizada pela interação por comando de voz ou clique do mouse. Para interação por comando de voz, os rótulos dos botões e dos menus são os comandos disponíveis para interação. Por exemplo, a opção Configurar pode ser acessada falando o comando de voz "Configurar".

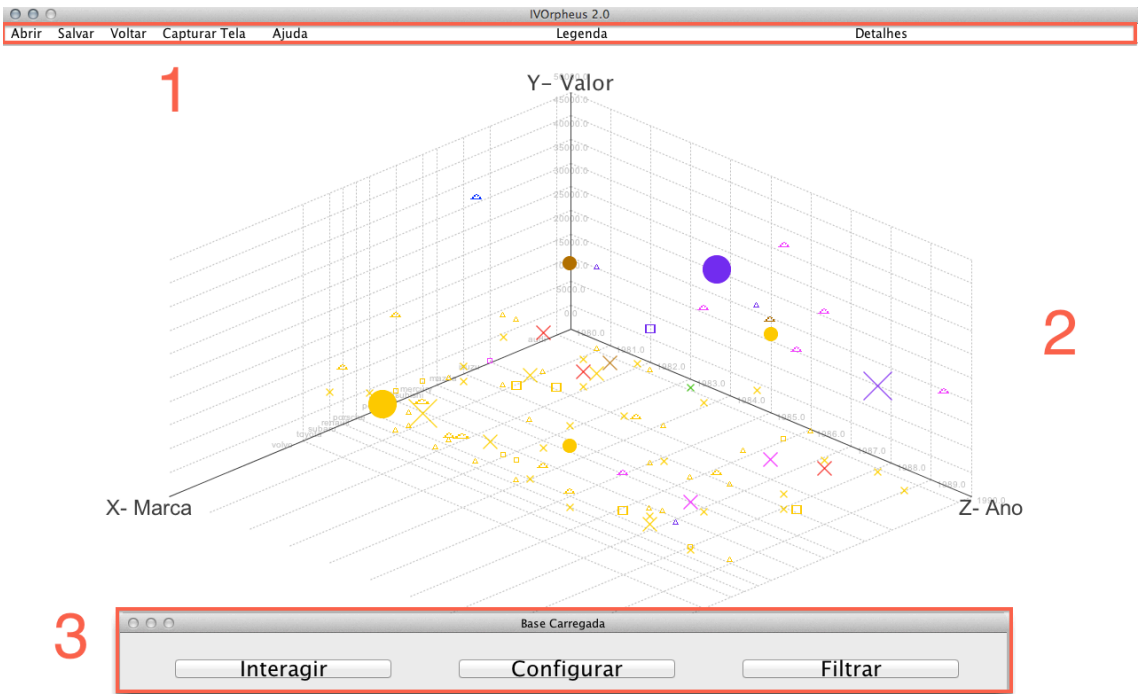

Figura 4.2. Tela inicial com base carregada.

Na barra de Opções, o usuário pode utilizar os comandos de voz “Abrir”, “Salvar”, “Voltar”, “Capturar Tela”, “Ajuda”, “Legenda” e “Detalhes”, todos exibidos visualmente.

A barra de Menu fica habilitada somente depois que uma base de dados é carregada na ferramenta, e possui os seguintes comandos iniciais: “Configurar”, “Filtrar” e “Interagir”.

Após um desses comandos serem utilizados, outros sub-menus são habilitados.

4.2.2 Funcionalidades

As funcionalidades do IVOrpheus seguem as principais características de uma boa ferramenta de visualização. A seguir são apresentadas tais funcionalidades.

Configurar/Filtrar: é possível configurar/filtrar os eixos X, Y, Z, e os canais visuais Cor, Forma e Tamanho. A Figura 4.3 apresentam a visão geral das funcionalidades configurar/filtrar. A configuração/filtragem dos eixos pode ser tanto aplicada a dados categóricos (valores discretos), quanto contínuos (valores flutuantes), e os canais cor, forma e tamanho somente podem ser aplicados para dados categóricos.

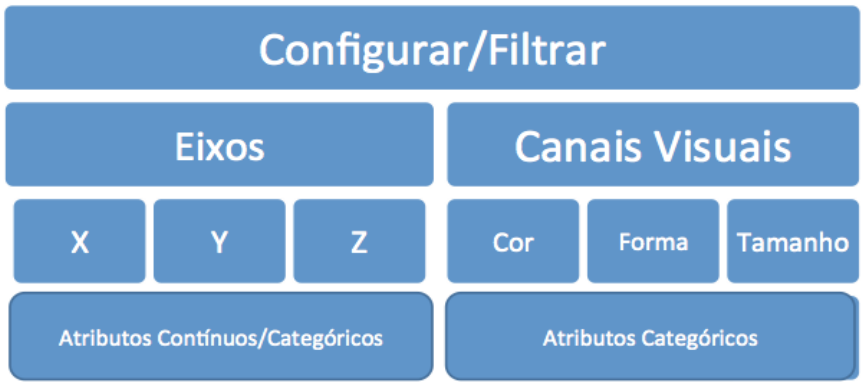

Figura 4.3. Ações do menu Configurar/Filtrar.

Interação: com o objetivo de manipular o gráfico como todo, a interação pode ser aplicada para girar, transladar, aumentar ou diminuir o tamanho do gráfico. Além disto, os sub-menus de interação possuem as funcionalidades de estado inicial e parar, pois, como os usuários de voz não necessariamente sabem quantos graus querem aplicar na rotação da visualização ou quanto deseja aumentar a escala da mesma.

O sistema IVOrpheus aplica de forma contínua o aumentar ou diminuir de

escala, o girar, e o transladar. Sendo que o usuário pode utilizar o botão parar, bem como para cessar o incremento ou decremento destes valores. Enquanto o botão estado inicial retorna a visualização para os seus valores de escala, giro e posição iniciais.

Figura 4.4. Ações do menu interação.

Barra de opções: as funcionalidades contidas nesta barra estão presentes durante toda a execução da ferramenta (ver Figura 4.5), sendo estas as regras globais que podem ser ditas a qualquer momento pelo usuário. A exemplo, “Abrir”, a qualquer momento o usuário poderá carregar uma base nova. Já os comandos Capturar tela, Legenda, e Detalhes, só estarão habilitados após uma base já estar carregada e configurada.

A opção de capturar tela permitirá ao usuário tirar um *Screenshot* da tela atual, assim podendo compartilha-la por e-mail ou outro meio de comunicação. Enquanto a opção de detalhes chamará o painel detalhes onde o usuário poderá configurar quais atributos deseja receber informações extras e ao selecionar um ponto tais detalhes extras ficaram visíveis.

Figura 4.5. Ações da barra de opções.

IVOrpheus assim como toda ferramenta de visualização de informação deve

atender as funcionalidades básicas para atender a definição estabelecida por (Shneiderman, 1996): "*Overview first, zoom and filter, then detail on demand*". Sendo que todas estas funcionalidades foram comtempladas pela aplicação IVOrpheus, como apresentado nas Figuras 4.3, 4.4 e 4.5.

4.3. Arquitetura

No IVOrpheus foi utilizado o padrão arquitetural MVC (*Model View Controller*), no qual a aplicação é dividida em três componentes principais, cada um com sua própria funcionalidade (Bushman, 1996).

A Visão (*View*) é responsável por mostrar uma resposta visual para as ações do usuário, o Controlador (*Controller*) trata da comunicação entre o Modelo e a Visão, e o Modelo (*Model*) é responsável pelo tratamento dos dados e lógica de acesso aos dados.

O diagrama de classes da ferramenta IVOrpheus pode ser visto na Figura 4.6, e foi dividido em três pacotes: *Model*, *View* e *Controller*, correspondendo aos três pacotes do MVC.

O pacote *View* contém a classe *Interface* que é responsável pelo que será exibido em tela aos usuários, representando o módulo "Visão" no padrão MVC.

A classe *Interface* determina como será criada a interface de usuário e o que será exibido nela. Por não possuir lógica de negócios aplicada à classe, a mesma é acionada pela classe *BtnListner* para a chamada dos painéis e da atualização da interface. Enquanto a classe *InfoVisModule* envia a visualização por pontos de dispersão para o painel principal da classe *interface,* onde será apresentada na área de visualização.

O pacote *Controller* representa o módulo Controlador da arquitetura, que é responsável por fazer a comunicação entre o pacote *View* com o pacote *Model*. Este módulo é composto pelas classes componentes do módulo de voz: *Recgonizer*, *RecognizerListener* e *Translator*. E as classes *BtnListener*, *InfoVisModule* com o subsistema *JMathplot*.

A classe *Recognizer* carrega e gerência o Coruja e todas as gramáticas. Enquanto o *RecognizerListener* fica "ouvindo" as entradas de voz do usuário e com o uso das gramáticas definidas em *Grammars* interpreta quais entradas (regras) serão repassadas

em forma de comandos para o *Translator*.

A classe *Translator*, faz a associação de um comando de voz recebido com um botão abstrato, sendo este passado para o *BtnListener* que chamará a função pertinente ao botão. Enquanto a classe *InfoVisModule* e o subsistema *JMathplot* são responsáveis pelas técnicas de visualização disponíveis e seus respectivos métodos para manipulação e geração das mesmas.

O pacote *Model* gerencia a base de dados e os atributos contidos nela. E é composto principalmente pelas classes *Attribute*, *Directory*, *Grammars* e a API Coruja.

A classe *Directory* retorna os nomes das bases de dados presentes no diretório raiz da aplicação. A classe *Attribute* lê as bases e trata os dados da base selecionada na ferramenta, atribuindo um tipo de dado correspondente para cada atributo, de acordo com os valores existentes por atributo, tornando possível a manipulação desses valores pelos métodos de *Grammars*.

A classe *Grammars* representa todas as gramáticas presentes na aplicação IVOrpheus e seu sistema de escrita dinâmico. Ela recebe os dados das classes *Directory* e *Attributes*, assim, escrevendo estes dados nas gramáticas dinâmicas: Carregar, Filtrar Categóricos e Atributos. Sendo que a gramática Carregar recebe os nomes das bases de dados presentes no diretório raiz da aplicação. E as gramáticas Atributos e Filtrar Categóricos são responsável por receber os atributos da base de dados escolhida e os valores únicos de cada atributo respectivamente.

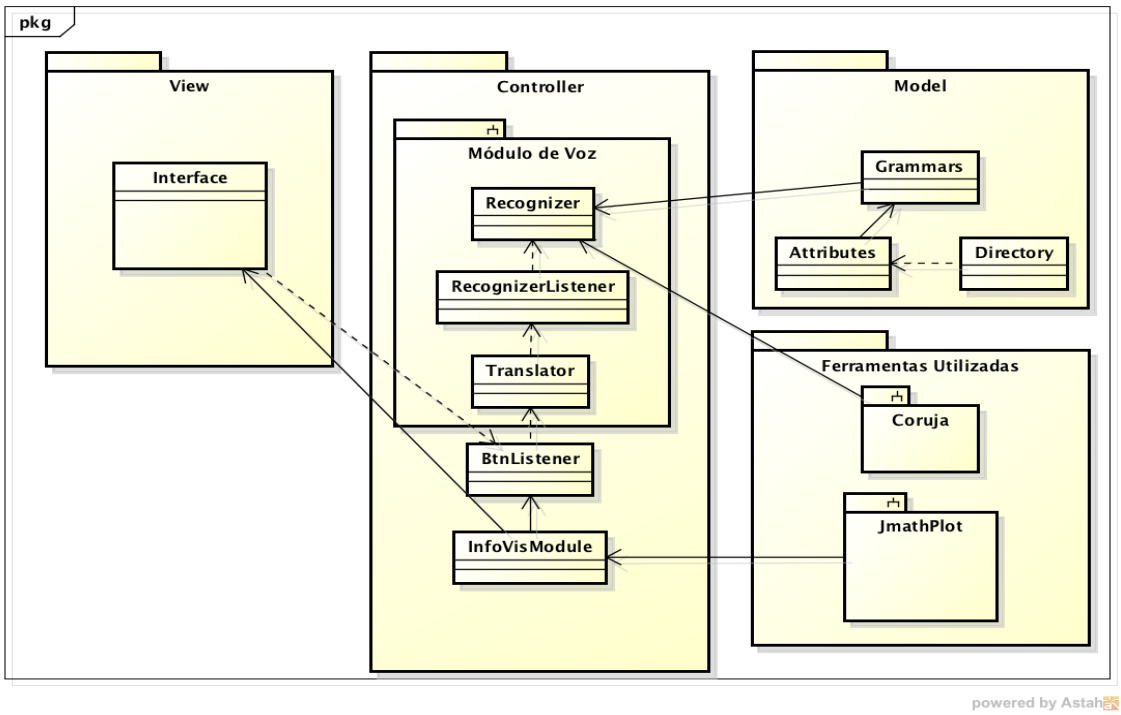

Figura 4.6. Diagrama de classes do IVOrpheus.

4.4. Fluxo De Telas

Esta seção visa apresentar a aplicação IVOrpheus e mostrar através de um fluxo de telas, suas funcionalidades e suas gramáticas. Sendo que tal fluxo também apresentará a solução das tarefas presentes no segundo teste com os usuários (ver seção 5.6).

Ao iniciar a tarefa, o usuário se depararia com Figura 4.7, que apresenta uma visualização vazia, pois até o momento não há nenhuma base carregada e configurada. Neste estado da aplicação é carregada a gramática dinâmica “Carregar”. Tendo, esta, todos os comandos de voz disponíveis na tela, a saber, “Abrir”, “Salvar” ,“Voltar”, “Capturar Tela”, “Ajuda”, “Legenda” e “Detalhes”. Além destes, a gramática possui os nomes dos arquivos bases a serem carregado, ver Figura 4.8. Estes são apresentados ao usuário, quando o mesmo chama a funcionalidade “Abrir”.

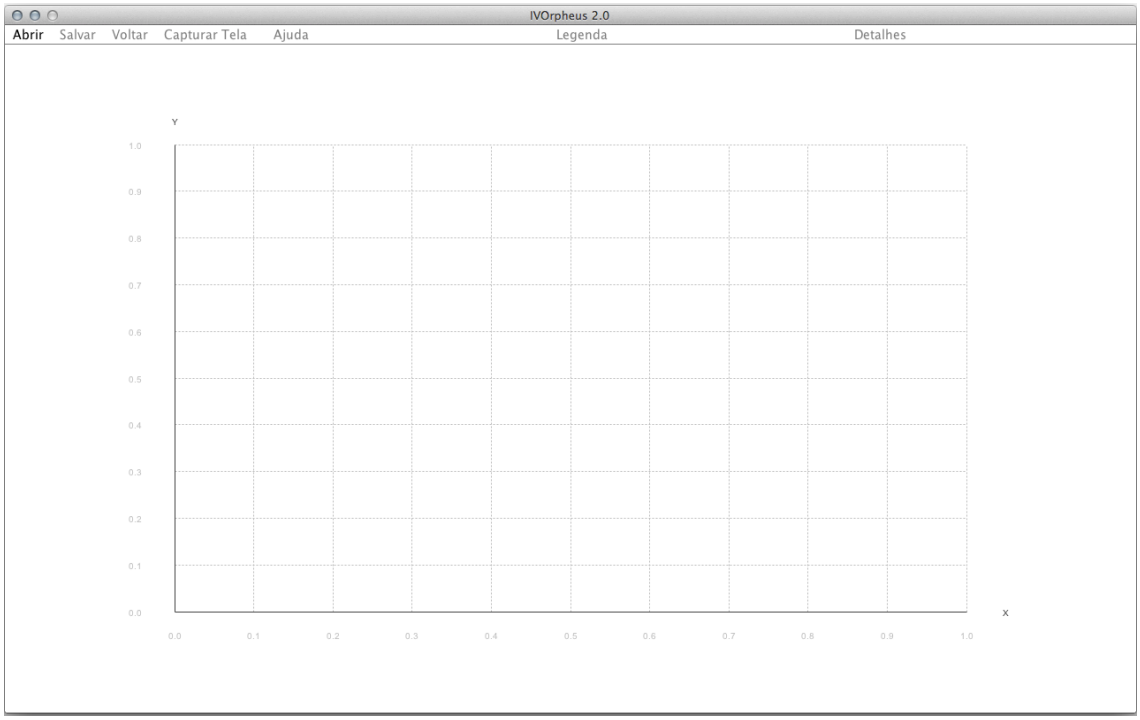

Figura 4.7. Início da aplicação.

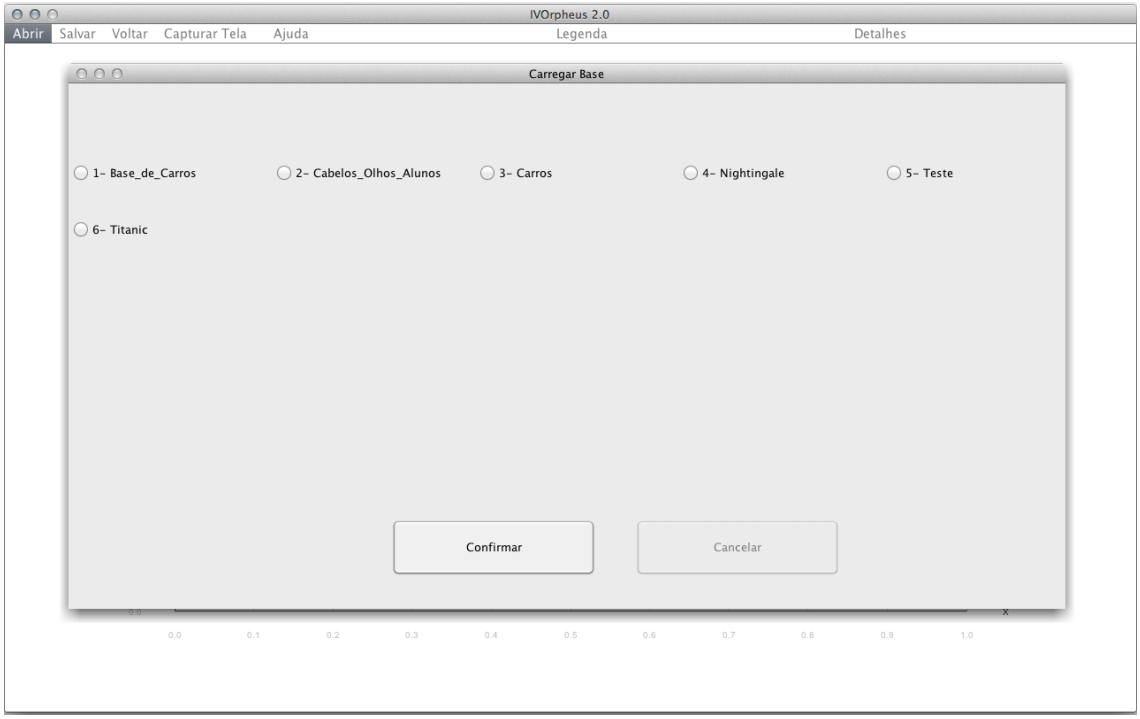

Figura 4.8. Carregar base.

Na Figura 4.8, é apresentado ao usuário todas as bases contidas no diretório raiz da aplicação.

Sabendo que todos os nomes das bases foram escritos na gramática “Carregar”, sendo que cada nome é escrito como regra local desta gramática. Quando o usuário seleciona a base desejada, que neste exemplo é a “Base_de_Carros”, o mesmo será redirecionado para a Figura 4.9.

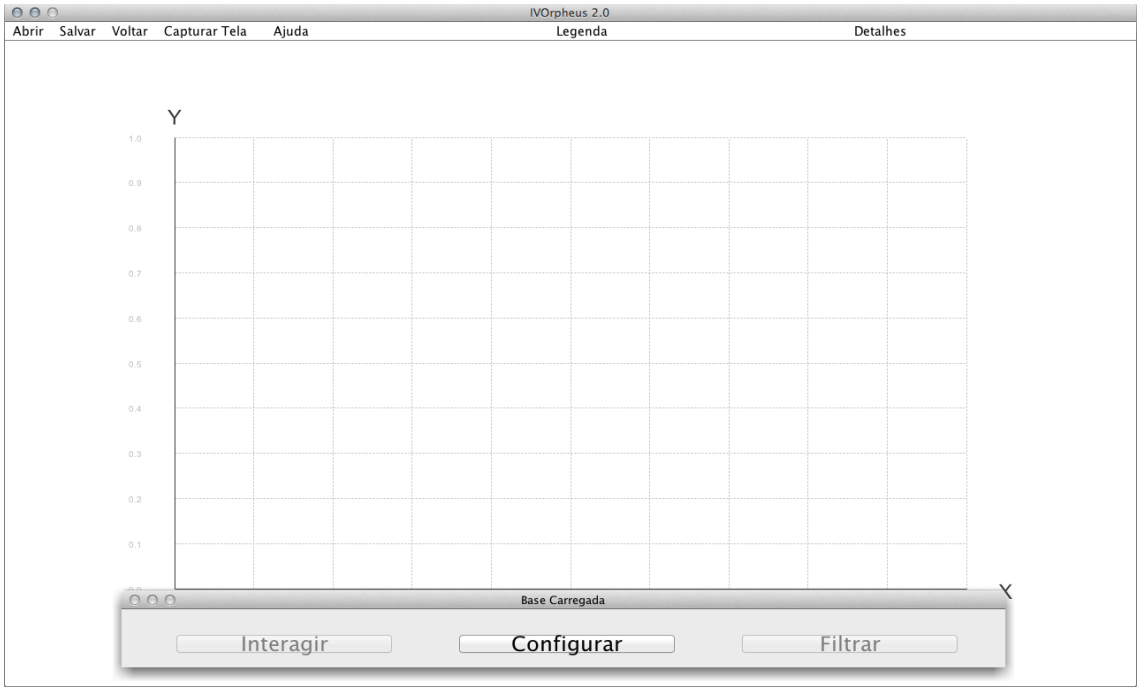

Figura 4.9. Interagir, configurar e filtrar.

Neste estado do IVOrpheus é carregada a gramática estática “IFC”, sendo o acrônimo de interagir, filtrar e configurar. E como pode ser observado na Figura 4.9, o botão de filtrar e interagir estão desabilitados. Isto se dá, por que o programa ainda não tem nenhuma base configurada. Em decorrência disso, o próximo passo que o voluntário deveria executar para a completude da tarefa, seria configurar os eixos X, Y e Z, para os seus respectivos valores, Marca, Valor e Ano, sendo que quando o participante entrar com o comando configurar, a Figura 4.10 seria disposta.

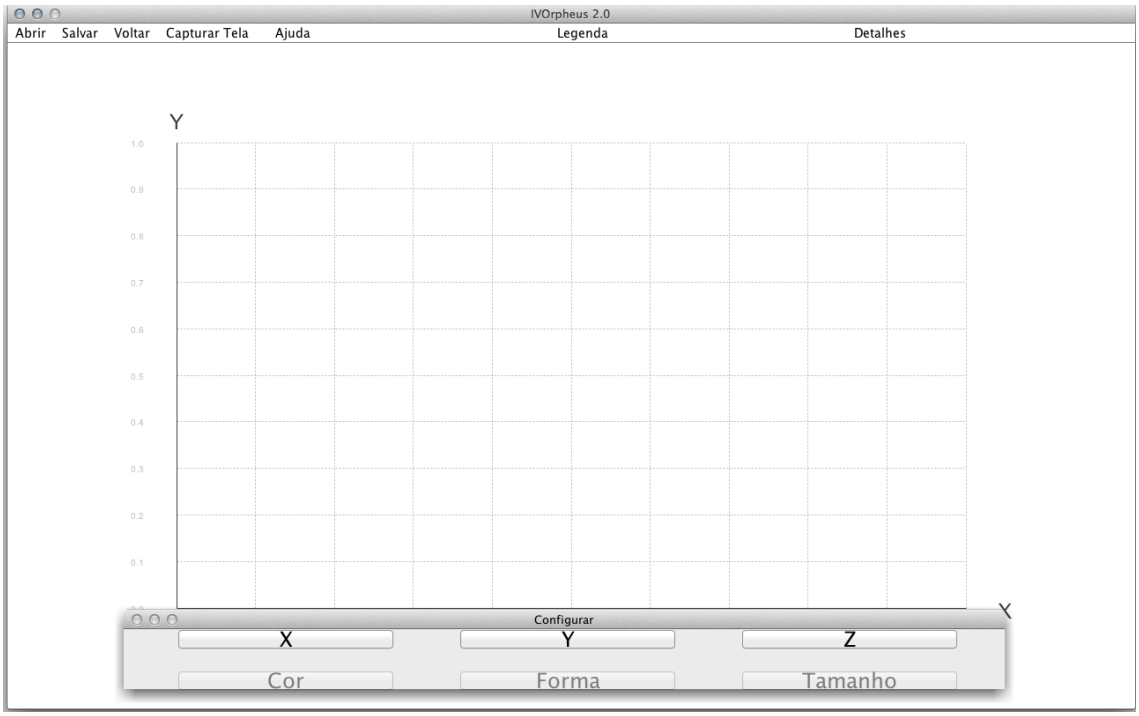

Figura 4.10. Eixos.

Na Figura 4.10 a gramática estática “FC” é carregada, sendo o acrônimo para Filtrar e Configurar. Tal nome é devido ambas funcionalidades possuírem as mesmas regras locais, no caso tais regras são, “Eixo X”, ”Eixo Y”, “Eixo Z”, “Cor”, “Forma” e “Tamanho”. Após o usuário escolher qual eixo deseja configurar a Figura 4.11 é apresentada.

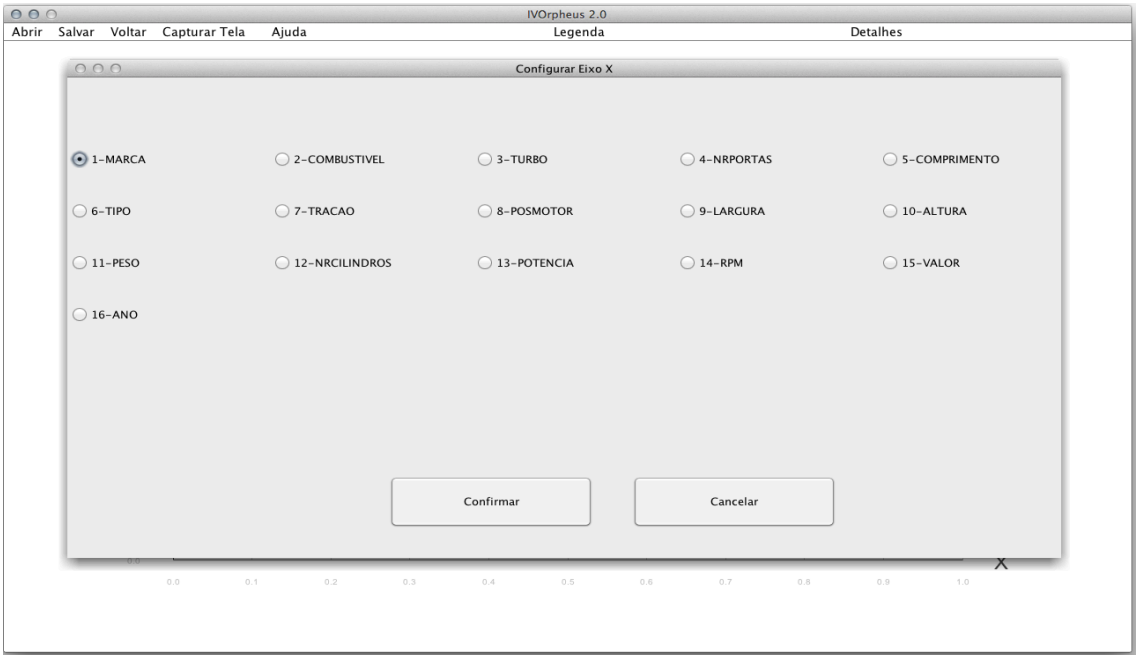

Figura 4.11. Eixo X.

Figura 4.12. Eixo Y.

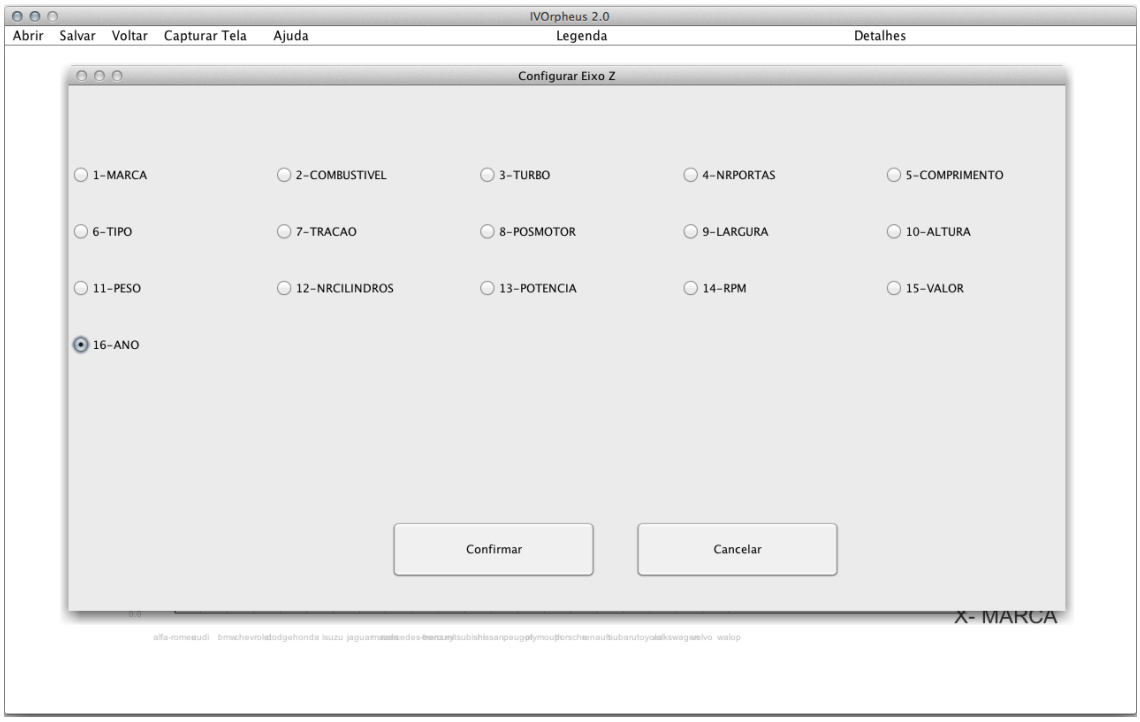

Figura 4.13. Eixo Z.

As Figura 4.11, Figura 4.12 e Figura 4.13 são respectivamente os atributos que foram configurados para os eixos X, Y e Z. É possível notar que os mesmos não estão acentuados devido ao processo de serem escritos automaticamente na gramática "Atributos" e o software coruja ter certa dificuldade no reconhecimento de caracteres

especiais. Após o usuário acertar o botão Confirmar ou entrar com o comando de voz "Confirmar", a visualização por dispersão de pontos em três dimensões será apresentada na área principal da aplicação, como pode ser visto na Figura 4.14.

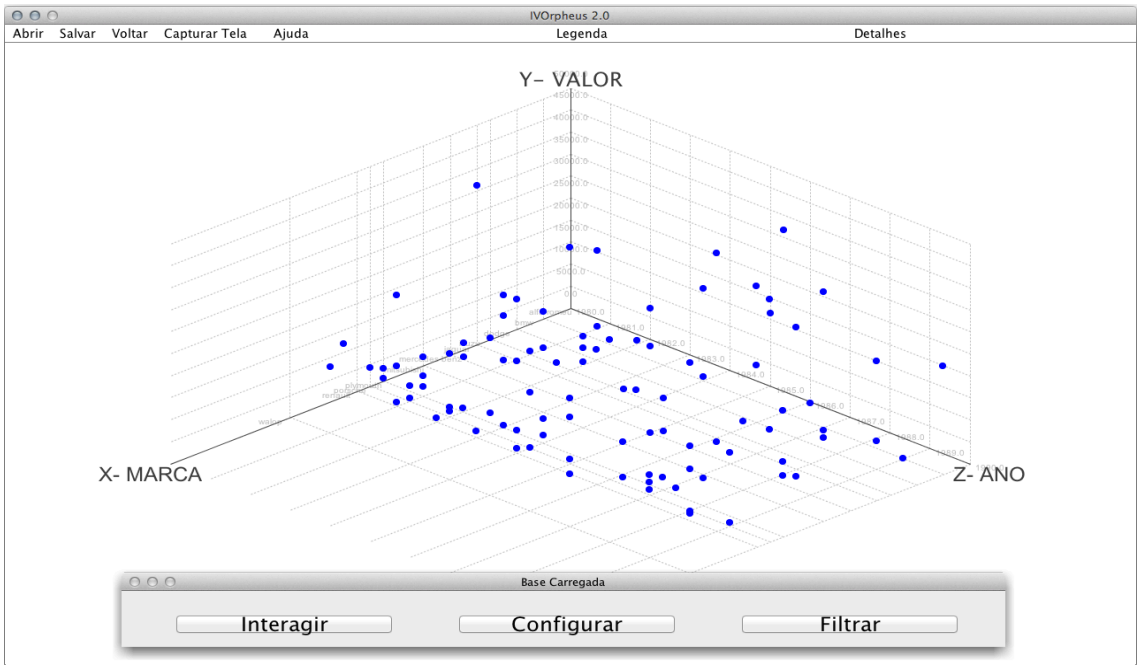

Figura 4.14. Base configurada.

Após configurar a base, o participante deveria aplicar filtros nos dados. Ao entrar com o comando "Filtrar", as mesmas opções apresentadas na Figura 4.10, são mostradas, porém o usuário desta vez deveria selecionar o eixo desejado para aplicar o filtro.

Na Figura 4.15, é apresentado o filtro categórico que apresenta os valores únicos do atributo configurado para o eixo X que recebeu o atributo "Marca". Dentro deste atributo, os valores únicos são: "Alfa-Romeu", "Audi", "BMW" entre outras marcas. O usuário deveria selecionar os valores "BMW", "Dodge "e "Isuzu", assim removendo da visualização todos os demais valores de marcas, deixando apenas os valores selecionados visíveis.

Na Figura 4.16 apresentam a aplicação do filtro no eixo Z que recebeu o atributo Ano. O voluntário deveria selecionar o ano de 1986.

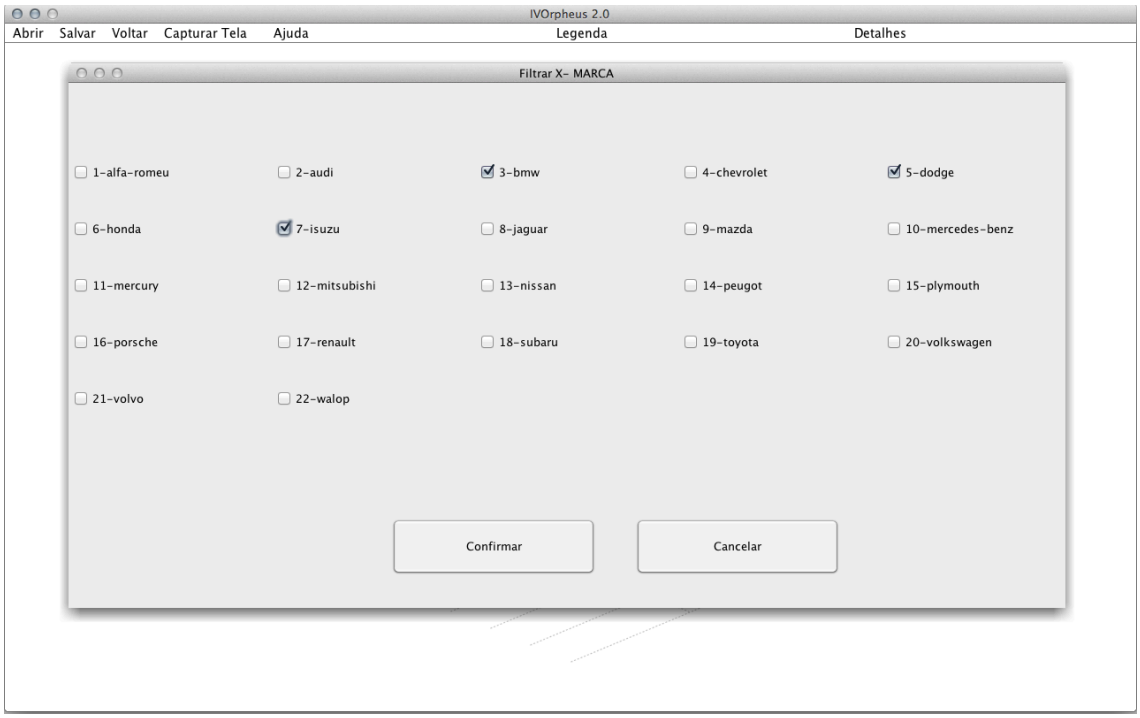

Figura 4.15.Filtrar X Marca (Categórico).

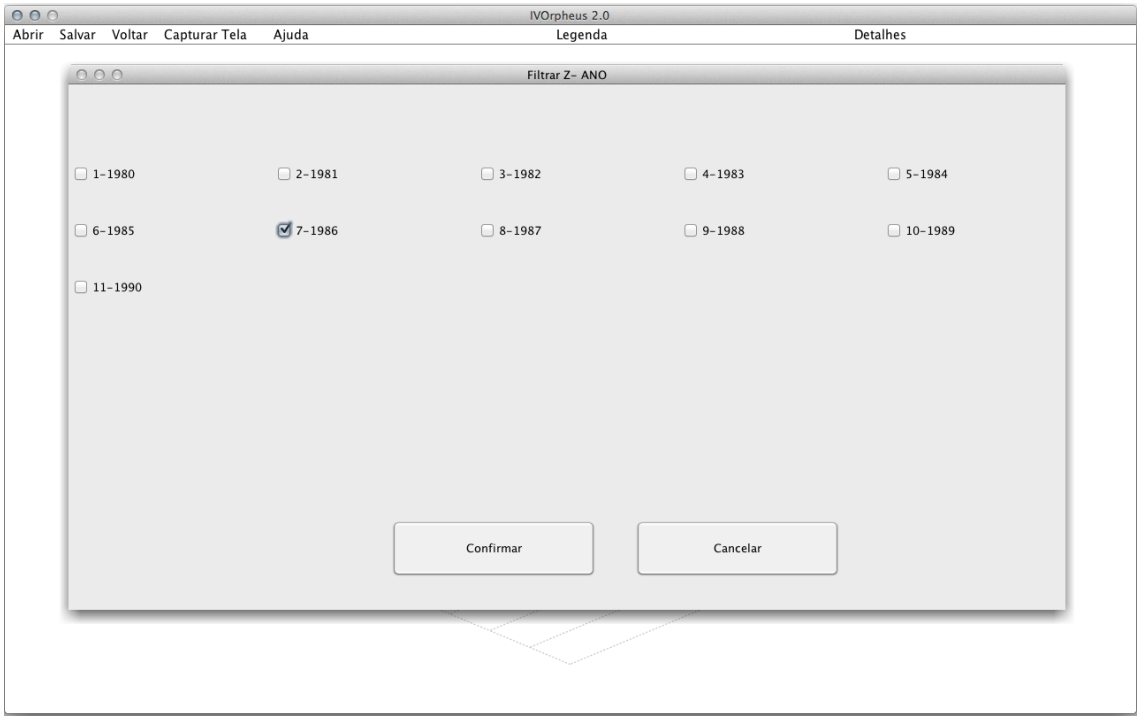

Figura 4.16.Filtrar Z Ano (Categórico).

No momento que a Figura 4.15 e Figura 4.16 são carregadas a gramática "FiltrarCategórico" é chamada, sendo que esta recebe os comandos dos dígitos de um até vinte e dois. Estes podem ser ditos nesta tela para selecionar os valores desejados. Após a aplicação dos filtros os usuários deveriam centralizar na tela o ponto de maior

valor. Para isso, era necessário utilizar o menu Interagir, para aplicar uma rotação e um aumento de escala na tela. Na Figura 4.17 apresenta as opções do menu interagir. Enquanto que na Figura 4.18 e Figura 4.19 são demonstradas as telas de Girar e Escala.

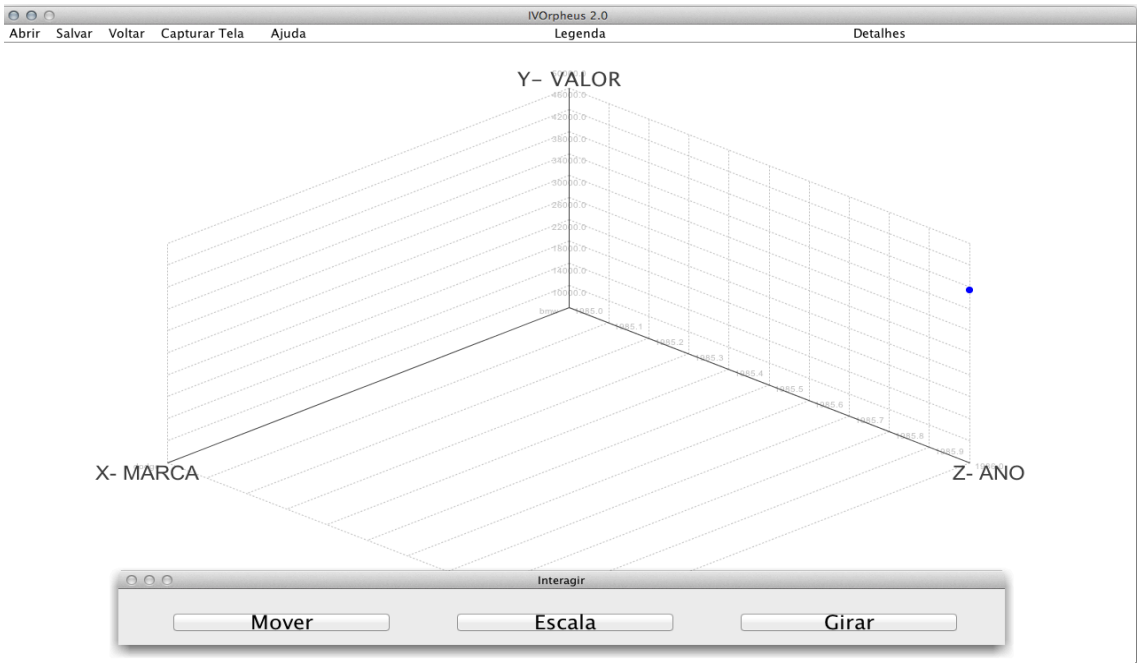

Figura 4.17.Opções do menu Interagir.

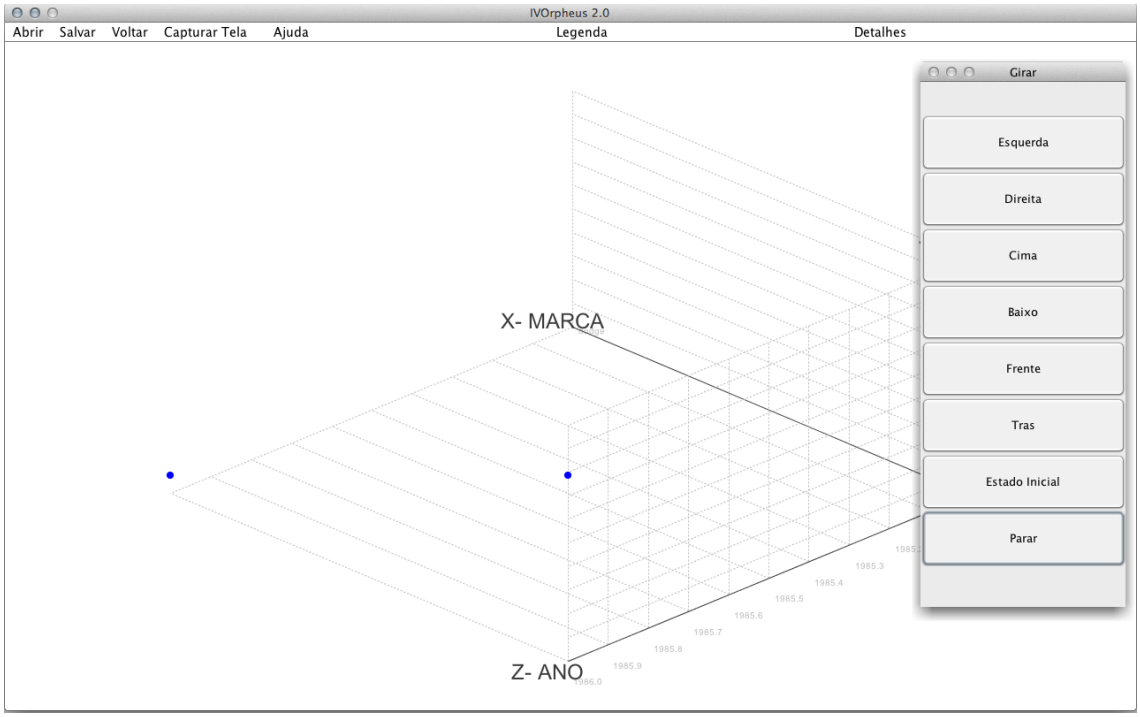

Figura 4.18.Girar.

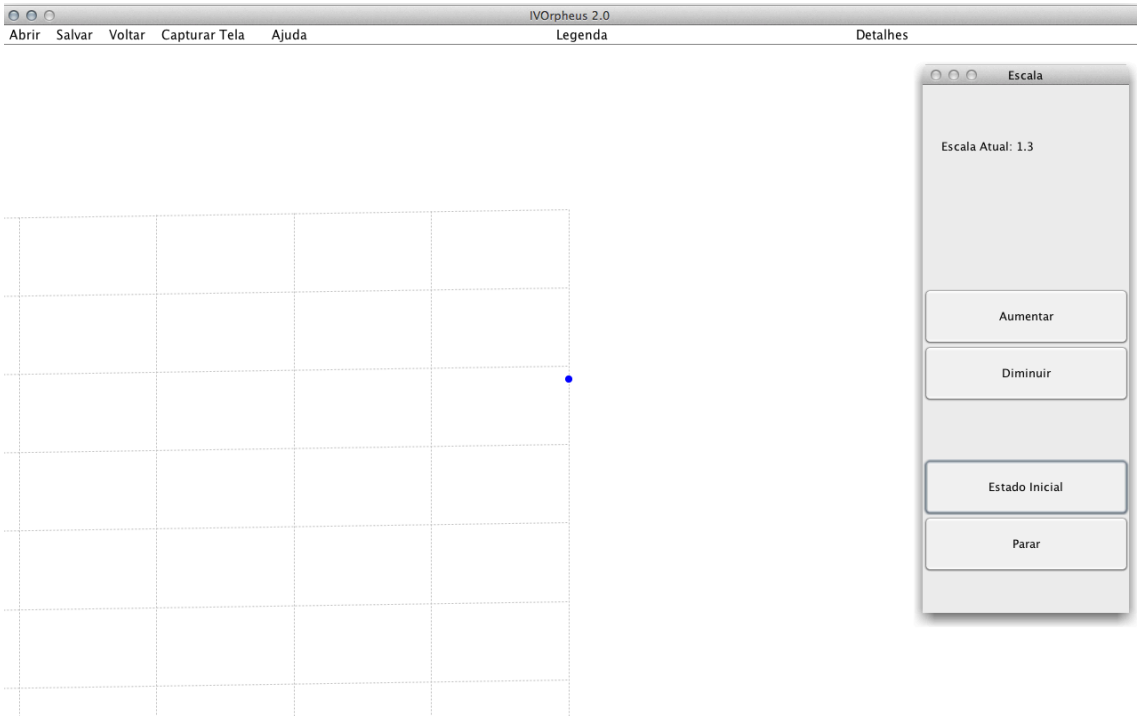

Figura 4.19.Escala.

Para a tarefa 2 o usuário deveria configurar os canais visuais Cor, Forma e Tamanho para os respectivos atributos, Cilindros, Tipo e Tração. O resultado após a configuração é apresentado na Figura 4.20. Para auxiliar o usuário na leitura dos dados, a opção legenda foi habilitada. Com isso o voluntário deveria aplicar o filtro no eixo Y que está configurado para representar o atributo valor.

Por valor se tratar de dados contínuos, ao filtrar o usuário se deparará com a tela de filtrar contínuos, ver Figura 4.21. Tal tela chama a gramática "FiltrarContínuo" que possui as regras locais, "início", "termino", "Dígitos de 0..9", "ponto", "limpar", "apagar", "confirmar" e "cancelar", as quais são descritas para o usuário na seção intitulada "Comandos de Voz" presente na própria tela. Além disso, a tela informar através das *labels* que se localizam abaixo dos botões inicio e término os valores de início e termino originais do eixo escolhido.

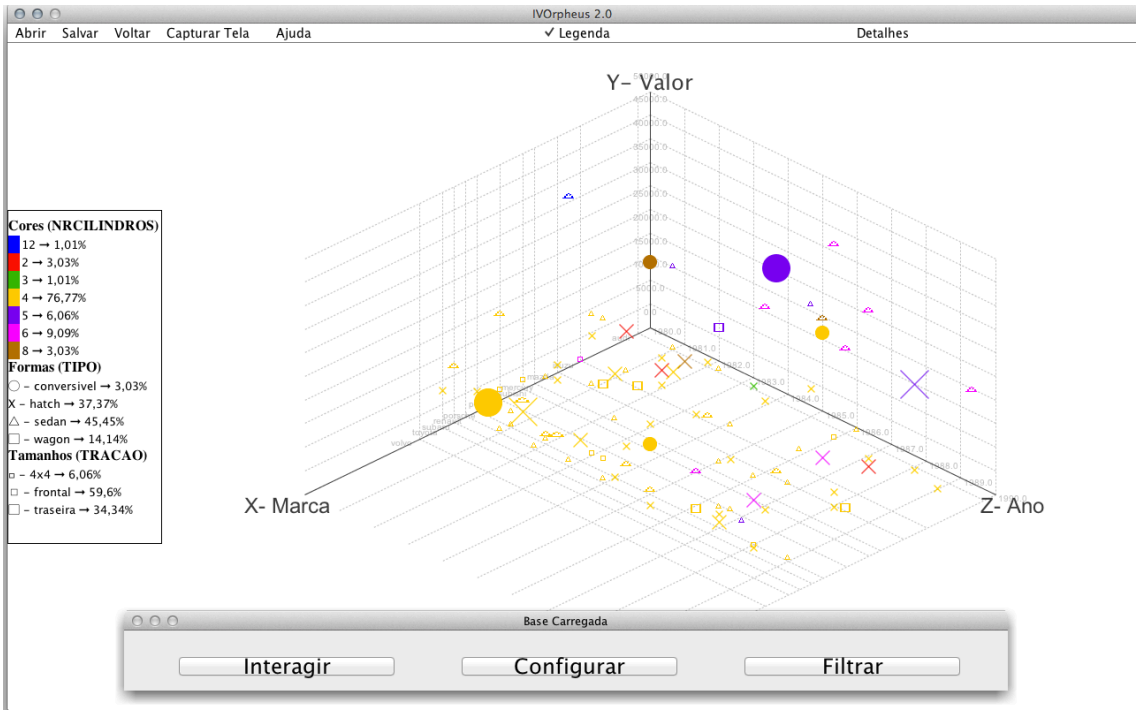

Figura 4.20.Base Configurada com as 6 dimensões mais legenda.

Como a tarefa 2 exige o carro de maior valor no intervalo de $22965 e $34875, o usuário deverá inserir o comando de voz “início” e ditar o início desejado. Por exemplo, o usuário deverá ditar os dígitos: “dois”, “dois”, “nove”, “seis/meia”, “cinco”. O mesmo processo deve ser aplicado no término.

Sabendo que a lógica do filtro contínuo é a seguinte, todos os valores iguais ou abaixo de início e todos os valores iguais ou maiores que o término serão retirado da visualização, deixando apenas os dados no intervalo entre os dois. A Figura 4.22 apresenta o resultado do filtro contínuo.

Também é possível observar nesta tela que a legenda é dinâmica, adaptando-se a visualização disposta na tela.

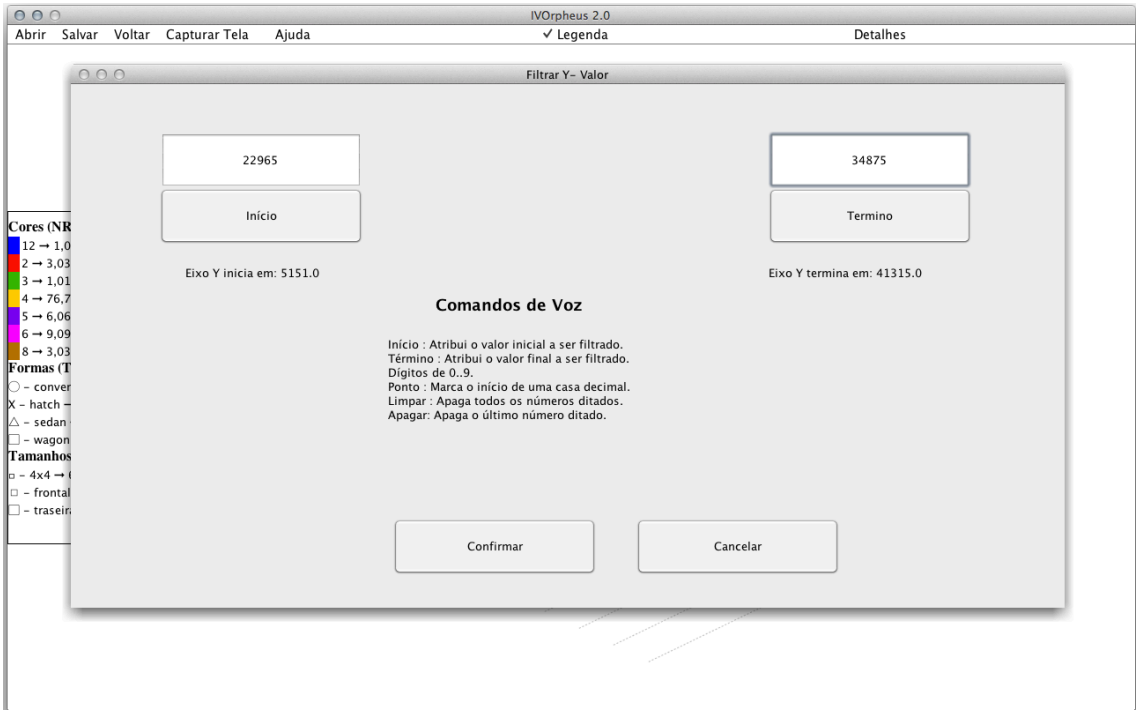

Figura 4.21.Filtrar Y Valor (Contínuo).

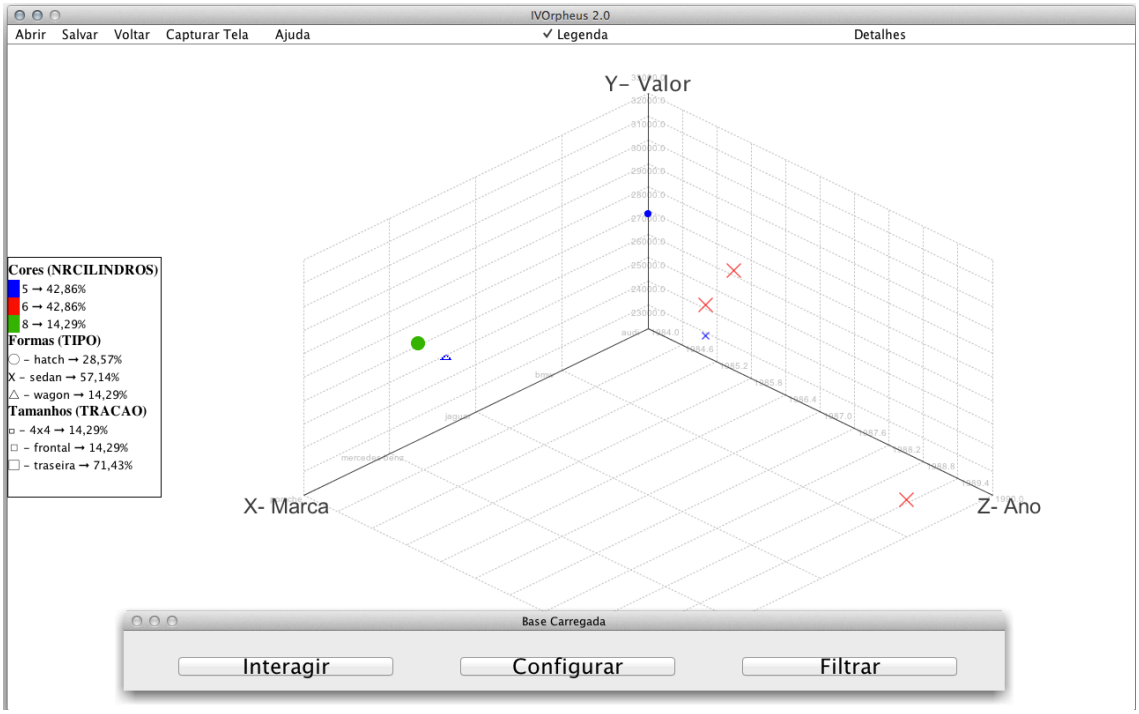

Figura 4.22. Após o filtro Contínuo.

Na Figura 4.23 é apresentado a legenda antes de depois da aplicação do filtro contínuo.

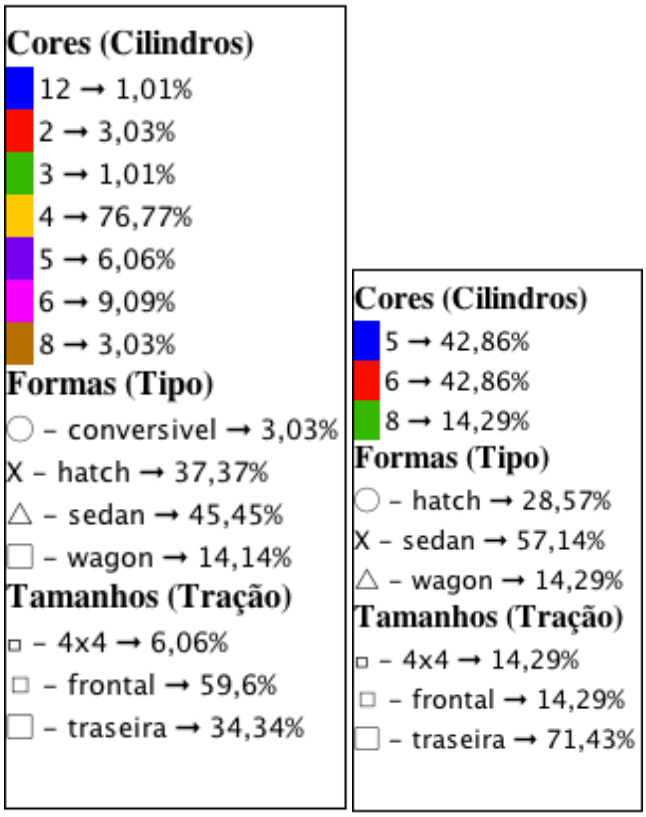

Figura 4.23. Legenda antes do filtro (Esquerda). Legenda após o filtro (Direita).

Como defendido por (Shneiderman, 1996), toda ferramenta de *InfoVis* deve possuir o mínimo de interações, como aponta seu mantra "*Overview first, zoom and filter, then details-on-demand*". Para atender o mantra, foi desenvolvida a funcionalidade detalhes, esta faz uso de todo o mantra. Como exemplo, na terceira tarefa o usuário deveria aplicar detalhes sobre um ponto específico. Então para o usuário completar sua tarefa, o mesmo deveria apertar no botão Detalhes na barra de opções e marcar os atributos Combustível e NRPortas (ver Figura 4.24).

Após apertar em confirmar estes atributos, 8 (oito) quadrantes são desenhados na área de visualização, apresentando a visão geral de dados, ou seja o *overview* do mantra (ver Figura 4.25). Estes quadrantes servem para dividir a visualização em 8 partes para que por meio de voz o usuário possa escolher um destes quadrantes e, ao escolher, o sistema aplique simultaneamente o zoom e o filtro na área desejada, como pode ser visto na Figura 4.26, onde o usuário seleciona o terceiro quadrante.

Após o usuário selecionar mais uma vez o terceiro quadrante, o mesmo é levado à Figura 4.27 e devido a quantidade de pontos na tela ser menor que 10, ao invés do sistema desenhar os quadrantes, o sistema enumera os pontos para que o usuário possa inserir um comando de voz de 0.9 para selecionar o ponto desejado e apresentar os detalhes sobre demanda para aquele ponto.

Como pode ser visto na Figura 4.28 o usuário selecionou o ponto de número 5. Assim, o sistema desenhou as coordenadas e os detalhes daquele ponto. Para finalizar a tarefa o usuário deveria escrever o número de portas (4) e combustível do carro (Gasolina).

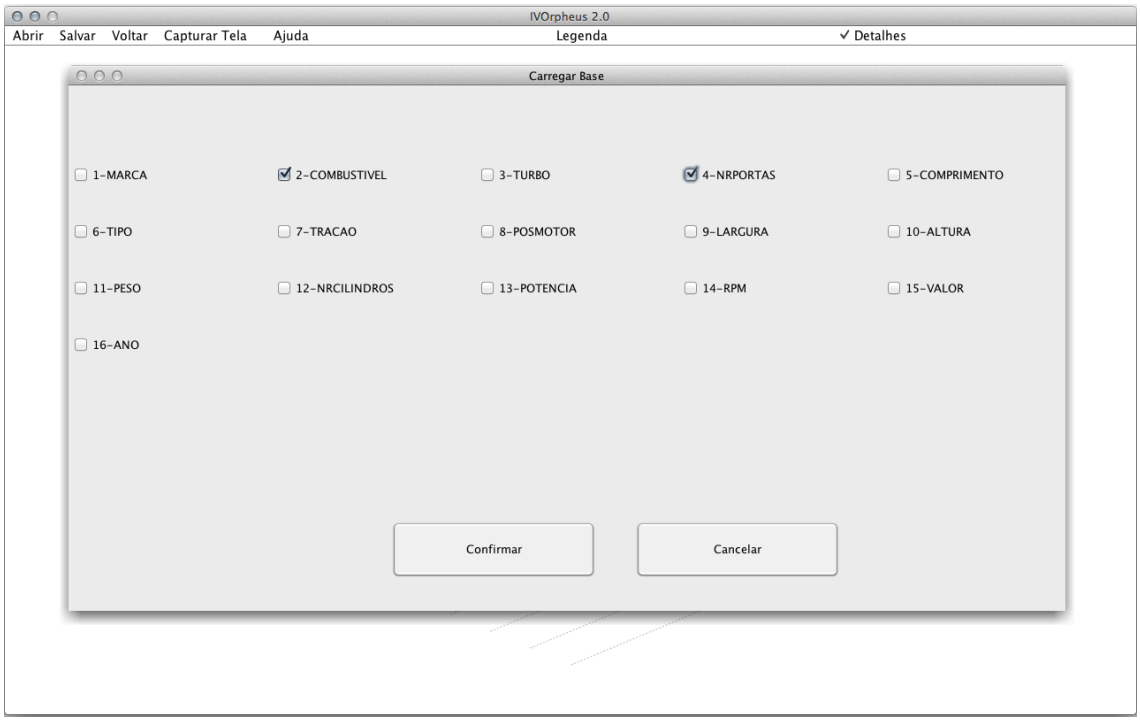

Figura 4.24. Selecionado os atributos para os detalhes sobre demanda.

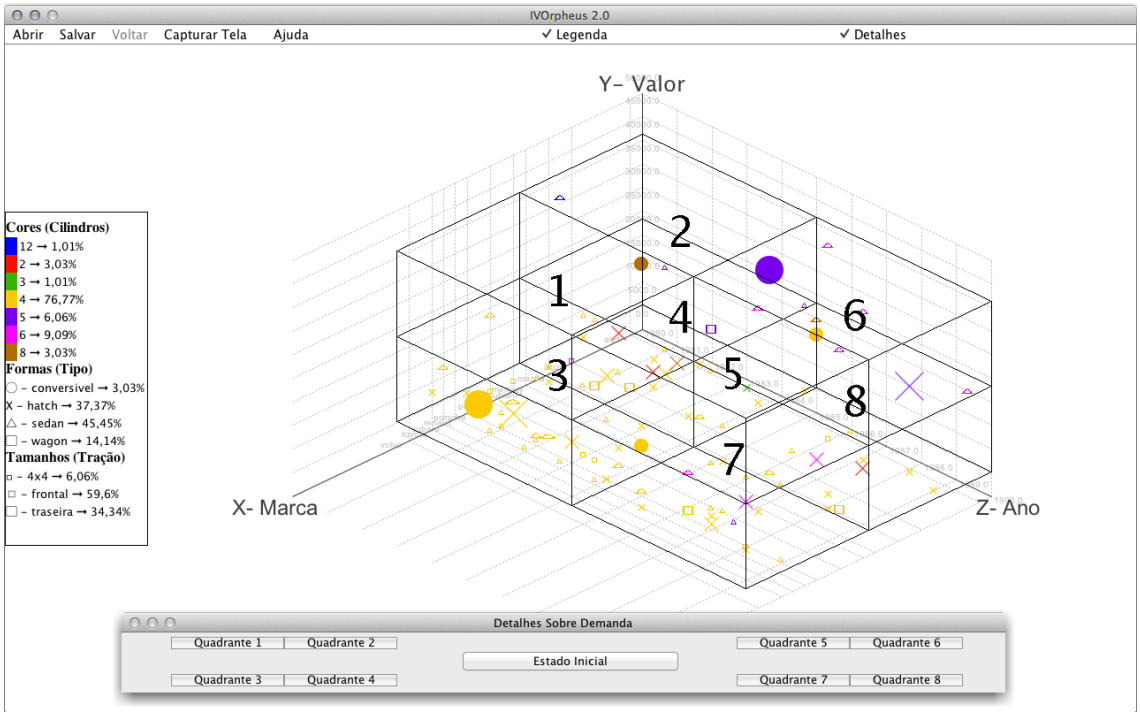

Figura 4.25. Detalhes sobre demanda quadrantes primeiro nível.

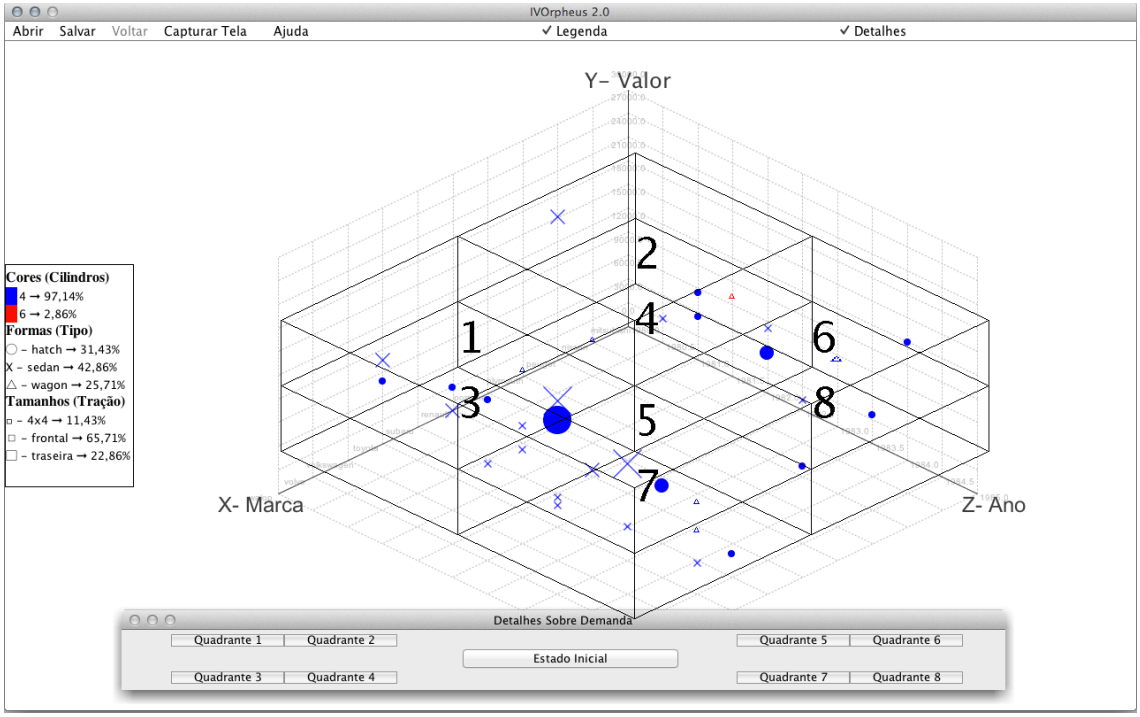

Figura 4.26. Detalhes sobre demanda quadrantes segundo nível.

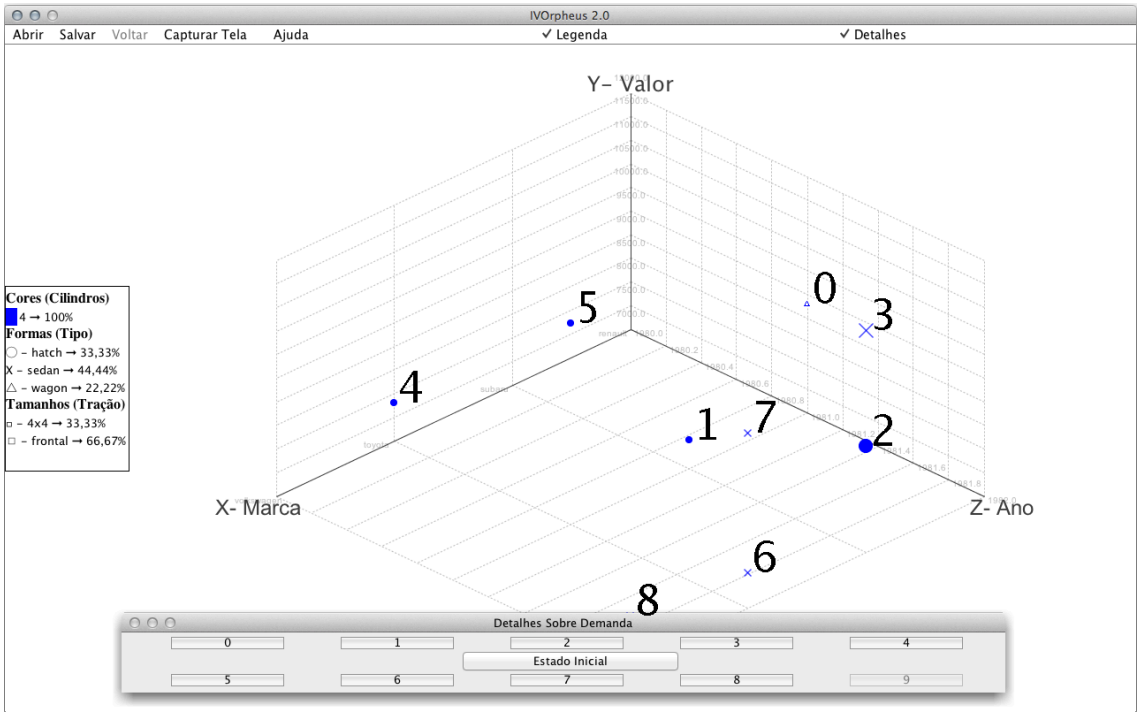

Figura 4.27. Ponto enumerados.

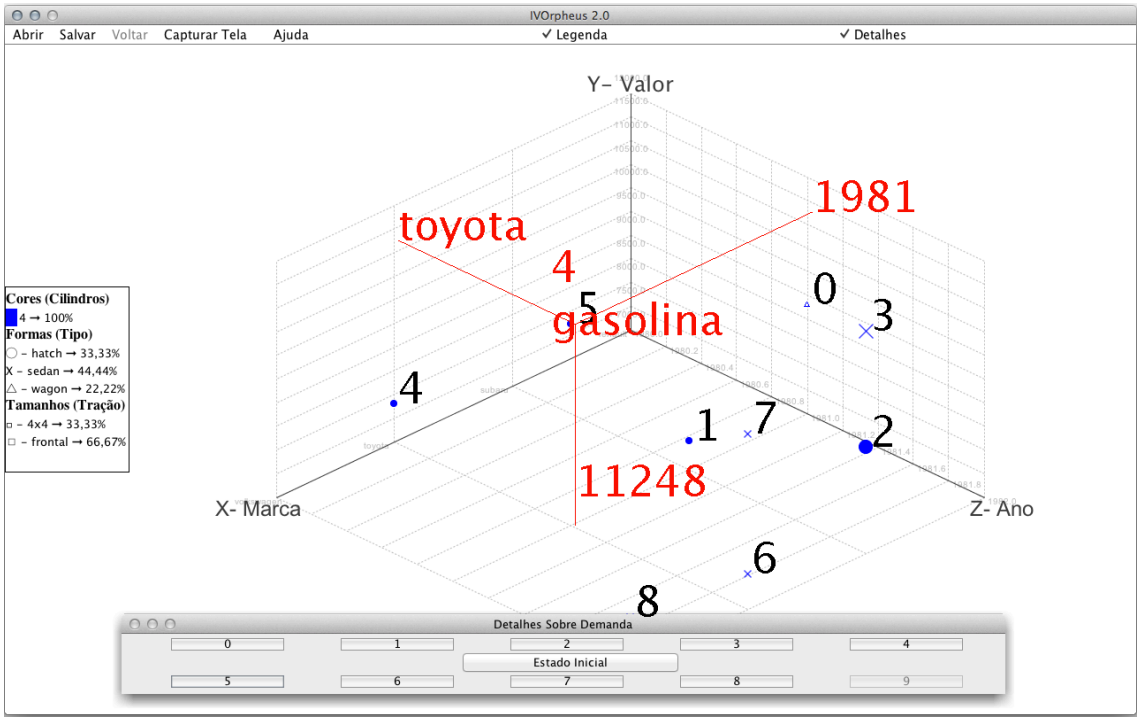

Figura 4.28. Detalhes Sobre Demanda.

4.5. Gerenciamento de Gramáticas

Todas as palavras que vão ser reconhecidas pelo Coruja estão organizadas em gramáticas. Sendo estas um conjunto de regras, utilizadas no reconhecimento de voz.

As gramáticas foram divididas em duas categorias, as dinâmicas e as estáticas.

As dinâmicas são as gramáticas Carregar, Atributos, e Filtrar Categórico. Estas tem a característica de carregar as regras a serem reconhecidas dinamicamente. A exemplo, na gramática Carregar, o diretório raiz do programa é lido a procura dos nomes das bases de dados, sendo estes escritos dinamicamente em forma de regras na gramática Carregar. Enquanto que na gramática Atributos, é escrito os atributos da base escolhida pelo usuário. Estes vão ser utilizados para configurar a visualização. E na gramática Filtrar Categórico os valores dos atributos são escritos nessa gramática para que o usuário possa aplicar o filtro categórico na base de dados.

As demais gramáticas são estáticas, o que implica que todas tem o número constante de regras desde o início da aplicação. Todas as gramáticas tanto dinâmicas e estáticas possuem regras globais e locais. As regras são as palavras que podem ser reconhecidas pelo ASR, e quando são de nível global, estão presentes em todas as gramáticas. Sendo as regras globais as seguintes, “Abrir”, “Salvar”, “Voltar”, “Capturar Tela”, “Ajuda”, “Legenda” e “Detalhes”. Enquanto que as regras locais são as palavras específicas de cada gramática estas serão apresentadas após a Figura 4.7 que tem como objetivo apresentar de forma geral a organização das gramáticas.

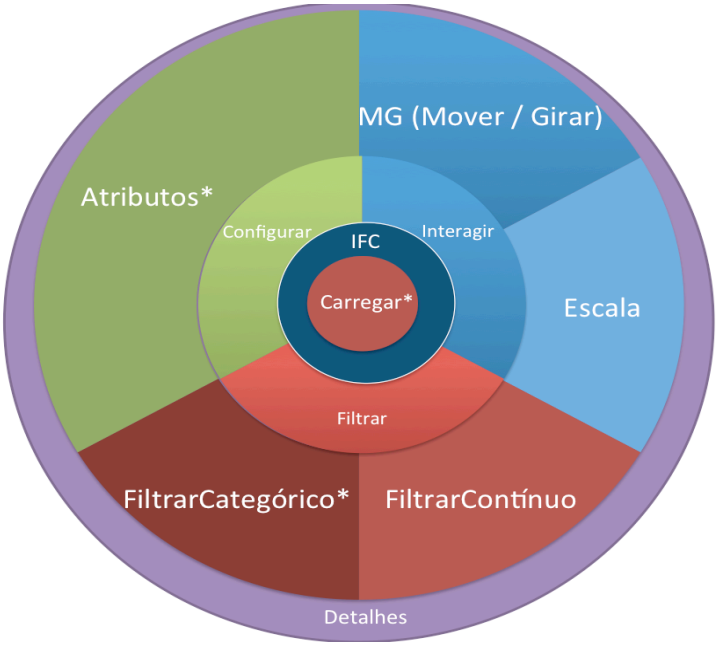

Figura 4.29. Organização das gramáticas. O asterisco representa as gramáticas dinâmicas.

- Carregar: Esta gramática inicia em conjunto com a aplicação, e é responsável por conter os comandos iniciais, sendo também uma gramática dinâmica, isso

quer dizer que o componente *Directory* retorna os nomes das bases de dados, que estão armazenados no diretório raiz da aplicação, para o componente *Recognizer* escrever estes dados na gramática Carregar. Além disto, a gramática Carregar Possui as regras locais de “Confirmar”, “Cancelar”;

- IFC: Acrônimo para Interagir, Filtrar e Configurar. É uma gramática estática que tem suas regras locais homônimas. Que acionam as gramáticas de Interagir e FC;
- Interagir: Tem as regras locais, “Mover”, “Girar” e “Configurar”;
- MG: Acrônimo para Mover e Girar. Tal gramática é utilizada para aplicar translação ou rotação na visualização de informação de dispersão de pontos em 3D, sendo composta pelas regras locais de mover e girar nos sentidos, “esquerda”, “direita”, “cima”, “baixo”, “frente”, “atrás”, e os comandos “parar” e “estado inicial”.
- Escala: Gramática que possui regras locais de “aumentar”, “diminuir”, “parar” e “estado inicial”. A quantidade de zoom aplicado na visualização quando mínimo permite uma visão geral da visualização, e quando máximo permite com que o usuário tenha o foco sobre um ponto ou conjunto de pontos;
- FC : Acrônimo para Filtrar e Configurar. Serve como intermediária para filtrar ou configurar um atributo na visualização. Tem suas regras locais, “Cor”, “Forma”, “Tamanho”, “Eixo X”, “Eixo Y” e “Eixo Z”;
- Atributos: Gramática dinâmica que age de forma similar ao que foi visto na gramática Carregar, criando suas regras locais a partir dos atributos pertencentes a base de dados escolhida pelo usuário. Além de possuir as regras locais “Confirmar” e “Cancelar”;
- FiltrarContínuo: Gramática estática ativada quando aos atributos a serem filtrados são flutuantes ou inteiros acima de 20 valores único. Com as regras locais, dígitos de 0..9, “início”, “término”, “apagar”, “ponto/virgula”, “confirmar” e “cancelar”;

- FiltrarCategórico: gramática dinâmica que tem em suas regras locais os valores únicos da dimensão espacial ou canal visual escolhido para ser filtrado;
- Detalhes: gramática estática que tem suas regras locais os números de 0..9, e o comando Estado Inicial, que retorna à aplicação ao estado anterior à aplicação dos filtros por quadrante.

CAPÍTULO 5- TESTES DE USABILIDADE

Neste capítulo, é apresentado o teste com os usuários, o aparato utilizado no teste, participantes e seus perfis, procedimento para realização do teste e a análise dos resultados obtidos.

5.1 Aparato

Para os testes com o usuário foi utilizado um Macbook Pro, 13 polegadas, com resolução de 1280x800 pixels, processador 2,4 GHz Intel Core i5, memoria RAM de 4 GB 1333 MHz DDR3, sendo o processador gráfico Intel HD Graphics 3000 384 MB e o sistema operacional Mac OS X Lion 10.7.5 e Ubuntu 14.04 32 bits.

5.2 Plano De Teste

Foi seguido o plano de teste baseado no trabalho de Rocha (2003). Tal plano de teste foi utilizado devido sua simplicidade, confiabilidade do autor e por apresentar questões relevantes as necessidades presentes no teste com os usuários proposto neste trabalho. Abaixo são apresentados os questionamentos e as respostas levantadas no plano de teste.

5.2.1 O objetivo do teste: o que se deseja obter?

O objetivo do teste é observar a carga de trabalho, o tempo despendido na utilização de uma interface de interação natural (Interface por Voz) e as diferentes dificuldades apresentadas pelos usuários em cada tarefa apresentada. E assim, poder analisar a funcionalidade, eficácia, eficiência, usabilidade e utilidade da ferramenta com o meio de entrada convensional e com o meio de entrada por voz.

5.2.2 Onde o teste irá acontecer?

O teste será realizado em um laboratório com poucos ruídos externos, para assim o reconhecedor automático de voz, ou seja o Software Coruja não venha sofrer interferências externas durante a execução dos mesmos.

5.2.3 Qual a duração prevista de cada sessão de teste?

O tempo de duração para cada seção de teste por usuário é de 30 minutos. Sendo que este tempo esta dividido em 18 minutos para a execução das tarefas, 7 minutos para o treinamento e 5 minutos para o preenchimento dos questionários pré e pós tarefa.

5.2.4 Qual software vai ser utilizado nos testes?

O software que será utilizado nos testes vai ser a ferramenta de visualização de informação, IVOrpheus.

5.2.5 Quais tarefas serão apresentadas nos teste?

As tarefas que serão apresentadas nos testes seguem abaixo:

- Configurar as dimensões espaciais;
- Filtrar valores categóricos;
- Aumentar escala e/ou aplicar um giro na visualização;
- Configurar canais visuais;
- Filtrar valores contínuos;
- Interpretar a visualização;
- Ler legenda;
- Detalhes sobre demanda.

5.2.6 Quem serão os usuários e quantos serão necessários?

Os usuários que irão participar nos testes terão como requisito necessário, o conhecimento básico em como interagir com um computador. Enquanto ao número de usuários será utilizado cinco para cada meio de interação proposto, ou seja, cinco usuários para interagir por mouse/teclado e cinco usuários para interagir por voz. Tendo como base o trabalho de (Nielsen, 2000), que afirma que o número de pessoas a executar um teste não deve ser superior a cinco.

Segundo a teoria de Nielsen, os vários dados gerados durante a execução do primeiro teste, se repetirão nos demais testes. Assim cada teste apresentará uma contribuição menor e por isso não se faz necessário uma quantidade de usuários maior que cinco.

5.2.7 Quando o avaliador poderá ajudar o usuário durante o teste?

Como o que está sendo testado é interação e não o conhecimento do público alvo na área de visualização de informação. O avaliador poderá prestar auxílio caso o usuário apresente dificuldades recorrentes sobre a visualização.

5.2.8 Quais dados serão coletados?

Os dados que serão coletados segue abaixo:

- Tempo: Os testes serão filmados visando identificar o tempo dispensado na execução de cada uma das tarefas e também será verificado a quantidade de palavras reconhecidas e as não reconhecidas pelo ASR nestas tarefas;
- Nível de dificuldade: Após a realização das tarefas o usuário passará por um questinário que utiliza a escala Likert (Likert , 1932), com intuito do usuário acusar os níveis de dificuldades presenciados em cada tarefa;
- Carga de Trabalho: Ao final dos testes o usuário será submetido ao NASA-TLX, onde apontará sua impressão subjetiva sobre as diferentes cargas de trabalho apresentadas no NASA-TLX.

5.2.9 Qual o critério para determinar o sucesso da interface?

O sucesso da interface será determinado ao não encontrar nenhum problema de usabilidade com grau de severidade elevado (acima de 3).

Serão avaliados três fatores para determinar o grau de severidade, sendo eles: impacto, frequência e persistência (Rocha, 2003).

Impacto, o usuário consegue continuar a usar o programa sem maiores dificuldades após entrar em contato com o problema. Frequência, quantidade de vezes que o problema é apresentado durante a execução das tarefas. E por fim, persistência, após o usuário encontrar o problema verificar se o mesmo persiste entre as telas da aplicação.

5.3 Procedimentos

Os participantes foram apresentados a ferramenta IVOrpheus através de um vídeo de treinamento e após isso, foi aplicado um questionário para identificação do perfil do usuário. Por conseguinte, foi dado início as tarefas, onde o tempo do usuário era cronometrado e seu desempenho gravado. Após o término das tarefas era administrado o questionário para identificar o nível de dificuldade em cada cenário. E, por fim, o questionário pós-tarefa, NASA-TLX, era apresentado ao usuário para que a carga de trabalho do mesmo fosse capturada.

5.4 Perfis dos Usuários

Para identificar os diferentes perfis de usuário, foi aplicado o questionário pré-tarefa. Com as seguintes perguntas:

1- Você conhece o plano/espaço Cartesiano para 2 e 3 Dimensões?

2- Você já utilizou algum software com objetivo de analisar uma tabela de dados, a exemplo, Excel® (Microsoft), Numbers® (Apple) ou outros?

3- Você faz uso ou já fez uso de algum aplicativo no seu celular ou computador que utilize a voz como meio de entrada, por exemplo, aplicativos de assistência pessoal, como, a SIRI® da Apple ou o S Voice® da Samsung? Caso sim, qual a frequência (raramente ou ocasionalmente ou frequentemente)?

4- Você tem familiaridade com aplicativos que utilizem as três dimensões espaciais na locomoção do ponto de vista do usuário, ou seja, da câmera, como Blender®, Zbrush®, Autodesk 3D Max®, entre outros?

Abaixo é apresentado na tabela I, o resultado do questionário tendo as cinco primeiras colunas representando aos usuários de mouse e as demais os usuários que utilizaram a interface de interação natural.

Tabela I. Questionário para identificar os perfis de usuários. Legenda, R- raramente, O- ocasionalmente e F- frequentemente.

Perguntas	**Usuário 1(Mouse)**	**Usuário 2(Mouse)**	**Usuário 3(Mouse)**	**Usuário 4(Mouse)**	**Usuário 5(Mouse)**	**Usuário 1(Voz)**	**Usuário 2(Voz)**	**Usuário 3(Voz)**	**Usuário 4(Voz)**	**Usuário 5(Voz)**
1	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim
2	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
3	Sim(R)	Sim(O)	Sim(R)	Sim(O)	Não	Sim(R)	Sim(R)	Sim(O)	Sim(R)	Sim(F)
4	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Sim

Na tabela I, pode ser observado que devido à popularização dos assistentes pessoais nos mais diversos sistemas abarcados, os usuários tiveram um contato prévio com a tecnologia de interação por voz. Mesmo assim, a frequência de utilização dos softwares por voz ainda é um meio de interação secundário, como pode ser visto através da frequência de uso dos usuários, que em sua maioria faz uso raramente desta tecnologia. Enquanto a utilização de ambientes tridimensionais é notória a subutilização

desta abordagem, devido ao fato da maioria dos softwares e aplicativos fazerem uso de duas dimensões.

5.5 Vídeo de Treinamento

Foi apresentado a todos os voluntários que realizaram os testes um vídeo de treinamento com duração de 7 minutos, o qual apresentava os seguintes pontos, introdução a ferramenta de visualização IVOrpheus, mostrando a interface e suas funcionalidades.

Logo após, era apresentado um exemplo de como abrir uma base, e como configurar as dimensões espaciais (eixos x, y e z) e os canais visuais (cor, forma e tamanho). Em seguida os usuários eram apresentados à utilização de filtros categóricos e contínuos. Por fim, era mostrado ao usuário como utilizar os detalhes sobre demanda.

5.6 Tarefas

No teste foi utilizada uma base de dados sobre carros da década de 80, que continham 789 registros e 16 atributos (7 contínuos e 9 discretos). Utilizando esta base foram apresentadas as seguintes tarefas aos usuários:

1- Configure os eixos X para Marca ,Y para Valor e Z para Ano. E encontre o carro de maior valor no ano de 1986 entre as marcas BMW, Isuzu e Dodge. Após encontrá-lo, aumente a escala para dar zoom no mesmo, o centralizando na tela.

2- Você começará com uma base configurada com os seguintes atributos nos eixos X - Marca, Y - Valor e Z - Ano. Com isso, Configure a Cor para NRCilindros, Forma –para Tipo e Tamanho para Tração. Após isso, descubra qual é o carro de maior valor no intervalo de valor de $22965 e $34875 e utilizando a legenda escreva a cilindrada, tipo e tração do mesmo.

3- Selecione o carro com as seguintes dimensões X - Marca (Toyota), Y - Valor (11248), Z - Ano (1981), Cor - NRCilindros (4), Forma – Tipo (Hatch) e Tamanho - tração (4x4). Utilize a técnica de detalhes sobre demandas para selecionar o ponto desejado, apresentando o número de portas e o combustível do carro citado.

5.7 Tempo

O usuário tinha a seu dispor 18 minutos para cumprir todas as tarefas apresentadas na lista de tarefas. Sendo que este tempo foi definido, a partir do tempo despendido pelo avaliador na execução das tarefas pelo meio de interação por voz em um pré teste. O avaliador tinha o perfil desejado para os testes, pois o mesmo tinha familiaridade com interação por voz e com a área de visualização de informação.

Sendo que o avaliador teve o tempo de 2 minutos e 05 segundos na 1˚ Tarefa, 2 minutos e 51 segundos na 2˚ Tarefa e 1 minuto e 02 segundos na 3˚ Tarefa. Somando um tempo total de 5 minutos e 58 segundos. Após o pré teste, o tempo total do avaliador fui utilizado como base para determinar o tempo dos testes com os usuários. Para isso, o tempo total do teste com o avaliador foi multiplicador pelo fator de folga (3) e arredondado. Assim dando um tempo total de 18 minutos para a execução dos testes com os usuários.

Apesar do trabalho fazer a abordagem de realizar testes prévios com o avaliador para determinar o tempo dos teste com o usuário. É notório que uma prática recomendável para estabelecer o tempo de teste dos usuários é realizar um ou mais testes pilotos onde o tempo dos usuários serão cronometrados. Assim com os resultados obtidos é aplicado uma folga no tempo médio para determinar o tempo total do teste com os usuários.

Para medir quantitativamente o desempenho dos usuários, os seus testes eram cronometrados. E caso o usuário não terminasse no tempo proposto as tarefas, as não finalizadas ficariam sem respostas e o voluntário era convidado a preencher os questionários pós-tarefa NASA-TLX e o questionário para identificar o nível de dificuldade dos cenários.

A Figura 5.1, ilustra o tempo do grupo de usuários que utilizaram o mouse/teclado para completar a tarefas 1 e a Figura 5.2 apresenta o tempo do grupo de usuários que utilizaram o meio de entrada por voz para completar a tarefa 2. Enquanto as Figuras 5.3 e 5.4 ilustram respectivamente o tempo dos usuários de mouse e dos usuários de voz para completar a tarefa 2. E, por fim, as Figuras 5.5 e 5.6 ilustram o tempo para completar a tarefa 3 e o tempo total do teste.

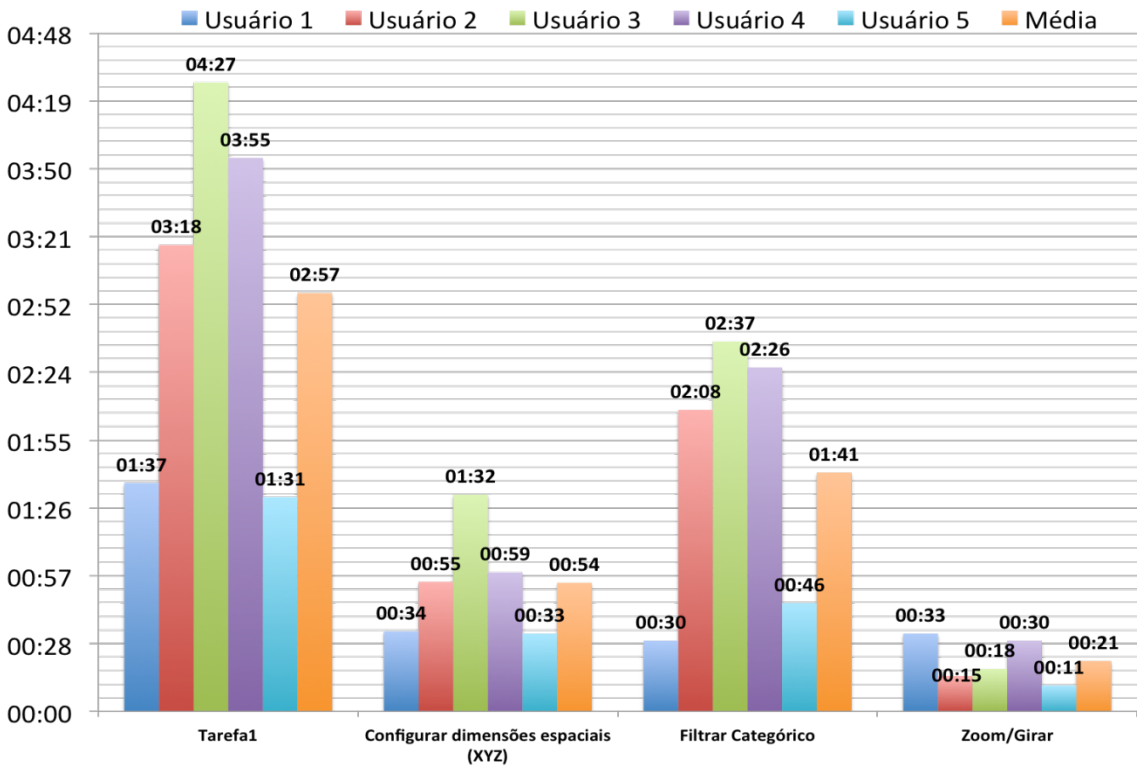

Figura 5.1. Tempo da tarefa 1 e suas sub-tarefas (Mouse).

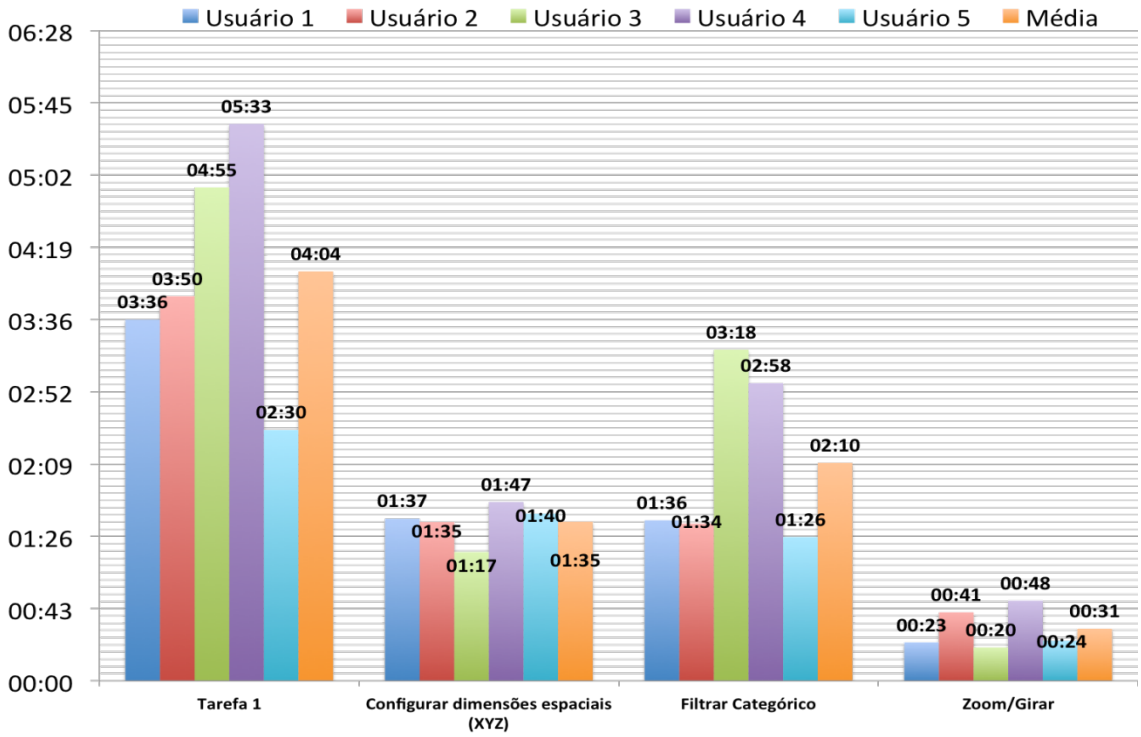

Figura 5.2. Tempo da tarefa 1 e suas sub-tarefas (Voz).

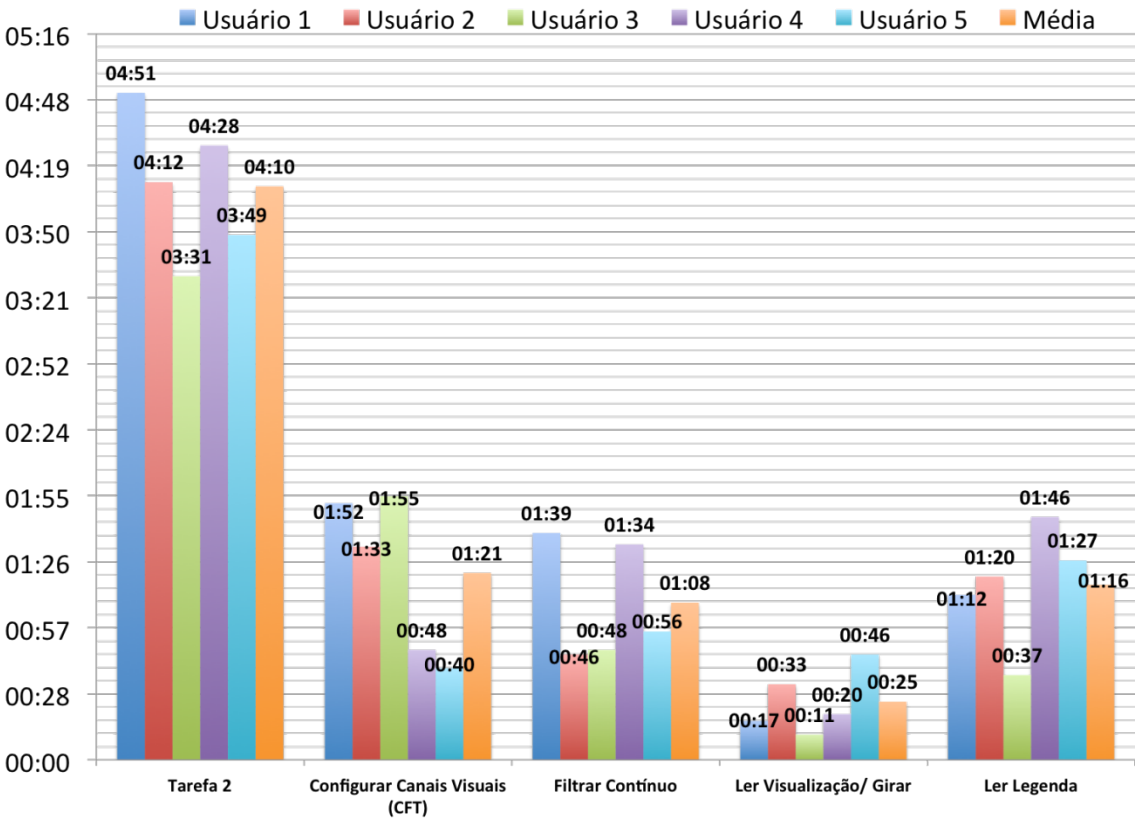

Figura 5.3. Tempo da tarefa 2 e suas sub-tarefas (Mouse).

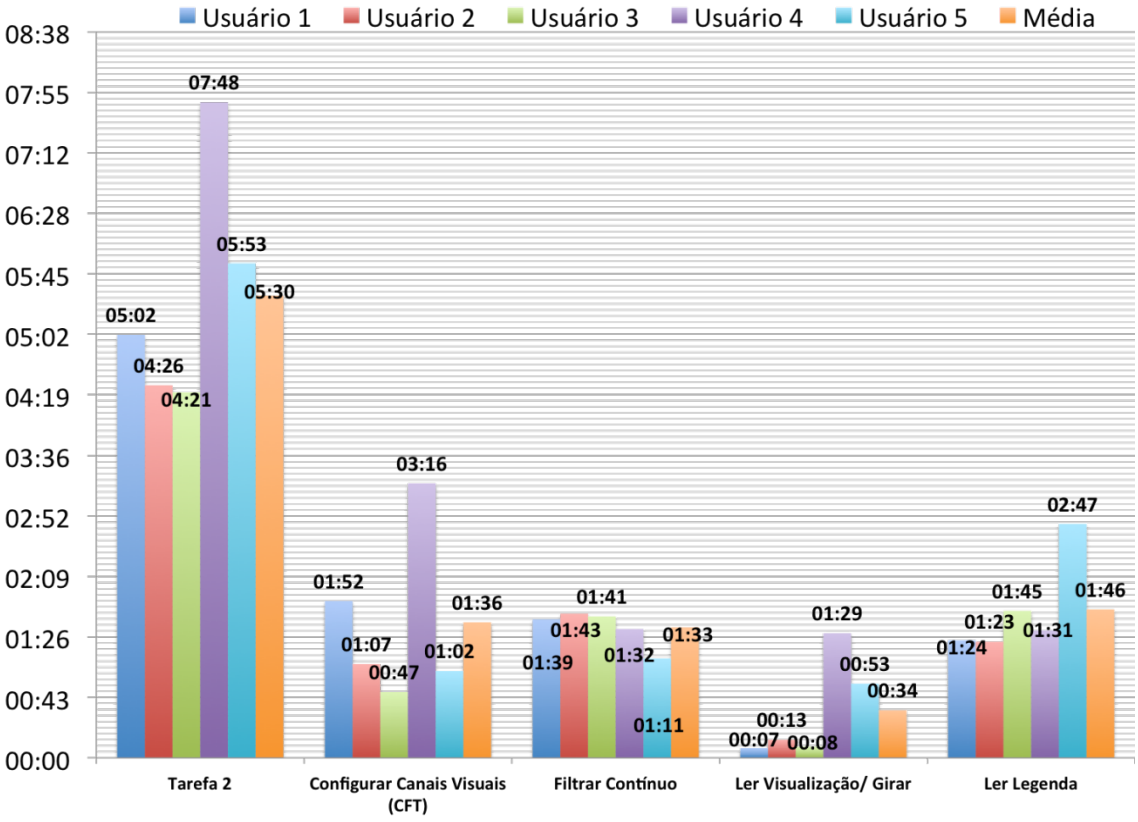

Figura 5.4. Tempo da tarefa 2 e suas sub-tarefas (Voz).

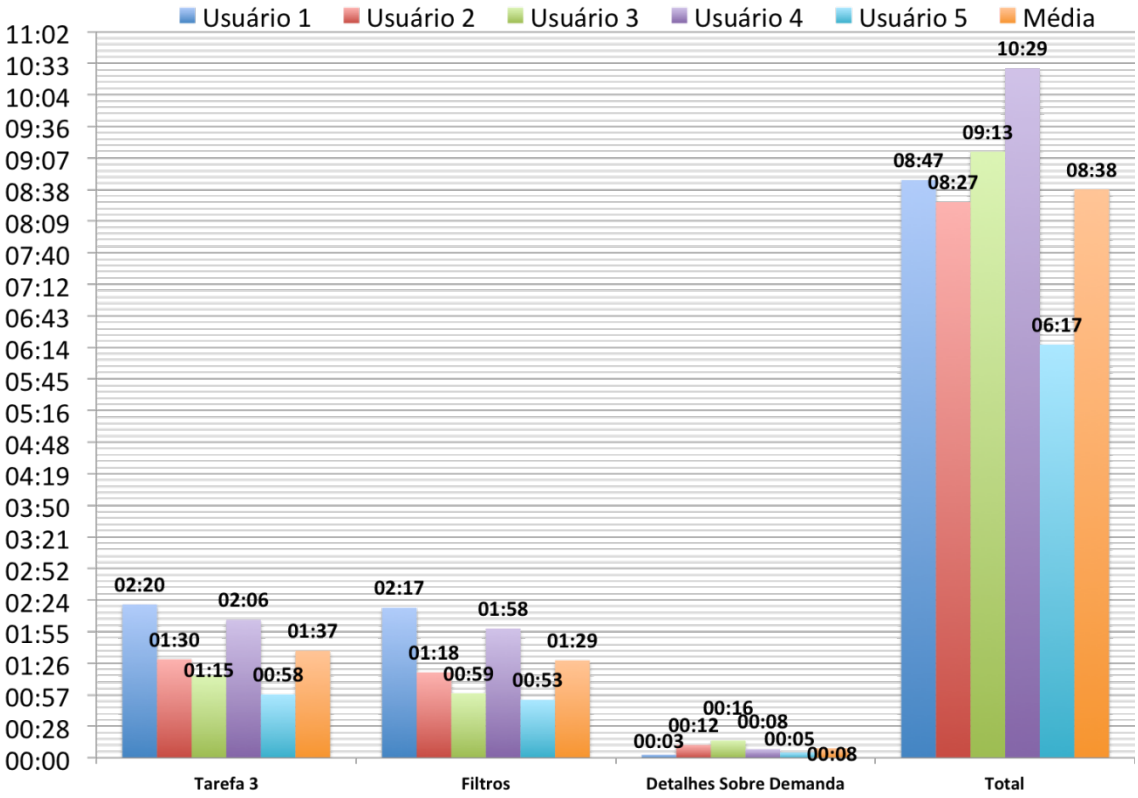

Figura 5.5. Tempo da tarefa 3 e suas sub-tarefas mais o tempo total do teste (Mouse).

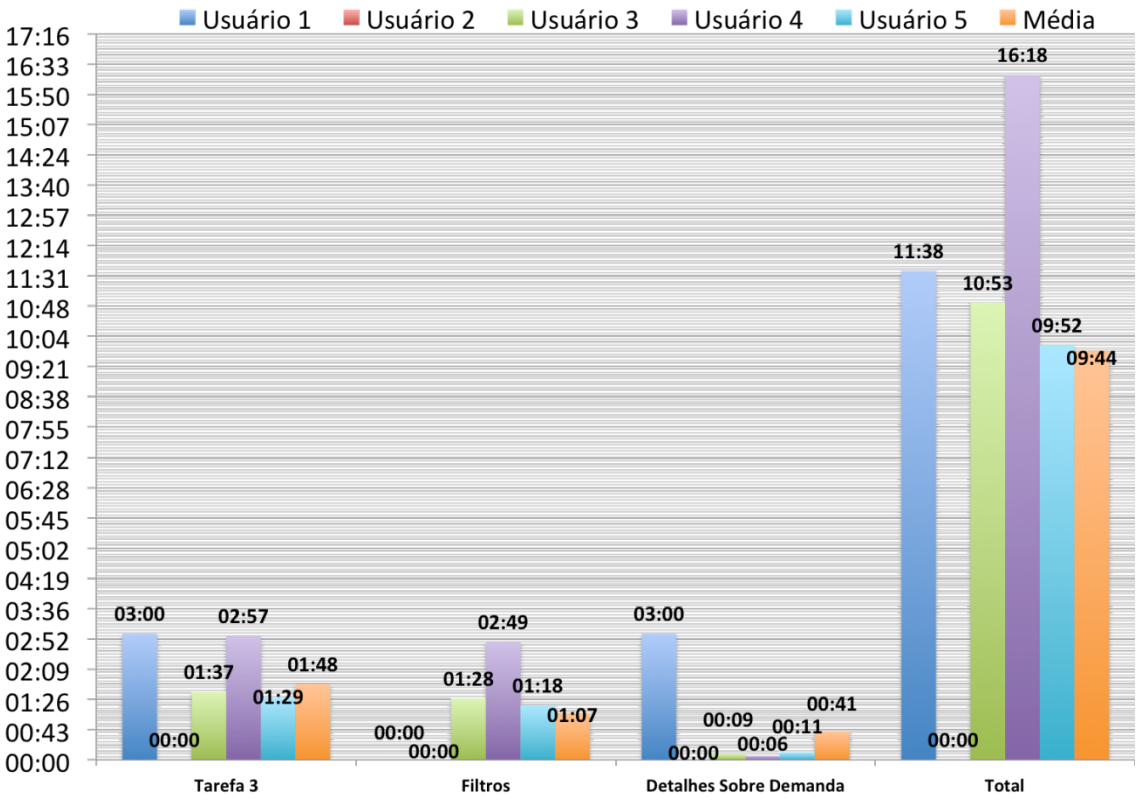

Figura 5.6. Tempo da tarefa 3 e suas sub-tarefas mais o tempo total do teste (Voz).

5.8 Questionário para Identificar o Nível de Dificuldade Subjetiva Durante a Execução das Tarefas e Sub-Tarefas

Após a completude das tarefas, os usuários receberam o questionário pós-tarefa para identificar o nível de dificuldade subjetiva durante a execução das tarefas e sub-tarefas (Cenários). Este questionário consistia de uma escala Likert (Likert, 1932) com cinco níveis, sendo eles: muito fácil, fácil, médio, difícil e muito difícil.

Ao usuário era a apresentado a uma tabela com todas as tarefas e cenários presentes no teste. Para cada tarefa e cenário o usuário deveria acusar um dos cinco níveis da escala Likert que melhor representasse sua impressão subjetiva em relação a tarefa executada. Nas tabelas II e III é apresentado respectivamente os resultados do questionário para o grupo de mouse e para o grupo de voz.

Tabela II. Resultado do questionário pós tarefa para os usuários do Mouse.

Nível de dificuldade	Usuário 1	Usuário 2	Usuário 3	Usuário 4	Usuário 5	Média
Tarefa 1	Fácil	Fácil	Médio	Fácil	Difícil	Médio
Configurar Dimensões Espaciais (XYZ).	Fácil	Fácil	Médio	Muito Fácil	Médio	Fácil
Filtrar Categórico	Médio	Fácil	Médio	Muito Fácil	Médio	Fácil
Zoom/Girar	Fácil	Muito Fácil	Médio	Médio	Muito Fácil	Fácil
Tarefa 2	Difícil	Médio	Médio	Difícil	Muito Fácil	Médio
Configurar Canais Visuais (CFT).	Médio	Médio	Médio	Muito Fácil	Muito Fácil	Fácil
Filtrar Contínuo.	Difícil	Fácil	Médio	Muito Fácil	Muito Fácil	Fácil
Ler Legenda	Muito Fácil	Médio	Médio	Muito Difícil	Muito Fácil	Fácil
Tarefa 3	Médio	Fácil	Médio	Muito Fácil	Muito Fácil	Fácil
Detalhes Sobre Demanda	Médio	Fácil	Médio	Muito Fácil	Muito Fácil	Fácil

Tabela III. Resultado do questionário pós tarefa para os usuários da Voz.

Nível de dificuldade	Usuário 1	Usuário 2	Usuário 3	Usuário 4	Usuário 5	Média
Tarefa 1	Fácil	Médio	Fácil	Fácil	Fácil	Fácil
Configurar Dimensões Espaciais (XYZ).	Fácil	Muito Fácil	Fácil	Fácil	Fácil	Fácil
Filtrar Categórico	Fácil	Médio	Fácil	Fácil	Fácil	Fácil
Zoom/Girar	Fácil	Muito Fácil	Médio	Fácil	Fácil	Fácil
Tarefa 2	Médio	Difícil	Médio	Médio	Médio	Médio
Configurar Canais Visuais (CFT).	Médio	Difícil	Fácil	Fácil	Fácil	Médio
Filtrar Contínuo.	Fácil	Médio	Médio	Fácil	Fácil	Fácil
Ler Legenda	Médio	Difícil	Fácil	Médio	Médio	Médio
Tarefa 3	Difícil	Difícil	Médio	Fácil	Fácil	Médio
Detalhes Sobre Demanda	Difícil	Difícil	Médio	Fácil	Fácil	Médio

5.9 NASA-TLX

A Carga de trabalho subjetiva dos usuários foi avaliada com o NASA Task-Load-Index (Hart et al. 1988), que é utilizado para identificar a carga de trabalho global nas diferentes tarefas e as principais fontes de carga de trabalho. A carga de trabalho é definida como uma "construção hipotética que representa o custo incorrido por um operador humano para alcançar um determinado nível de desempenho." O NASA TLX é estimado com seis sub-escalas (demanda física, mental, temporal, esforço, frustração e desempenho).

Cada uma dessas dimensões é apresentada por meio de escala que possui vinte níveis discretos para imprimir a carga de trabalho do usuário (ver Figura 5.7). Sendo que quanto menor o valor de cada escala, menor esforço que aquela carga de trabalho representou e vice e versa (Ver Tabela IV). Logo após o usuário identificar o valor de cada escala, é utilizada uma sequência de 15 perguntas, que apresenta duas escalas e indaga qual contribui mais para a completude das tarefas, com o objetivo de balancear a resposta do usuário, para melhor se aproximar da carga de trabalho sentida pelo voluntário.

É apresentado na tabela V os resultado do NASA TLX para os usuário que utilizaram o mouse e na tabela VI os resultados para os usuário que utilizaram a interface de interação natural. E por fim é apresentado na tabela VII as médias dos dois grupos.

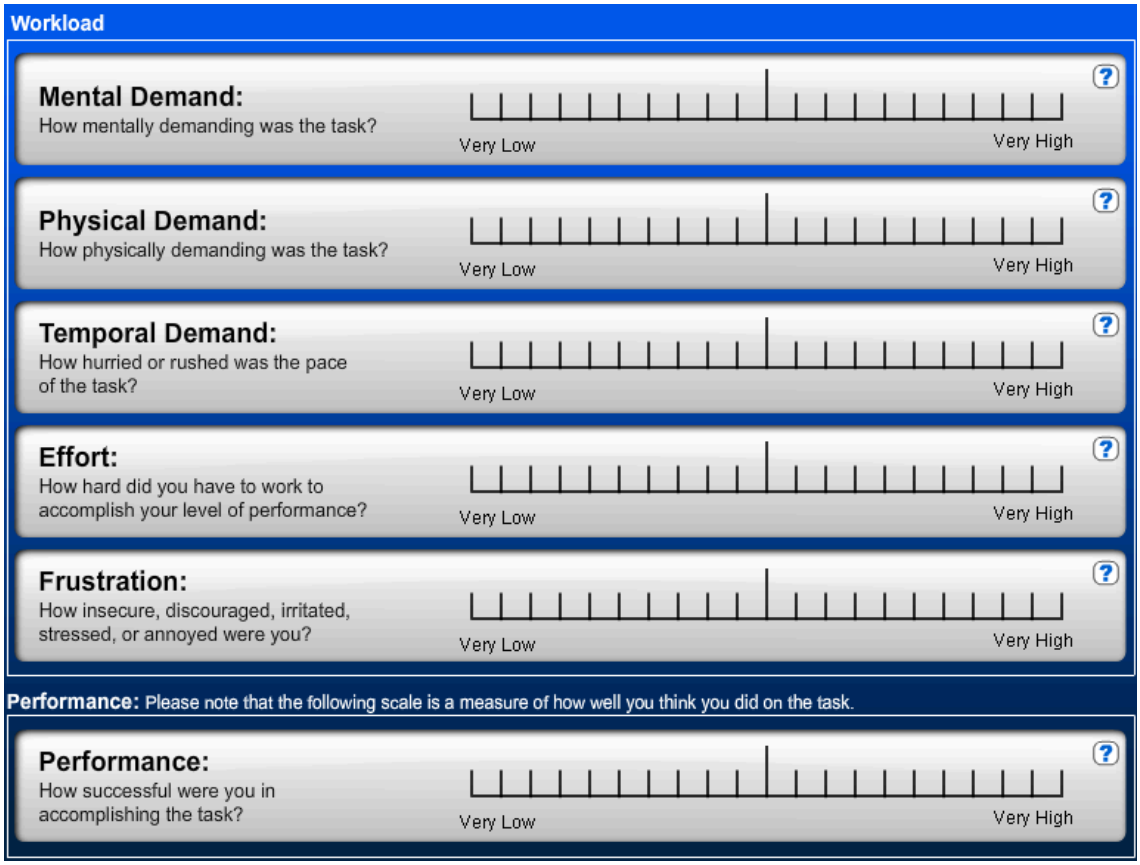

Figura 5.7. Interface do NASA-TLX.

Tabela IV. Escalas consideradas na avaliação do NASA-TLX.

Escalas	**Limite Mínimo**	**Limite Máximo**
Demanda Mental	Tarefas com objetivos alcançados sem dificuldades.	Tarefas exige muito esforço mental para se atingir o objetivo.
Demanda Física	Tarefa leve, lenta, facilmente realizada.	Tarefa pesada, rápida, vigorosa e agitada.
Demanda Temporal	Ritmo de trabalho com baixa pressão exercida para o término das atividades.	Ritmo rápido com muita pressão exercida para o término das atividades.
Esforço	Tarefa exige baixa demanda mental e pouco esforço físico.	Tarefa exige alta demanda mental e muito esforço físico.
Frustração	Alto grau de segurança durante a tarefa.	Baixo grau de segurança durante a tarefa.
Desempenho	Baixa satisfação do usuário com seu desempenho na completude das tarefas.	Alta satisfação do usuário com seu desempenho na completude das tarefas.

Fonte: Soares, 2012. p. 82.

Tabela IV. Resultado do NASA TLX para usuários do Mouse.

NASA-TLX	Pontuação-TLX	Demanda Mental	Demanda Física	Demanda Temporal	Esforço	Frustração	Desempenho
Usuário 1	24.16	71	3	3	23	2	43
Usuário 2	29.16	71	3	2	45	3	51
Usuário 3	19.84	49	0	0	3	0	67
Usuário 4	17.66	20	0	75	0	0	11
Usuário 5	22.49	39	5	20	30	15	26

Tabela V. Resultado do NASA TLX para usuários da voz.

NASA-TLX	Pontuação-TLX	Demanda Mental	Demanda Física	Demanda Temporal	Esforço	Frustração	Desempenho
Usuário 1	34.82	60	5	36	50	35	23
Usuário 2	73.47	80	16	92	70	80	42
Usuário 3	19.66	25	0	6	11	50	26
Usuário 4	14.17	64	11	19	11	16	16
Usuário 5	9.82	54	2	2	5	17	24

Tabela VI. Média do NASA TLX.

Médias do NASA-TLX	Pontuação-TLX	Demanda Mental	Demanda Física	Demanda Temporal	Esforço	Frustração	Desempenho
Média do Grupo do Mouse.	22.49	50	2.2	20	20.2	4	39,6
Média do Grupo de Voz.	30.38	37.2	6.8	31	29.4	39.6	26.2

5.10 Análise Do Teste

A análise dos resultados verificará os quesitos apontados por Mazza (2009), sendo eles: funcionalidade, eficácia, eficiência, usabilidade e utilidade. Segundo Mazza estes quesitos são indicadores que atestam boas característica de usabilidade em um software de visualização de informação.

5.10.1 Quanto à Funcionalidade

É possivel através do uso de interação por interface natural fazer todas as ações que pelo meio convencional? Esta é a proposta defendida neste trabalho, para isso foi criada a ferramenta de visualização de informação IVOrpheus. Tal ferramenta contempla todas as funcionalidades previstas no mantra de Shineidermman. Sendo elas: Visão geral, permiti ao usuário visualizar em sua totalidade uma base configurada. Zoom (Escala), permiti ao usuário aumentar/diminuir a escala de uma visualização, para

assim poder observar agrupamentos e tendências nos dados.

Filtros: propicia que o usuário remova informações da visualização, assim deixando para ser analisada apenas a informação relevante ao mesmo. Detalhes Sobre Demanda, esta funcionalidade tem o objetivo de facilitar o processo de obter informações extras sobre um ponto, além dos valores dos eixos e dos canais visuais.

Todas estas funcionalidades foram contempladas tanto para o mouse quanto para a voz. Sendo que a resposta para indagação inicial, é que sim, é possivel através do uso de uma interface de interação natural fazer todas as ações realizadas pelo meio de interação convencional.

5.10.2 Quanto à Eficácia

Será que a interação proporciona com que os usuários completem todas as tarefas? Durante os teste com os usuários todos os participantes que utilizaram o mouse conseguiram completar todas as tarefas.

Enquanto os que utilizaram a voz, quatro de cinco (4/5) dos usuários conseguiram terminar todas as tarefas, salvo um voluntário que apresentou dificuldades na terceira tarefa, mais precisamente na sub-tarefa de aplicar detalhes sobre demanda em uma visualização já configurada, devido o usuário ter extrapolado o tempo a tarefa foi dada como incompleta.

Conclui-se que a eficácia da interação por voz em 80% dos casos estudados. Isto se deu devido a interface por voz não ter apresentado nenhum problema de usabilidade com grau de severidade elevado (acima de 3). Ou seja, não ocorreu nenhum problema de usabilidade com impacto, frequência e persistência, que impedisse o progresso do usuário perante as tarefas.

5.10.3 Quanto à Eficiência

O uso da entrada por voz possibilita um menor tempo e uma menor carga de trabalho na execução das tarefas? Como pode ser visto nas Figuras 5.1 e 5.2, a primeira tarefa teve um tempo médio para sua completude de 02:57 segundos para mouse e 04:04 segundos para voz. Isso quer dizer que o mouse foi 28% mais rápido do que a voz, e nas demais subtarefas da tarefa 1 é possível acompanhar esta mesma tendência. Por exemplo, para configurar os eixos x, y e z, o mouse teve um tempo médio 44% menor

do que a voz. Enquanto na subtarefa de aplicar um filtro categórico foi apenas de 23% e para escala (zoom) e girar foi de 33% mais rápido.

O padrão se repete na segunda tarefa (Figuras 5.3 e 5.4) com uma média de tempo onde o mouse foi 25% mais rápido. Também na sub-tarefa de configurar cor, forma e tamanho, o mouse teve um tempo 16% menor. O mesmo serve para filtrar valores contínuos, que tiveram um tempo 27% menor pelo mouse. Para ler a visualização e interagir com ela o mouse persistiu mais rápido com uma diferênça de 27% em relação a voz. E na ultima sub-tarefa (ler legenda) da segunda tarefa, novamente, o mouse se sobressaiu com 29% mais rápido que a voz.

Na terceira tarefa, (Figuras 5.5 e 5.6) o mouse teve a vantagem de ser em média 11% mais rápido. O caso mais expressivo onde o mouse teve um tempo muito menor a voz foi no cenário detalhes sobre demandas 81%. Isso é compreensivo, pelo fato de selecionar um ponto por voz exige uma complexidade bem maior do que apenas clicar em um ponto com o mouse. Sublinhando isso, o tempo total para a execução de todas as tarefas teve a diferênça de 24% entre o mouse e a voz, mostrando o mouse como um meio de entrada mais eficiênte para a execução das tarefas propostas.

5.10.4 Quanto à Usabilidade

A utilização por voz foi simples ou intutiva o suficiênte para o a completude das tarefas? Para atestar a usabilidade do software foi proposto aos usuários responderem a dois questionários pós tarefa.

Sendo o primeiro o questionário para identificar o nível de dificuldades subjetivas durante a execução das tarefas e sub-tarefas, o qual o resultado pode ser visto nas tabelas II e III, como apresentado nestas tabelas a frequência de dificuldade para mouse foi, muito fácil (15), fácil (10), médio (20), difícil (4) e muito difícil (1). Enquanto para voz os a frequência de dificuldade foi, muito fácil (2), fácil (26) médio (15), difícil (7) e muito difícil (0).

De modo geral, na tabela II uma maior quantidade de valores estão concentrados entre os níveis muito fácil , fácil e médio. Por isso a média dos valores aponta que a interação por mouse foi de nível fácil. Enquanto, a interação por voz (tabela III), devido possuir uma frequência maior entre os níveis fácil, médio e difícil. Obteve um nível

médio de dificuldade. Também na tabela II e III é possível observar as tarefas e sub tarefas que os usuários tiveram maiores dificuldades.

O grupo da interação por mouse apresentou dificuldades nas tarefas de aplicar filtro contímuo e ler legenda. Quanto a sub tarefa de ler legenda os usuários, tanto de mouse quanto de voz tinham dificuldade de relacionar a cor, a forma e o tamanho dos pontos apresentados na área de visualização com a legenda de cores, formas e tamanhos do menu legenda.

Uma primeira hipótese para o entendimento desta dificuldade está em saber que os usuários vinham das mais diversas áreas do conhecimento. Sendo assim, muitos não estavam habituados com a atividade de ler a legenda presente em um gráfico.

Quanto a sub tarefa de filtrar valores contínuos, mesmo esta tendo sido apresentada no vídeo de treinamento. Alguns usuários ficaram confusos na hora da execução da mesma, isto pode ser explicado devido a aplicação de filtros em valores continuos ser diferente da aplicação de filtros em valores categóricos. Pois, os valores categoricos exigem apenas que os usuários selecionem os valores desejados a serem filtrados.

Enquanto os valores contínuos exigem que o usuário descrimine um intervalo de valores onde será aplicado o filtro, e este intervalo de valores é inserido pelo teclado ou comando de voz. Por ter esta complexidade extra os usuários demoravam um pouco mais para entender o seu funcionamento.

Enquanto, o grupo de interação por voz apresentou dificuldades em ler legenda, configurar canais visuais, e na tarefa de detalhes sobre demanda. Quanto a sub tarefa de configurar canais visuais, alguns usuários tinham dificuldades de identificar que tal funcionalidade atuava de forma separada da configuração dos eixos. Sendo que os mesmo tendiam a entrar na opção de configurar os eixos “X” ou “Y” ou “Z” e dentro destas opções procuravam a opção de configurar a cor, forma ou tamanho.

Isto se deve a inexperiência no uso de ferramentas de visualização, como na primeira tarefa lhes eram exigidos a configuração dos eixos, os mesmos tendiam a executar este passo na busca de configurar os canais visuais.

Quanto à tarefa de detalhes sobre demanda, os usuários utilizaram diferentes

abordagens para a execução da mesma, uma delas foi a de aplicar filtros continuos e categóricos na visualização, assim diminuindo o conjunto de dados dispostos na visualização. Para assim, selecionar o ponto especificado, tal abordagem foi utilizada pelos usuários 3, 4 e 5. Já os usuários 1 e 2 utilizaram a estratégia de selecionar o ponto especificado sem nenhuma aplicação prévia de filtro.

Isto se mostrou ineficiênte, pois como mostra a Figura 5.6 o primeiro usuário teve o maior tempo na execução desta tarefa e o segundo usuário extrapolou o tempo permitido, assim ficando com a tarefa inconclusa. Como os usuários 1 e 2 não aplicaram filtros na visualização de informação a mesma apresentou uma massiva quantidade de pontos o que dificultava a seleção do ponto especificado.

Após o questionário que visava mensurar o nível de dificuldade das tarefas, foi pedido aos usuários que preenchessem o questionário pós tarefa NASA-TLX. Sendo que este mede a carga de trabalho para a execução das tarefas e os seus resultados podem ser acompanhados nas tabelas IV, V e VI. Como pode ser visto nas tabelas IV,V e VI, a demanda física foi 68% inferior na utilização do mouse, e a demanda temporal do mouse foi 36% inferior a da voz, e consequentemente o mouse teve um esforço 32% inferior na sua utilização.

Tais resultados podem ser explicados pelo fato do usuário estar mais habituado a utilizar o mouse para interagir com o computador do que a voz. Em função desta familiaridade, é notável que os usuários terão uma demanda física, temporal e esforço conseideravelmente menores na utilização do meio de interação convencional do que com o meio de interação proposto.

Ademais, um dos fatores de maior peso durante a avaliação do NASA-TLX foi a frustração, que teve um resultado 90% inferior no mouse em comparação com a voz. A frustração identifica a insegurança ou desconforto do usuário durante a execução das tarefas e a mesma tem relação direta com o desempenho subjetivo do usuário e a pontuação TLX.

Logo, quanto maior a frustração maior será a carga de trabalho sentida pelo usuário, conforme demonstra o desempenho subjetivo e a pontução TLX, que foram respectivamente de 51% (maior) e 26% (menor) no mouse em comparação com a voz.

Sintetizando os resultados obtidos no quesito usabilidade, é possível concluir que o mouse foi mais fácil de ser utilizado do que a voz. Por ser um meio de interação com maior familiaridade, com menor demanda física e temporal, menor esforço, menor frustração, e consequentemente menor carga de trabalho final. Assim mostrando o mouse como um meio de interação eficiênte, enquanto o meio de interação por voz mostra ser eficaz.

5.10.5 Quanto à Utilidade

Em qual contexto a voz melhor beneficia o usuário? A entrada por voz tem o papel de atender uma gama maior de usuários. A exemplo, usuários com problemas motores ou usuários sem sensibilidade nas mãos, ambos podem fazer uso do sistema IVOrpheus para interagir com uma base de dados. Assim, alcançando o objetivo apresentado por Alan Kay (1972), que é quanto mais "amigável" for a interação entre homem e máquina, maior a gama de pessoas alcançadas.

CAPÍTULO 6- CONSIDERAÇÕES FINAIS

Este capítulo apresenta os desafios e limitações encontradas, conclusão e trabalhos futuros e por fim, as publicações aceitas.

6.1 Desafios Encontrados e Limitações

Durante o desenvolvimento da ferramenta IVOrpheus foi perceptível algumas limitações que influenciaram negativamente no desempenho da ferramenta e desafios que dificultaram a execução desta pesquisa. Entre as adversidades encontradas estavam:

- Foi mostrado na revisão sistemática adotada no capitulo 3, que são poucos os trabalhos que exploram a interação por voz em ambientes de visualização de informação. Acarretando em lacunas na utilização deste meio de entrada aplicado nesta área de pesquisa;
- Adaptação da interface para melhorar a experiência dos usuários. Na aplicação IVOrpheus foram evitados a utilização de comandos compostos por mais de uma palavra, como “Carregar Base”. Devido a testes preliminares mostrarem que os usuários tinham preferência por comandos simples que contivessem apenas uma palavra, como “Abrir”;
- O maior desafio encontrado foi a seleção de um ponto por voz. É evidente que a utilização da voz para esta atividade é de difícil execução. Principalmente por esta atividade está aplicada em uma visualização de dispersão de pontos, que faz uso de uma nuvem de pontos selecionáveis;
- Mesmo este trabalho tendo apresentado uma solução eficaz para a tarefa de seleção. Ainda é discrepante a diferença de tempo entre a seleção por voz e a seleção por mouse. Sendo o mouse 81% mais rápido para execução desta tarefa. Assim se faz necessário verificar outras possíveis soluções para tal dificuldade que diminua essa diferença e assim melhore a experiência do usuário;
- A ausência do reconhecimento de palavras estrangeiras pelo software Coruja. Com isso palavras como Plymouth ou Volkswagen, não eram reconhecidas;
- Falta de precisão no reconhecimento de palavras com poucos fonemas, como,

vogais e consoantes isoladas. Por exemplo, uma das palavras que menos reconhecidas era a consoante zê (Z). Isto acontecia devido o sistema Coruja ter pouco treinamento com palavras monofônicas;

- O ritmo de reconhecimento era lento e pausado, ou seja, o usuário tinha de esperar de dois a três segundo para um comando ser reconhecido e assim poder inserir o próximo comando;
- A não implementação do ditado contínuo. Caso o usuário desejasse inserir o número 35865. O mesmo deveria ditar: “três”, esperar de dois a três segundos para o comando ser reconhecido e após isso falar “cinco” e assim consecutivamente. Sabendo que possivelmente o usuário teria uma menor carga de trabalho se ditasse de forma contínua, a exemplo, “trinta e cinco mil e oitocentos e sessenta e cinco”;
- O não reconhecimento de siglas, por exemplo, ao inserir em uma gramática a sigla BMW, o sistema Coruja não reconhecia tal palavra. Assim, será necessário escrever por extenso a sigla dentro da gramática correspondente, por exemplo, era necessário escrever “Bemedabliou”;
- Geração de hipóteses erradas baseadas em mesmo prefixo, como exemplo, ao usuário inserir o comando “configurar” ou “filtrar”, era disposto na barra de menu as opções: “Eixo X”, “Eixo Y” e “Eixo Z”. Quando o usuário inserir o comando “Eixo Z”, o ASR gerava a saída de forma aleatória qualquer um destes comandos que tivessem o mesmo prefixo, ou seja, a saída poderia ser “Eixo X” ou “Eixo Y” ou “Eixo Z”. Para contorna tal situação as palavras com mesmo prefixo foram alteradas, removendo os prefixos;
- Reconhecimento impreciso para palavras com fonemas parecidos. Como exemplificação, pode-se mencionar a tela de filtrar contínuo. Nela, os comandos de voz utilizados para inserir os dados no campo de texto, eram “início” e “fim”. Porém, devido ao ASR gerar um reconhecimento equivocado da palavra “fim”, geralmente reconhecendo o número “cinco” no lugar, o comando de voz foi alterado para “término”, com o objetivo de evitar este obstáculo;

- Comando ou botão “parar” o reconhecimento de voz. Durante a execução dos testes, os usuários sentiam a necessidade de comunicar alguma informação. Porém, como o motor de reconhecimento não parava de “ouvi-los”, isso acarretava em um comportamento inesperado. Esta questão foi pesquisada nos estágios iniciais da aplicação IVOrpheus. Todavia, não foi encontrada uma solução, devido ao fato de que quando o Sistema Coruja era parado/retomado, o mesmo não recarregava as gramáticas. Assim parando o reconhecimento de voz na aplicação;
- Tratamento dos ruídos ambientes, os quais frequentemente causavam uma saída inesperada. Por isso os testes foram realizados em ambientes controlados, porém, até a respiração do usuário poderia acarretar em saída indesejada. Por isso, o usuário era instruído a manter o microfone a 18 centímetros de distância da boca.

6.2 Conclusão e Trabalhos Futuros

Neste trabalho foi desenvolvida uma ferramenta de visualização da informação em ambiente tridimensional, utilizando a técnica de visualização por dispersão de pontos, com interface baseada em entrada por voz, visando melhorar a experiência do usuário e aumentar a produtividade com a qual ele realiza suas consultas em bases de dados, tornando os comandos mais naturais e confortáveis para o usuário.

No desenvolvimento da ferramenta foram aplicadas diretrizes para sistemas de reconhecimento de voz, sistemas de visualização da informação e sistemas da área de IHC de forma geral, assim como boas práticas de programação na linguagem Java. Foram realizados testes de usabilidade com usuários de perfis distintos, com a finalidade de determinar a eficiência ou eficácia da ferramenta.

Durante o processo de implementação da ferramenta, houve a observação de boas práticas relativas principalmente às gramáticas utilizadas na API Coruja. Essas práticas foram:

- Classificar os comandos por voz em dois tipos: comandos globais da ferramenta existentes independente da base selecionada e comandos

específicos por base de dados, relativos aos atributos pertencentes a cada base;

- Gramáticas para comandos globais: ao determinar os comandos por voz, levar em consideração possíveis restrições da API de voz utilizada, por exemplo, dificuldades de reconhecimento de alguns fonemas e escolher comandos próximos à Linguagem Natural (LN) do usuário;
- Gramáticas para comandos específicos: analisar a base de dados, e se necessário realizar pré-processamento, a fim de adaptar os atributos em uma gramática legível ao sistema;

Com base nos resultados dos testes de usabilidade, foi observado que a maioria dos usuários que utilizaram a voz, tiveram um nível de frustração e carga de trabalho elevado para a completude das tarefas. Estes podem ter sido ampliados, devido ao processo de reconhecimento de voz, que ocasionalmente gerava hipóteses erradas, em virtude de ruídos ambientes ou ao ritmo de fala do usuário, entre outros problemas previamente apontados no seção 6.1.

Dessa forma, esses acontecimentos direcionaram a pesquisa ao seguinte questionamento: “Qual abordagem na utilização de voz na técnica de visualização de informação por dispersão de pontos em três dimensões tem maior eficiência na diminuição do tempo e carga de trabalho do usuário?”.

Como apontado pelos testes com os usuários a abordagem na utilização de voz com o objetivo de mimetizar o mouse, possui menor eficiência enquanto ao tempo e a carga de trabalho que a utilização de interface de interação padrão (mouse). No entanto, a utilização de uma interface natural se mostra eficaz ao permitir com que a maioria dos usuários terminassem todas as tarefas apresentadas.

Esta observação guia o futuro da aplicação IVOrpheus, visando atender uma proposta de interação por voz mais eficiente e que demande menor carga de trabalho do usuário.

Sabendo disso, para trabalhos futuros, o sistema IVOrpheus abordará a utilização da interação por voz através de diálogo, utilizando do princípio apresentado nos trabalhos (Cox et al. 2001), (Sun et al. 2010) e (Sharma et al. 2003), que deixam a

interface de interação invisível ao usuário, onde o mesmo não deve se ater a comandos de interface, como botões e painéis, todavia, apenas introduzirá a pergunta desejada e o sistema apresentará uma visualização de resposta.

Além disto, foram determinadas melhorias e adições nas funcionalidades atuais da ferramenta. Entre elas, encontram-se:

- A implementação de configurar canais visuais para valores contínuos, como a cor assumir gama de cores baseadas no valor contínuo que a mesma está configurada.
- Adicionar a funcionalidade de ordenar os dados nos eixos de forma crescente ou decrescente de acordo com a escolha do usuário.
- Possibilitar a derivação de dados a partir da base selecionada, como exemplo, gerar tabela com média de valores contínuos, assim gerando novas tabelas dentro da base selecionada.
- Implementar a funcionalidade de salvar o estado do usuário, assim, o usuário poderá continuar seu progresso de onde parou, ou poderá compartilhar suas descobertas com outros usuários.
- Criar seção para anotações do usuário, ou seja, permitir com que o usuário possa comentar descobertas por voz na base de dados.
- E, por fim, estender os testes de usuário para a visualização por dispersão de pontos em duas dimensões e comparar com os testes atuais, averiguando quanto o incremento/decremento de uma dimensão influência no tempo, na dificuldade e na carga de trabalho exercida nos usuários.

6.3 Publicações em Anais de Congresso

Como resultados deste trabalho, temos a publicação de um artigo em anais de evento e outro artigo aceito para publicação, São eles:

- L. Furtado, B. Miranda, N. Neto and B. Meiguins, "*IVOrpheus - A Proposal for Interaction by Voice Commands in Three-Dimensional Environments of Information Visualization*," *Computer and Information Technology;*

Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on, Liverpool, 2015, pp. 878-883.

- L. Furtado, A. Marques, N. Neto, B. Meiguins, M. Mota, (Aceito para publicação). "*IVOrpheus 2.0 - A proposal for interaction by voice command-control in three dimensional environments of information visualization*". HCI International 2016. Toronto, Canadá.

REFERÊNCIAS

ALI M. M.; RAVI S. S. **A new paradigm for human-building interaction: the use of CDF and augmented reality**. Automation in Construction, Elsevier 14, pgs 71–84, 2005.

Apple SIRI, assistente pessoal inteligente. Disponível em: <http://www.apple.com/br/ios/siri/>. Acesso em: jul. 2016.

BOHUS, D.; RUDNICKY, A. **LARRI: a language-based maintenance and repair assistant**. In: Proceedings of ISCA Tutorial and Research Workshop on Multi-modal Dialogue in Mobile Environments (IDS'02), Kloster Irsee, Germany, 2002.

BUSCHMAN F. **Pattern-Oriented Software Architecture Volume 1: A System of Patterns**, Willey, pp. 125, 1996.

CARD, S.; MACKINLAY, J.; SHNEIDERMAN, B. Readings in Information Visualization - **Using Vision to Think**, Morgan Kaufmann, 1999.

CORRADINI, A.; WESSON, R.M.; COHEN, P.R. **A map-based system using speech and 3D gestures for pervasive computing**. Proceedings. Fourth IEEE International Conference on Multimodal Interfaces, pp 191-196, 2002.

COX, K.; GRINTER, R.E.; HIBINO, S.L.; JAGADEESAN, L.J.; MANTILLA; D. **A Multi-Modal Natural Language Interface to an Information Visualization Environment**. J. of Speech Technology 4, 297–314 (2001) .

CHEN F.; CHOI E.; EPPS J.; LICHMAN S.; RUIZ N.; SHI Y.; TAIB R.; WU M. **A study of manual gesture-based selection for the PEMMI multimodal transport management interface**, Proceedings of the 7th international conference on Multimodal interfaces, Torento, Italy, 2005.

CHEN, C. **Information Visualization**. Vol. 1.Palgrave Macmillan, 2002.

DFKI, **The MARY Text-to-Speech System**, Disponível em: <http://mary.dfki.de/> Acesso em: fev. 2015.

Dados da população de Luxemburgo. Disponível em: <http://countrymeters.info/pt/Luxembourg>. Acesso em: fev. 2015.

DUTOIT T. **An introduction to text-to-speech synthesis**. Kluwer Academic, Dordrecht, 2001.

FINKE, M.; FRITSCH, J.; KOLL D.; WAIBEL A. **Modeling and Efficient Decoding of Large Vocabulary Conversational Speech**. in Proceedings, Eurospeech-99. Budapest: pp. 467-470, 1996.

FINKE M.; FRITSCH, J.; GEUTNER P.; RIES K.; ZEPPENFELD T.; WAIBEL A. **The JanusRTk Switchboard/Callhome 1997 Evaluation System**" in The DARPA Large Vocabulary Conversational Speech Recognition Hub5e Workshop, Baltimore,

1997.

GKESOULIS, D.; VASSILIADIS, P.; MANOUSIS, P. **CineCubes: aiding data workers gain insights from OLAP queries**. Inf. Syst., 2015.

Google Now Assistente Pessoal Inteligente. Disponível em: <https://www.google.com/landing/now/>. Acesso em: fev. 2016.

HART, S. G.; STAVELAND, L. E. **Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research**. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press, 1988.

HUANG X.; ACERO A.; HON H. **Spoken language processing**. Prentice-Hall, New York, 2001.

HEEREN, W.; VAN DER WERFF, L.; ORDELMAN R.; VAN HESSEN, A.; JONG, F. **Radio Oranje: searching the queen's speech(es)**. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, Amsterdam, The Netherlands, 2007.

IBM, Big Data IBM and Data Scientists: Disponível em: <http://www.ibm.com/big-data/us/en/>. Acesso em: fev. 2015.

JANKOWSKI J.; HACHET M. **A survey of interaction techniques for interactive 3D environments**. In: Eurographics 2013 state of the art reports. p. 65–93. 2013.

JMathPlot, API do modulo de Visualização. Disponível em: <http://jmathtools.sourceforge.net/doc/jmathplot/html/main.html> Acesso em: mar. 2015.

Julius, Motor de reconhecimento de Voz. Disponível em: <http://julius.sourceforge.jp/en_index.php>. Acesso em: jan. 2015.

KRAMMES, H.; SILVA, M. M.; MOTA, T.; TURA, M. T.; MACIEL, A.; NEDEL, L. **The Point Walker Multi-label Approach**. In Proceedings of 3D User Interfaces (3DUI),IEEE Symposium, pp 189-190, 2014.

KERREN, A.; EBERT, A.; MEYER, J.. **Human-Centered Visualization Environments**. p.403. 2007.

KARAT C.M.; VERGO J.; NAHAMOO D. **The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications**. Conversational interface technologies, L. Erlbaum Associates Inc., Hillsdale, NJ, 2003

KAY A. **A personal computer for children of all ages**. In Proceedings of the ACM Annual Conference, Volume 1, 1972.

LEE K., R.A. GRICE. **The Design and Development of User Interfaces for Voice Application in Mobile Devices**. IEEE International Professional Communication Conference, Saratoga Springs, New York, p. 308-320. 2006.

LIKERT, R. **A technique for the measurement of attitudes. Archives in Psychology**, 140, p. 1-55, 1932.

LUBOS, P.; BEIMLER, R.; LAMMERS, M.; STEINICKE, F. **Touching the Cloud: Bimanual Annotation of Immersive Point Clouds**. in Proceedings of 3D User Interfaces (3DUI),IEEE Symposium, pp 191-192. 2014.

MYERS B.; MALKIN R.; BETT M.; WAIBEL A.; BOSTWICK B.; MILLER R. C.; YANG J.; DENECKE M.; SEEMANN E.; ZHU J.; PECK C. H.; KONG D.; NICHOLS J.; SCHERLIS B.. **Flexi-modal and Multi-Machine User Interfaces**. In Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces ICMI'02, Pittsburgh, Pennsylvania, pgs 343-348. 2002.

MILLER C.; ROBINSON A.; WANG R.; CHUNG P.; QUEK F. **Interaction techniques for the analysis of complex data on high-resolution displays**. Proceedings of the 10th international conference on Multimodal interfaces, Chania, Crete, Greece, 2008.

MAZZA R. **Introduction to Information Visualization**. Springer Publishing Company, Incorporated, p. 125, 2009.

Microsoft Cortana, assistente pessoal inteligente. Disponível em: <http://www.microsoft.com/pt-br/celulares/experiences/cortana/>. Acesso em: fev. 2016.

NIELSEN, J. **Why you only need to test with 5 users**. Usability Engineering. Boston, 2000.

NETO, N.; SILVA, P.; KLAUTAU, A.; TRANCOSO. I. **Free tools and resources for Brazilian Portuguese speech recognition**. Journal of the Brazilian Computer Society. Springer. v 17. n 1. p 53--68. 2010.

NETO, N.; SILVA, C.; KLAUTAU, A.; BATISTA, P. **Coruja: Um Reconhecedor de Voz Livre para Português Brasileiro com Interface de Programação**. Disponível em: <www.laps.ufpa.br/falabrasil/downloads.php.>. Acesso em: jan. 2015.

OLIVEIRA R.; BATISTA P.; NETO N.; KLAUTAU A. **Recursos para desenvolvimento de aplicativos com suporte a reconhecimento de voz para desktop e sistemas embarcados**. 12˚ Forum Internacional de Software Livre, 2011.

PERUGINI, S.; MCDEVITT K.; RICHARDSON R.; PÉREZ-QUIÑONES M. A.; SHEN, R.; RAMAKRISHNAN, N.; WILLIAMS, C.; FOX E. A. **Enhancing usability in CITIDEL: multimodal, multilingual, and interactive visualization interfaces**. Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, Tuscon, AZ, USA, 2004.

ROCHA, H. V. da; BARANAUSKAS, M. C. C. **Design e Avaliação de Interfaces Humano-Computador**. Campinas, Instituto de Computação - Universidade Estadual

de Campinas, 2003.

RABINER L.; JUANG B. **Fundamentals of speech recognition**. PTR Prentice Hall, Englewood Cliffs, 1993.

SHARMA, R.; YEASIN, M.; KRAHNSTOEVER, N.; RAUSCHERT, I.; CAI, G.; BREWER, I.; MACEACHREN, A.M.; SENGUPTA, K. **Speech-Gesture Driven Multimodal Interfaces for Crisis Management**. Proc. IEEE, Vol. 91, No. 9, pp. 1327-1354, 2003.

SUN Y.; LEIGH J.; JOHNSON A.; LEE S. **Articulate: A Semi-automated Model for Translating Natural Language Queries into Meaningful Visualizations**. Proceedings of the 10th international conference on Smart graphics, EUA, p. 184-195, 2010.

SHINEIDERMAN ,B. **The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations**. In 1996 IEEE Symposium on Visual Languages. Page 336, 1996.

SABIR, K.; STOLTE, C.; TABOR, B.; O'DONOGHUE, S.I. **The Molecular Control Toolkit: Controlling 3D molecular graphics via gesture and voice**. *Biological Data Visualization (BioVis), 2013 IEEE Symposium on* , vol., no., p. 49,56, 2013.

SILVA, P.; BATISTA, P.; NETO, N.; KLAUTAU, A. **An open-source speech recognizer for Brazilian Portuguese with a windows programming interface**. The International Conference on Computational Processing of Portuguese (PROPOR), 2010.

SOARES, G. M. A. **Uma interface t-commerce com o auxílio de uma técnica de visualização da informação para o middleware brasileiro de iDTV**. 82 f. Dissertação (Mestrado em Ciência da Computação) – Instituto de Ciências Exatas e Naturais (Programa de Pós-Graduação em Ciência da Computação), UFPA, Belém, 2012.

TAYLOR P. **Text-to-speech synthesis**. Cambridge University Press, Cambridge, 2009.

TE'ENI D.; CAREY J. M.; ZHANG P. **Human-Computer Interaction: Developing Effective Organizational Information Systems**. John Wiley & Sons, pg 56, 2005.

VAN DAM A.; LAIDLAW D. H; SIMPSON R. M. **Experiments in Immersive Virtual Reality for Scientific Visualization**. Computers & Graphics 2002.

VANNEVAR, B. **As We May Think**. In Atlantic Monthly, p. 112-124. Publishing Press, 1945.

WARD M.; GRINSTEIN G.; KEIM D. **Interactive Data Visualization: Foundations, Techniques, and Applications**, A. K. Peters, Ltd., Natick, MA, p. 28, 2010.

WEGMAN E. **Affordable Environments for 3D Collaborative Data Visualization**. Computing in Science & Engineering, vol2, no 6, p. 68-72, 2000.