



DISSERTAÇÃO

Mestrado em Engenharia Eletrotécnica - Electrónica e Telecomunicações

**IF2LGP - Intérprete automático de fala em Língua
Portuguesa para Língua Gestual Portuguesa**

LUÍS ROMÃO GASPAR

Leiria, Setembro de 2015



DISSERTAÇÃO

Mestrado em Engenharia Eletrotécnica - Electrónica e Telecomunicações

IF2LGP - Intérprete automático de fala em Língua Portuguesa para Língua Gestual Portuguesa

LUÍS ROMÃO GASPAR

Dissertação de Mestrado realizada sob a orientação da Professora Doutora Carla Alexandra Calado Lopes, da Escola Superior de Tecnologia e Gestão (**ESTG**) do Instituto Politécnico de Leiria (**IPL**).

Leiria, Setembro de 2015

*Dedico este trabalho à minha
família e aos meus avós.*

Luís Romão Gaspar

Agradecimentos

Este trabalho só foi possível graças a inúmeras pessoas e entidades. A todos um muito obrigado.

Um especial agradecimento à minha família pelos ensinamentos e conselhos durante a minha vida académica e pessoal.

Um especial agradecimento a todos os professores e à **ESTG** pelos ensinamentos e pelas condições que proporcionaram.

Um especial agradecimento à Professora Doutora Carla Alexandra Calado Lopes pela sua orientação e ajuda no desenvolvimento deste trabalho.

Um muito obrigado à Associação de Surdos da Alta Estremadura (**ASAE**) pela sua receptividade e ajuda no desenvolvimento deste trabalho.

Por último, não menos importante, agradeço à minha namorada Daniela Coelho pelo apoio e muita paciência que teve comigo.

Resumo

Segundo informação da Organização Mundial de Saúde existem no mundo 360 milhões de pessoas, cerca de 5% da população mundial, que possuem algum nível de perda de audição. Em Portugal este número ascende a 150 mil pessoas, sendo que dessas cerca de 30 mil utilizam a língua gestual como forma de comunicação. A comunicação entre um surdo profundo e um ouvinte é, geralmente, muito difícil devido ao desconhecimento do ouvinte sobre a língua gestual, sendo necessária, na maioria das vezes, a mediação de um intérprete. No dia-a-dia de um surdo, torna-se difícil a presença permanente de um intérprete, especialmente em contextos fora do profissional ou escolar. O trabalho que se apresenta surge, numa tentativa, de resposta a este problema. Pretende-se desenvolver uma aplicação que permita fazer a conversão de fala em Língua Portuguesa (LP) para Língua Gestual Portuguesa (LGP), de forma automática, que permita auxiliar a comunicação entre ouvintes e surdos na ausência de um intérprete. O sistema captura o áudio produzido pelo emissor falante e apresenta ao surdo gestuante uma sequência de vídeos correspondente à mensagem produzida oralmente. O sistema é constituído por 3 módulos principais: (1) Um reconhecedor de fala, que parte do padrão acústico e fornece na sua saída a sequência de palavras reconhecida; (2) uma máquina de tradução, que parte do texto em LP e apresenta o mesmo texto segundo a estrutura frásica da LGP - GLOSA e (3) uma interface gráfica, que partindo da sequência de palavras em LGP, faz a animação por gestos através de vídeos.

O sistema foi avaliado por um pequeno grupo de pessoas, os resultados dessa avaliação demonstraram que o sistema é útil e fácil de utilização mas necessita ainda de alguns melhoramentos.

Palavras-chave: Reconhecimento automático de fala, GLOSA, tradução automática, Língua Gestual Portuguesa

Abstract

According to the World Health Organization information, there are 360 million people, around 5% of the world population, that have some level of hearing loss. In Portugal, there are about 150 thousand people affected by hearing impairment, of which 30 thousand use sign language to communicate. The communication between hard-of-hearing deaf and a hearing person is usually very difficult, often requiring the mediation of an interpreter, since most people are not familiarized with sign language. On the daily basis life of a deaf it is difficult to have an interpreter with him at all time, especially outside professional or academic contexts. The work presented on this dissertation is an attempt to respond to this problem. It is intended to develop an application that automatically converts Portuguese speech to Portuguese Sign Language, facilitating the communication between deaf and hearing in the case that an interpreter is not present. The system records the audio produced by the speaker and display the corresponding video sequence in Portuguese Sign Language to the deaf. The system consists of three main modules: (1) A speech recognizer, for decoding the spoken utterance into a word sequence; (2) a translation machine, for converting a Portuguese word sequence into a sequence of signs belonging to the Portuguese Sign Language and (3) a graphical interface, for playing back the gestures.

The system was evaluated by a small group of people, the results of this evaluation showed that the system is useful and easy to use but still require some improvements.

Keywords: Automatic Speech Recognition, GLOSS, machine translation, Portuguese Sign Language

Índice

Agradecimentos	iii
Resumo	v
Abstract	vii
Índice	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
Lista de Abreviaturas	xvii
1 Introdução	1
1.1 Objetivos e Metodologia de Desenvolvimento	2
1.2 Organização do documento	6
2 Estado da Arte	9
2.1 Projetos Inclusivos que usam a LGP	9
2.2 Projetos vocacionados para o Ensino da Língua Gestual Portuguesa	11
2.2.1 Ensino da Língua Gestual Assistido por Personagens 3D Virtuais	11
2.2.2 Avatares em Língua Gestual Portuguesa	13
2.3 Sistemas Automáticos de Tradução de Fala para Línguas Gestuais	16
2.3.1 Virtual Sign Project	16
2.3.2 iCommunicator	18
2.3.3 TESSA	18
2.3.4 Speech to sign language translation system for Spanish	19
2.3.5 Speech Translation into Pakistan Sign Language	22

2.3.6	Falibras	23
2.3.7	Tradutores Automáticos em Aplicações Móveis	25
2.4	Sumário	27
3	Desenvolvimento modular do sistema implementado	31
3.1	Módulo 1: Reconhecimento de Fala	31
3.1.1	Escolha do Reconhecedor de Fala	33
3.1.2	Desenvolvimento, Testes e Conclusões do Reconhecedor	35
3.1.2.1	Avaliação da <i>Web Speech API</i>	38
3.2	Módulo 2: Máquina de Tradução	43
3.2.1	Estudo da Língua Gestual Portuguesa	43
3.2.2	Escolha do Método para a Máquina de Tradução	46
3.2.3	Desenvolvimento da Máquina de Tradução	48
3.2.4	Testes e Conclusões	55
3.3	Módulo 3: <i>Interface</i> Gráfica	58
3.3.1	Escolha da <i>Interface</i> Gráfica	58
3.3.2	Desenvolvimento da <i>Interface</i> Gráfica	59
3.3.3	Testes e Conclusões	60
4	Funcionamento Global do Sistema	65
4.1	Integração dos 3 Módulos	65
4.2	Avaliação por alguns surdos e intérpretes da ASAE, alguns Especialistas e Público-Geral	68
4.2.1	Objetivos	69
4.2.2	Método de Apresentação do Sistema	69
4.2.3	Questionário de Avaliação ao Sistema	71
4.2.4	Resultados e Conclusões	71
5	Conclusões e trabalho futuro	77
	Bibliografia	79
A	Tutoriais	87
A.1	Tutorial Google Speech API	87
A.2	Tutorial Tradutor LP para LGP	105

Lista de Figuras

1.1	Comparação de alfabetos utilizados na Língua Gestual Portuguesa (LGP) e na língua gestual nos Estados Unidos da América, a <i>American Sign Language</i> (ASL).	2
1.2	Diagrama de blocos do sistema desenvolvido.	3
1.3	Arquitetura do sistema desenvolvido.	4
2.1	<i>Interface</i> do Construtor 3D	11
2.2	<i>Interface</i> do Abecedário 3D	12
2.3	<i>Interface</i> do Animações Vídeio	12
2.4	<i>Interface</i> do Gestuário 2D	13
2.5	Avatar final usado pelo autor	14
2.6	Adivinha gesto	14
2.7	Representação em LGP da palavra “verde”	15
2.8	Composição de frases	15
2.9	Exemplo de tradução de texto para gestos LGP do projeto <i>Virtual Sign</i> .	16
2.10	Exemplo de tradução de gestos LGP para texto do projeto <i>Virtual Sign</i> .	17
2.11	iCommunicator	18
2.12	Diagrama do sistema TESSA	19
2.13	Diagrama do sistema “ <i>Speech to sign language translation system for Spanish</i> ”	20
2.14	Diagrama da Máquina de Tradução do módulo tradução da fala para LSE de “ <i>Advanced Speech Communication System for Deaf People</i> ”	21
2.15	Diagrama do sistema “ <i>Speech Translation into Pakistan Sign Language</i> ”	22
2.16	Mimix: primeira utilização, reconhecimento da fala e representação do gesto “Hello” (“Olá”) previamente reconhecido.	25
2.17	ProDeaf: representação do gesto “Olá”.	26
2.18	Hand Talk: representação do gesto “Olá”.	27

2.19	Sistema automático de fala para língua gestual genérico.	28
3.1	Diagrama de blocos genérico para um sistema de reconhecimento de fala	32
3.2	Testes ao módulo 1: início do reconhecimento da fala.	36
3.3	Testes ao módulo 1: utilização do microfone.	36
3.4	Testes ao módulo 1: paragem do reconhecimento de fala.	37
3.5	Testes ao módulo 1: frase reconhecida.	37
3.6	Módulo 2 do sistema automático de fala LP para LGP.	48
3.7	Diagrama do funcionamento do algoritmo de tradução desenvolvido.	51
3.8	Testes ao módulo 3.	63
4.1	<i>Website</i> com integração dos 3 módulos.	66
4.2	<i>Website</i> com integração dos 3 módulos: frase LP reconhecida.	67
4.3	<i>Website</i> com integração dos 3 módulos: vídeo correspondente à frase LP reconhecida.	68
4.4	<i>Website</i> com integração dos 3 módulos: frases LP e respectivas GLOSA.	70
4.5	Resultados questionário: questão sobre utilidade do sistema.	72
4.6	Resultados questionário: questão sobre experiência/facilidade de usar o sistema.	72
4.7	Resultados questionário: questão sobre o visual atrativo.	73
4.8	Resultados questionário: questão de avaliação do reconhecedor de fala.	73
4.9	Resultados questionário: questão sobre a avaliação da conversão para GLOSA.	74
4.10	Resultados questionário: questão sobre a avaliação dos vídeos em LGP.	74
4.11	Resultados questionário: questão sobre o tamanho da janela de vídeo.	75
4.12	Resultados questionário: questão sobre a qualidade visual dos vídeos.	75

Lista de Tabelas

2.1	Comparação das <i>Interfaces</i> Gráficas	29
3.1	Frases de Teste usadas na avaliação do <i>Web Speech API</i>	39
3.2	<i>HResults</i> : frase 5 (tabela 3.1) nos ficheiros em formato “MLF”	40
3.3	<i>HResults</i> : frase 5 (tabela 3.1) alinhamento	40
3.4	Resultados das eficiências do <i>Web Speech API</i> para português europeu. . .	41
3.5	Resultados da eficiência do <i>Web Speech API</i> para inglês	42
3.6	GLOSA: frases em SOV e em OSV	44
3.7	GLOSA: traduções erradas	44
3.8	GLOSA: frases negativas	45
3.9	GLOSA: frases interrogativas	45
3.10	GLOSA: tempos verbais	46
3.11	Resultado simplificado do servidor da <i>Lemmatization, PoS and Parsing</i> na versão análise morfossintática: “Ontem fui ao mercado”	49
3.12	GLOSA das Frases de Teste usadas na avaliação do <i>Web Speech API</i> . . .	50
3.13	Descrição dos blocos com as principais regras da LGP implementadas . . .	53
3.14	Testes ao módulo 2: “olá!”.	55
3.15	Testes ao módulo 2: “o meu carro é amarelo” (tabela 3.1)	56
3.16	Especificações dos vídeos LGP guardados.	59
3.17	Concatenar vídeos com o FFMPEG: frase “Ela viaja no domingo”	60
3.18	Concatenar vídeos com o FFMPEG: frase “A cor do gorro é roxo escuro” .	61
4.1	Questionário: perguntas sobre o sistema.	71

Lista de Abreviaturas

ASAE Associação de Surdos da Alta Estremadura

ASL *American Sign Language*

BD Base de Dados

BSL *British Sign Language*

ESTG Escola Superior de Tecnologia e Gestão

HMM *Hidden Markov Models*

HTK *Hidden Markov Model Toolkit*

HTML *Hyper Text Markup Language*

IPL Instituto Politécnico de Leiria

LIBRAS Língua Brasileira de Sinais

LSE *Lengue de Signos Española*

LGP Língua Gestual Portuguesa

LP Língua Portuguesa

MT Máquina de Tradução

PHP *PHP: Hypertext Preprocessor*

PSL *Pakistan Sign Language*

OSV OBJETO-SUJEITO-VERBO

SOV SUJEITO-OBJETO-VERBO

SVO SUJEITO-VERBO-OBJETO

Capítulo 1

Introdução

Segundo informação da Organização Mundial de Saúde existem no mundo 360 milhões de pessoas, ou seja, cerca 5% da população mundial, que possuem algum nível de perda de audição [1]. Algumas causas para perda de audição são: hereditariedade, hemorragias no ouvido interno, meningite, ruído e envelhecimento [2]. Considera-se que uma pessoa possui deficiência auditiva a partir do 25 dB nível de audição em ambos os ouvidos [1]. Esta deficiência é avaliada por 4 níveis de intensidade: leve, moderada, severa e profunda [1]. As pessoas que sofrem de perda de audição suave a severa usam a fala em conjunto com os aparelhos auditivos ou os implantes cocleares¹ para comunicar. Por vezes utilizam também a língua gestual como meio de comunicação. Já as pessoas que possuem uma perda de audição profunda usam apenas língua gestual para comunicar, pois os aparelhos auditivos e implantes cocleares são incapazes de melhorar a sua condição auditiva [1].

Em Portugal, segundo a Associação Portuguesa de Surdos, existem 150000 pessoas que possuem algum nível de perda de audição e dessas cerca de 30000 utilizam a Língua Gestual Portuguesa (LGP).

A LGP é a língua gestual usada pela comunidade surda portuguesa que se caracteriza por ser uma forma de comunicação através dos movimentos das mãos e expressões corporais e faciais regidos por um vocabulário e gramática próprios, em que os seus significados aparecem sob a forma de gestos [2, 3].

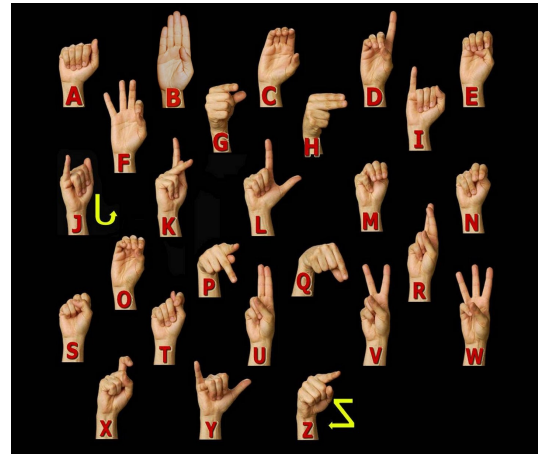
Enquanto que as pessoas portuguesas sem limitações utilizam a Língua Portuguesa (LP) como língua de comunicação, língua que utiliza os padrões acústicos para transmitir as ideias, as pessoas surdas portuguesas utilizam a LGP. Como as duas línguas possuem características únicas, isto é, as duas línguas possuem estruturas frásicas diferentes não se pode gestuar uma frase traduzindo palavra a palavra. Assim, quando um surdo e um ouvinte (pessoa sem limitações) querem dialogar necessitam na maioria das vezes dos intérpretes pois o ouvinte desconhece a LGP.

¹<http://www.neurelec.com/pt/implante-coclear>

À semelhança das línguas orais, as línguas gestuais variam de país para país [2]. Como forma de comparação na figura 1.1 apresentam-se os alfabetos utilizados na **LGP** e na língua gestual utilizada nos Estados Unidos da América, a *American Sign Language* (**ASL**) e como se pode constatar, por exemplo, a letra 'A' tem representações diferentes nas duas línguas.



(a) LGP [3]



(b) ASL [4]

Figura 1.1: Comparação de alfabetos utilizados na Língua Gestual Portuguesa (LGP) e na língua gestual nos Estados Unidos da América, a *American Sign Language* (ASL).

No dia-a-dia de um surdo português torna-se difícil a presença permanente de um intérprete o que este trabalho surge numa tentativa de resposta a este problema.

Com este trabalho pretende-se desenvolver uma aplicação que permita fazer a conversão de fala em **LP** para **LGP** de forma automática. O sistema captura o áudio produzido pelo emissor falante e apresenta ao surdo uma sequência de vídeos, através de uma pessoa gestuante, correspondente à mensagem produzida oralmente.

1.1 Objetivos e Metodologia de Desenvolvimento

Dos 360 milhões de pessoas que possuem algum nível de perda de audição [1] cerca de 150000 são portuguesas e dessas cerca de 30000 utilizam a **LGP** para comunicar, segundo a informação fornecida pela Associação Portuguesa de Surdos². De forma ajudar essas 30000 pessoas pretendeu-se com este trabalho desenvolver um sistema automático de tradução de fala em **LP** para **LGP** que permita auxiliar a comunicação entre ouvintes e surdos aquando da ausência dos intérpretes. O sistema terá que ter a capacidade de traduzir várias frases simples de temáticas do quotidiano e do curso de engenharia eletrotécnica, tais como, “Eu vou à escola” e “Calcula a tensão elétrica do circuito”, respetivamente.

²<http://www.apsurdos.org.pt/>

Na figura 1.2 apresenta-se o diagrama de blocos do sistema desenvolvido.

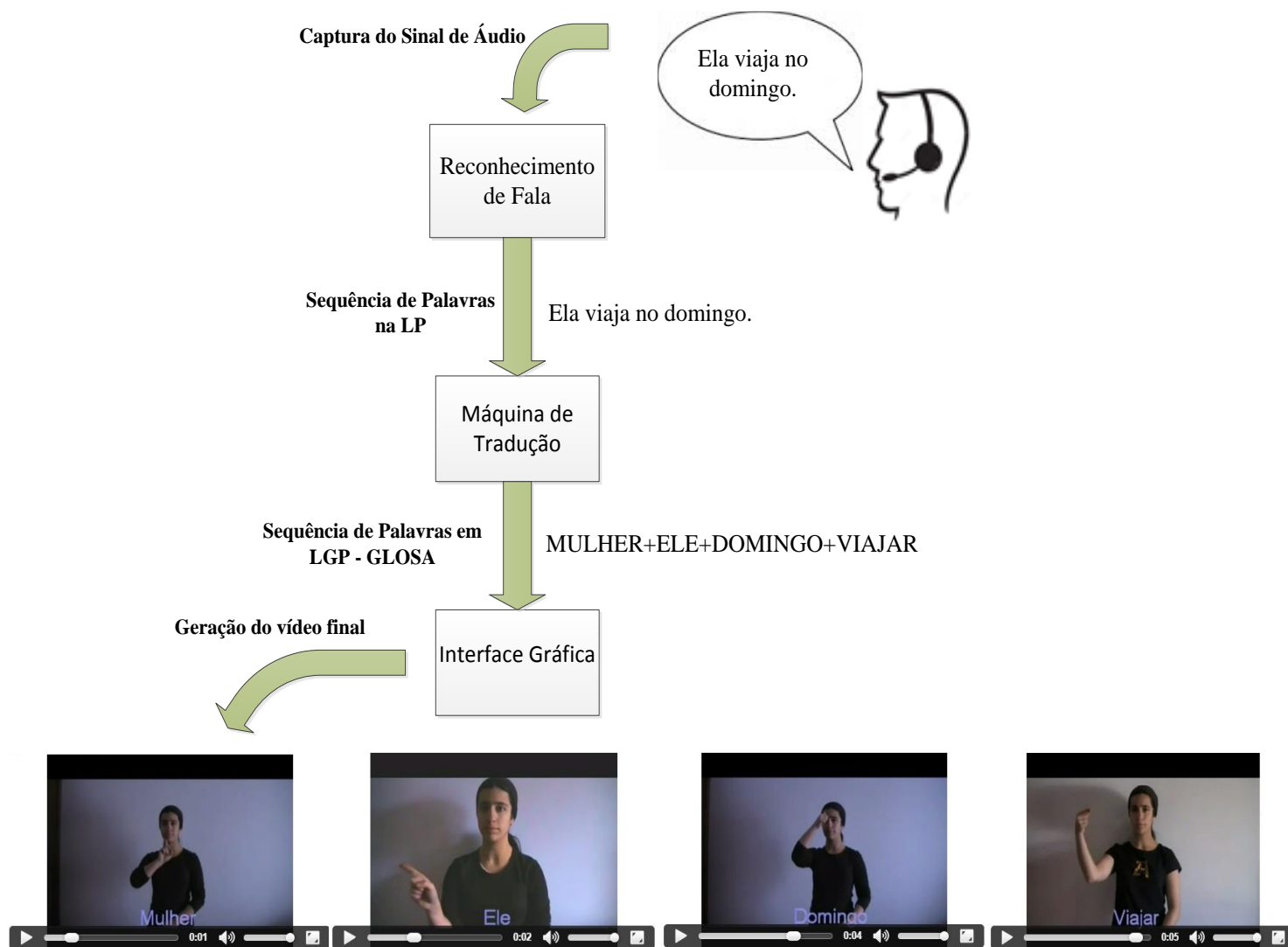


Figura 1.2: Diagrama de blocos do sistema desenvolvido.

Observando a figura 1.2 verifica-se que o emissor ouvinte discursa uma frase, por exemplo, “Ela viaja no domingo”, que é convertida para texto em **LP** pelo reconhecedor de fala - módulo 1. Depois essa frase entrará na Máquina de Tradução (**MT**) - módulo 2 que faz a conversão de texto em **LP** numa sequência de texto em **LGP - GLOSA** e, por fim cada vídeo das palavras da GLOSA é concatenado e é gerado um vídeo final na *interface* gráfica - módulo 3, para que possa ser visualizada a mensagem pelo surdo. Na figura 1.2 é possível verificar a concatenação dos vídeos dos gestos MULHER, ELE, DOMINGO e VIAJAR de acordo com a GLOSA.

Tendo como base o diagrama anterior (figura 1.2), na figura 1.3 apresenta-se a arquitetura do sistema desenvolvido.

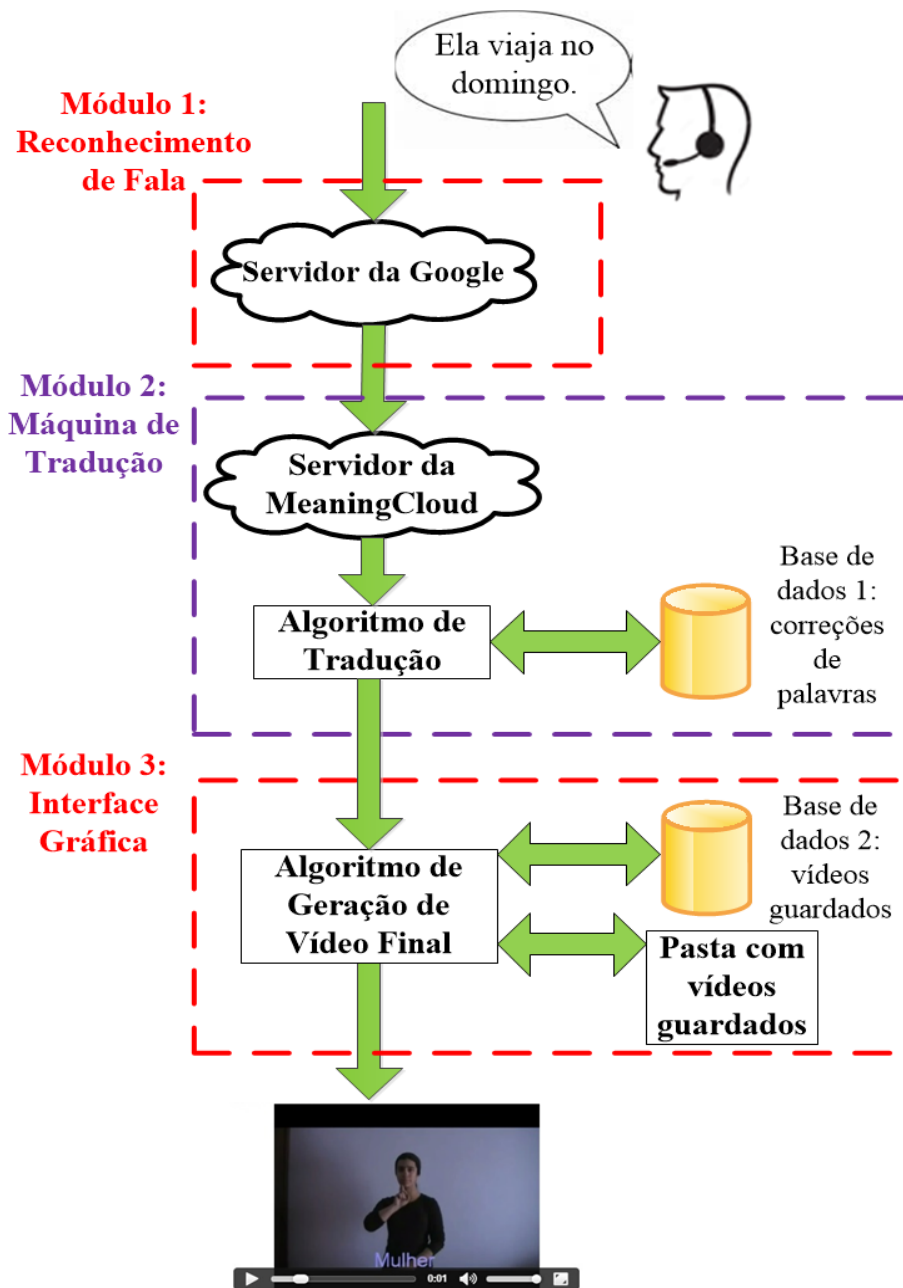


Figura 1.3: Arquitetura do sistema desenvolvido.

A captura do sinal de áudio e a sua conversão para texto em LP é realizada no módulo de reconhecimento de fala (figura 1.3). Esta conversão é realizada pela ferramenta *Web Speech API* da Google [5] (tema que será abordado em detalhe na secção 3.1).

O texto reconhecido terá que ser convertido numa sequência de palavras em LGP. Esse processo é realizado no módulo da MT (figura 1.3). O texto proveniente do módulo anterior será inicialmente analisado no servidor da *MeaningCloud* que o caracteriza

morfologicamente. Depois o texto caracterizado morfologicamente irá para um algoritmo de tradução (algoritmo desenvolvido no âmbito deste trabalho) que o converterá numa sequência de palavras de acordo com as regras da **LGP**.

A análise morfológica do texto reconhecido é realizada pela ferramenta *Lemmatization, PoS and Parsing* na versão análise morfossintática da *MeaningCloud* [6].

O algoritmo de tradução traduz o texto caracterizado morfologicamente para GLOSA (sequência de palavras em **LGP**) através de regras da **LGP**. Esse algoritmo usa uma base de dados, que contém uma lista de palavras, para corrigir eventuais erros do texto caracterizado morfologicamente. Na secção 3.2 será abordado em detalhe este tema.

A GLOSA, proveniente da **MT**, entrará no módulo da *interface* gráfica (figura 1.3). Esse módulo tem como finalidade a conversão de GLOSA para um vídeo final que a represente e, será mostrado ao surdo na aplicação. Esse processo é implementado através de um algoritmo de geração de vídeo final (algoritmo desenvolvido no âmbito deste trabalho).

O algoritmo desenvolvido que fizesse a conversão de GLOSA para um vídeo final consistiu: para cada palavra da GLOSA é verificada se existe o seu vídeo correspondente numa Base de Dados (**BD**) (figura 1.3). Essa **BD** contém uma tabela de todos os vídeos guardados que existem no servidor da aplicação. Se a palavra existir na **BD**, então o algoritmo cria um ficheiro de texto com a localização do vídeo correspondente para essa palavra; se não existir na **BD**, o algoritmo parte a palavra letra a letra. Cada letra dessa palavra é enviada à **BD** a fim de encontrar o seu vídeo respetivo e, coloca a localização do vídeo correspondente para essa letra no ficheiro de texto. Esse ficheiro de texto fica guardado no servidor. Deste modo, está a ser usada a datilologia (secção 3.2.1). Se não existir vídeo para essa letra, o sistema não coloca nenhuma localização do vídeo no ficheiro de texto e continua o algoritmo, pois não existem vídeos das letras acentuadas, tais como, ‘é’, ‘ç’ e ‘â’. Estando criado o ficheiro de texto com todas as localizações de todos os vídeos respetivos à GLOSA utilizou-se o *software* FFmpeg [7, 8, 89]. Esse *software* lê linha a linha do ficheiro de texto e faz a concatenação dos vídeos e gera um vídeo final que represente a GLOSA. Esse vídeo final é depois mostrado numa página *web* desta aplicação para o surdo a ver. Na secção 3.3 será abordado em detalhe este tema.

As etapas para concretização do trabalho foram:

1. Definição de um conjunto de frases simples com temáticas do quotidiano e do curso de engenharia electrotécnica com o objectivo de possibilitar uma avaliação ao módulo de reconhecimento de fala e de possibilitar o desenvolvimento das regras da **LGP** no módulo da **MT**
2. Estudo sobre alternativas que fizessem o reconhecimento de fala em **LP** para texto em **LP**. Realização de testes e conclusões.
3. Estudo da **LGP**, estudo sobre alternativas que possibilitassem o desenvolvimento da máquina de tradução que faça a conversão de texto em **LP** para texto em **LGP**. Realização de testes e conclusões.
4. Estudo de alternativas que fizessem a concatenação de vídeos e crie um vídeo final de acordo com o texto em **LGP** recebido no módulo anterior. Realização de testes e conclusões.
5. Desenvolvimento de um *website* com os três módulos e realização de inquéritos a algumas pessoas a fim de adquirir indicadores de melhoria ao trabalho realizado

1.2 Organização do documento

A estrutura desta dissertação está definida em 5 capítulos. O presente capítulo tem como objetivo fazer uma introdução ao tema do trabalho. São apresentados os objetivos do trabalho que se pretende desenvolver. De forma a permitir um melhor enquadramento nos capítulos seguintes são apresentados os diagramas de blocos do sistema, cujos detalhes de implementação são abordados nos capítulos seguintes.

No **capítulo 2** descreve-se toda a pesquisa relacionada com os sistemas de conversão de fala para língua gestual de modo a auxiliar o desenvolvimento do trabalho desta dissertação.

No **capítulo 3** é constituído por 3 secções principais: reconhecimento da fala, máquina de tradução (**MT**) e *interface* gráfica. Na secção do reconhecimento da fala (secção **3.1**) faz-se um levantamento das alternativas possíveis que possibilitem a tradução de fala em **LP** para texto em **LP**. Apresenta-se ainda a escolha do reconhecedor de fala, bem como os testes experimentais realizados e as suas conclusões. A secção **3.2** inicia-se pelo estudo da **LGP** e apresenta-se um estudo das alternativas possíveis de serem utilizadas para o desenvolvimento da **MT**. Faz-se uma análise dessas alternativas, apresenta-se a escolha e descreve-se todo o desenvolvimento, os testes realizados e as conclusões da **MT** deste

trabalho que faz a conversão de texto em **LP** para texto em **LGP**. Na secção **3.3** apresenta-se todo o trabalho desenvolvido na *interface* gráfica que faz a conversão de texto em **LGP** para um vídeo final. Apresenta-se ainda os testes realizados e as conclusões.

No **capítulo 4** apresenta-se o *website* com os 3 módulos que constituem o sistema, os testes experimentais efetuados, apresentam-se as opiniões de alguns surdos, intérpretes, especialistas e público em geral sobre o sistema e suas conclusões.

Finalmente, no capítulo (**capítulo 5**), anuncia-se as conclusões finais do trabalho desenvolvido e o trabalho futuro que poderá ser realizado.

Capítulo 2

Estado da Arte

A **LGP** foi reconhecida como uma das línguas oficiais de Portugal em 1997 [2]. Apesar desse reconhecimento a **LGP** não é comum integrar a oferta formativa das escolhas ao nível das línguas. Geralmente, o ensino da **LGP** é realizado em centros de formação ou em escolas de referência onde se ensinam as duas línguas [2], por isso, existem poucos ouvintes a saber a **LGP**.

Como apenas uma minoria dos ouvintes sabe a **LGP**, a comunidade surda portuguesa possui diversas dificuldades para se integrar no mundo orientado para os ouvintes. Este problema de integração da comunidade surda não está apenas em Portugal mas também nos outros países, assim a comunidade científica tem abordado este problema e apresentado algumas soluções.

Neste capítulo faz-se uma revisão da literatura, procurando evidenciar o estado da arte em projetos que usam quer a **LGP** quer línguas gestuais de outros países.

2.1 Projetos Inclusivos que usam a LGP

Têm surgido ao longo dos anos várias iniciativas empresariais e académicas com objetivo de aproximar a comunidade surda portuguesa e os ouvintes portugueses.

Em 2005 a rádio Antena 1, depois em 2013 a rádio TSF, emitiram em direto na Internet vários conteúdos informativos, tais como, relato de um jogo de futebol, notícias e música traduzidos para **LGP** através de intérpretes [9–11].

Em 2011, desenvolvido pela *ZonAdvanced* surge o *Serviin*. Caracteriza-se por ser um serviço que permite fazer a comunicação entre surdos e empresas/entidades aderentes ao serviço através da utilização de um intérprete presente na empresa *Serviin* [12]. O surdo realiza uma videochamada para o *call center* da *Serviin* onde se encontra um intérprete de **LGP**, este redireciona a chamada para empresa ou entidade aderente que o surdo

quer contactar e é estabelecida uma comunicação bidirecional entre o surdo, a assistente e a empresa aderente [12]. Este serviço já conta com diversas empresas e entidades aderentes, nomeadamente, empresas de telecomunicações, escolas, comércio local, táxis, INEM, bombeiros, PSP e GNR [13].

Em 2008 a RTP passou a transmitir os noticiários com legendagem automática que é útil para deficientes auditivos [14]. Depois em 2011, passou a disponibilizar conteúdos em LGP, sejam de programas emitidos com intérprete incluído na emissão num dos cantos inferiores do ecrã, sejam programas emitidos com intérprete em grande plano (na RTP2 é emitido um programa diário com interpretação em língua gestual com o intérprete em grande plano) sejam programas emitidos em direto, pela Internet, em duplo ecrã [15, 16].

A Vodafone Portugal, cria em 2012, um serviço de apoio a clientes em LGP. Baseia-se numa ligação de uma videochamada por parte dos surdos para o *call center* desta empresa que depois são atendidos por um assistente da Vodafone fluente em LGP [17].

Desde 19 Maio de 2013, todos os domingos, na Igreja da Santíssima Trindade, as missas das 15 horas são acompanhadas por dois intérpretes de LGP. Esta iniciativa parte de uma proposta dos responsáveis do IPL, tratando-se de mais uma ação que integra o projeto IPL(+)*Inclusivo* [18, 19].

Através do projeto IPL(+)*Inclusivo* do IPL foi realizada uma emissão de rádio em direto para LGP através de um intérprete no dia 8 maio de 2013 na rádio IPLay, a rádio do IPL. Além dessa iniciativa, o IPL(+)*Inclusivo* já realizou a tradução para LGP das músicas das tunas do IPL, missas em Leiria e Fátima e peças de teatro³ [20].

Alunas do curso de Jornalismo e Comunicação da Escola Superior de Educação de Portalegre criaram, em outubro de 2013, uma *webradio*, uma rádio na Internet, que se caracteriza pela tradução dos conteúdos jornalísticos para LGP através de intérpretes⁴ [20].

A união europeia também tem feito esforços no sentido de dar a conhecer as línguas gestuais. Financiado pelo programa Leonardo da Vinci, surge o projeto internacional *Spread the Sign*. O *Spread the Sign* é um dicionário *online* de línguas gestuais, multilingue e gratuito. Apresenta alfabetos, vídeos de palavras e frases em 25 línguas gestuais estando incluída a LGP. Este dicionário nasceu com o objetivo de melhorar as competências linguísticas dos alunos surdos de escolas vocacionais quando estes viajassem para fora do país em trabalho [21]. Disponibiliza aplicações para o computador, iPhone e Android [21].

³<https://www.youtube.com/user/Radioplay>

⁴<http://radiomaosaconversa.uphero.com/index.php/>

2.2 Projetos vocacionados para o Ensino da Língua Gestual Portuguesa

Nesta secção apresentam-se dois trabalhos académicos portugueses que tiveram como finalidade o ensino da **LGP** através do uso de avatares.

2.2.1 Ensino da Língua Gestual Assistido por Personagens 3D Virtuais

Em 2002, foi desenvolvido um trabalho [22] com a finalidade de divulgar a **LGP**. Para isso o autor criou diversas ferramentas em cada uma com um objectivo diferente: *Construtor 3D*, *Abecedário 3D*, *Animações Víde*o e *Gestuário 2D*.

O *Construtor 3D* tem como finalidade a construção interativa de gestos de **LGP**. O utilizador tem ao seu dispor 3 modelos gráficos (figura 2.1): uma mão esquerda, um avatar (humano virtual) e uma mão direita. Cada um desses modelos dispõe de manipuladores tridimensionais para rotações e translações de forma a compor o gesto. Depois cada gesto criado poderá ser associado a uma palavra, letra ou número. O autor refere que esta ferramenta auxilia a memorização dos gestos, pois cada gesto construído está associado a uma única palavra.

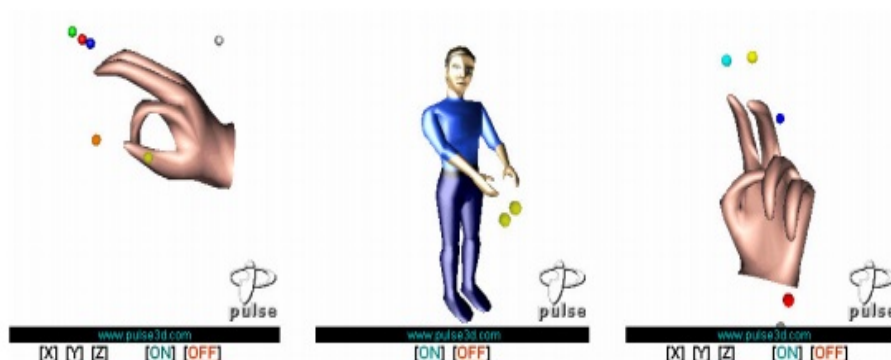


Figura 2.1: *Interface* do Construtor 3D [22].

O *Abecedário 3D* foi desenvolvido a fim de permitir a visualização a três dimensões (3D) de todas as letras do alfabeto. Com esta ferramenta o utilizador observa o modelo do braço e mão de um avatar (figura 2.2), com a possibilidade de ver a animação de cada letra sobre quatro pontos de vista diferentes: frente, cima, do ponto de vista do avatar e uma vista rotativa sobre o braço.



Figura 2.2: *Interface* do Abecedário 3D [22].

Animações Vídeo tem como objetivo a visualização de animações 3D dos verbos da **LGP**. Com esta ferramenta um avatar completo realiza os gestos dos verbos selecionados (figura 2.3).



Figura 2.3: *Interface* do Animações Vídeo [22].

O *Gestúário 2D* apresenta ao utilizador imagens relativas à numeração (figura 2.4). As imagens são acompanhadas com descrições sobre a orientação, localização e movimentos necessários para definir o gesto.



Figura 2.4: *Interface* do Gestuário 2D [22].

O autor refere que o seu sistema precisa de melhor qualidade gráfica de detalhe nos modelos gráficos de forma a ser melhor aceite pela comunidade surda [22].

2.2.2 Avatares em Língua Gestual Portuguesa

Em 2013, foi desenvolvido um outro trabalho [23] com a finalidade de ensinar conceitos básicos de **LGP** a crianças através de um avatar. A aplicação possui diversos temas, tais como: as cores, as profissões, os animais, os verbos e os pronomes pessoais e interrogativos [23].

O autor desenvolveu uma aplicação chamada “Os meus primeiros e-Gestos” que possui um jogo de adivinhar gestos (“Adivinha Gesto”), um dicionário, um compositor de frases e um portal exclusivo aos administradores.

O autor realizou vários testes em vários avatares de modo a criar um avatar com maior aceitação possível pelos surdos portugueses [23]. O avatar que reúne as melhores características desde a aparência, expressões faciais, movimentos corporais e fluidez dos gestos pode ser visualizado na figura 2.5.

A captura dos movimentos corporais e expressões faciais para o avatar foram através da utilização do sensor Microsoft Kinect e utilizou o método de animação por *keyframe* a fim de adicionar o movimento dos dedos devido às limitações da tecnologia de baixo custo (Microsoft Kinect) [23]. A animação por *keyframe* consiste em definir várias posturas ou poses distintas no avatar ao longo de instantes temporais e o *software*, que possui uma função de interpolação, faz as transições de cada articulação do avatar para todas as



Figura 2.5: Avatar final usado pelo autor [23].

frames entre uma dada postura inicial e final [23].

Em todos os tratamentos das animações, bem como na composição de frases o autor utilizou a ferramenta *MotionBuilder*. Ao utilizar o *MotionBuilder* o autor conseguiu concatenar animações, também conhecido como *blend*, para gerar frases [23]. O que resultou numa animação final fluida e consistente [23].

O “Adivinha Gesto” tem como objetivo testar os conhecimentos do utilizador. O utilizador escolhe uma palavra e o sistema apresenta várias animações de gestos, sendo que apenas uma delas corresponde ao gesto dessa palavra (figura 2.6). O utilizador terá que adivinhar a animação correta.

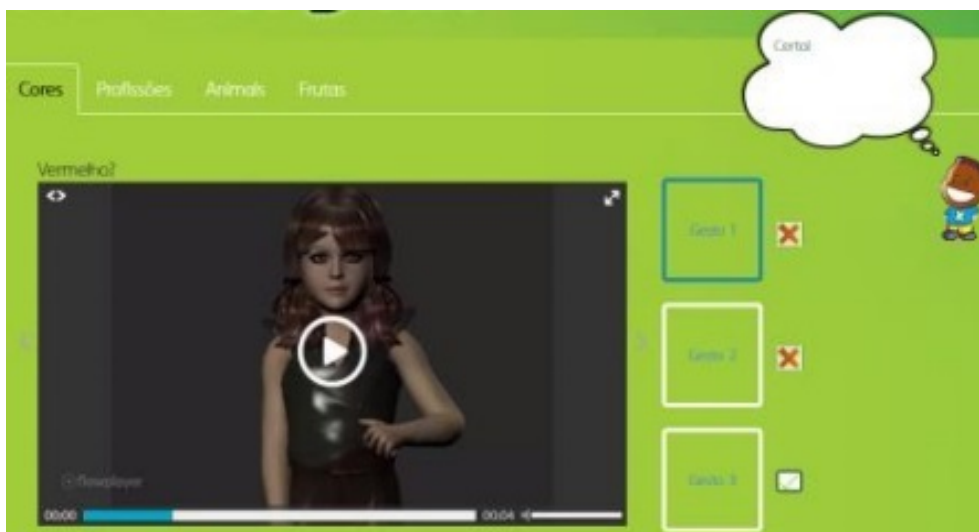


Figura 2.6: Adivinha gesto [23].

O dicionário permite a consulta de palavras e sua tradução em gestos em LGP. Na figura 2.7 pode-se ver o avatar a representar a palavra “verde” em LGP.

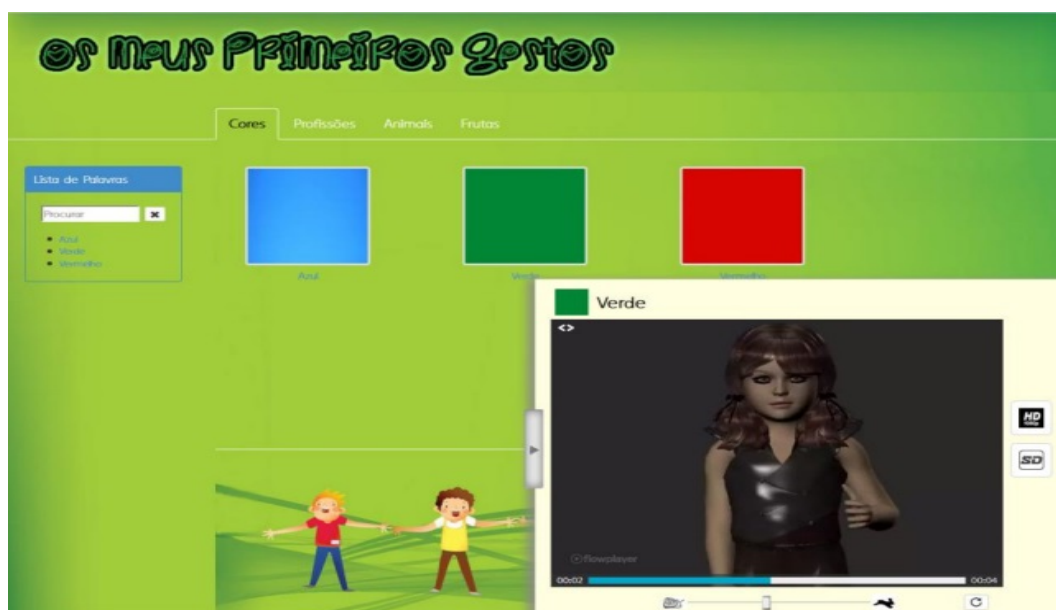


Figura 2.7: Representação em LGP da palavra “verde” [23].

No compositor de frases (figura 2.8), o utilizador deve arrastar uma palavra de cada vez de acordo com a estrutura sintática da LGP de forma a compor as frases. Depois da frase estar completa o avatar faz a sua representação. Caso o utilizador não saiba a estrutura sintática da LGP existem sugestões de frases. Uma limitação do sistema é o facto de só serem possíveis representações de frases que constem da base de dados do sistema [23].



Figura 2.8: Composição de frases [23].

No portal exclusivo aos administradores é possível adicionar, remover, editar gestos, palavras, frases e utilizadores.

Os resultados da avaliação da usabilidade do sistema foram bons, destaque para o compositor de frases em que foi considerado útil e fácil de usar [23]. Na avaliação da qualidade do avatar e gestos os resultados também foram bons [23].

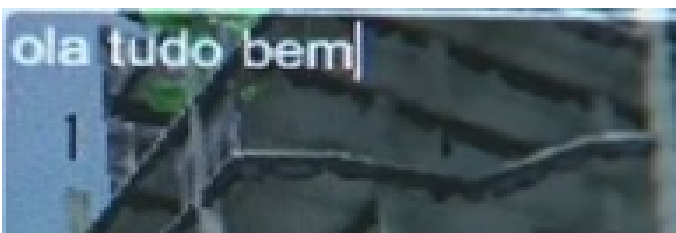
2.3 Sistemas Automáticos de Tradução de Fala para Línguas Gestuais

A investigação na área dos tradutores automáticos de fala para língua gestual é relativamente recente, sendo que, a maioria dos trabalhos surge na última década. Atualmente, existem vários sistemas um pouco por todo o mundo tais como: nos Estados Unidos da América [24–26], no Brasil [27–30], na Europa [31–37], na Malásia [38], na China [39] e no Japão [40] a funcionarem em diversas plataformas. Nesta secção apresentam-se detalhadamente alguns tradutores automáticos já existentes, por fim é realizado um resumo da literatura de modo a auxiliar o desenvolvimento do trabalho desta dissertação.

2.3.1 Virtual Sign Project

No que respeita à **LGP**, que seja do nosso conhecimento, existe somente um tradutor automático. Esse projeto designa-se por *Virtual Sign* [41,42] e tem como objetivo auxiliar a comunicação, de forma bidirecional entre surdos e ouvintes em contexto de sala de aula. Existe bidirecionalidade pois o sistema permite a tradução de texto para gestos e de gestos para texto.

A tradução de texto para gestos em **LGP** é através de um avatar. O utilizador ouvinte escreve a mensagem que pretende (num computador) e o avatar reproduz em **LGP** de modo a que o surdo a perceba (figura 2.9).



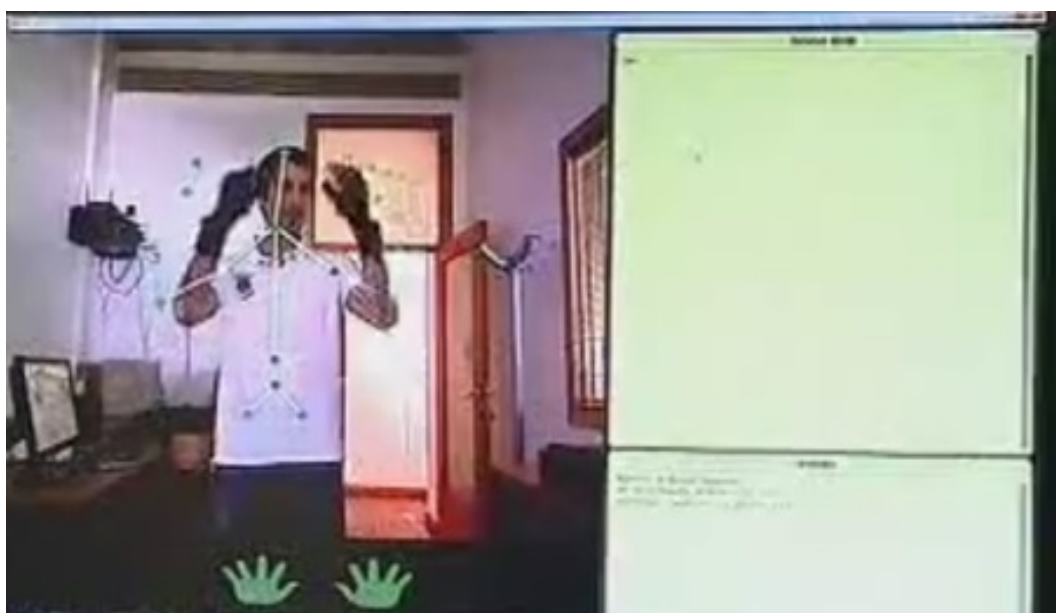
(a) texto [42]



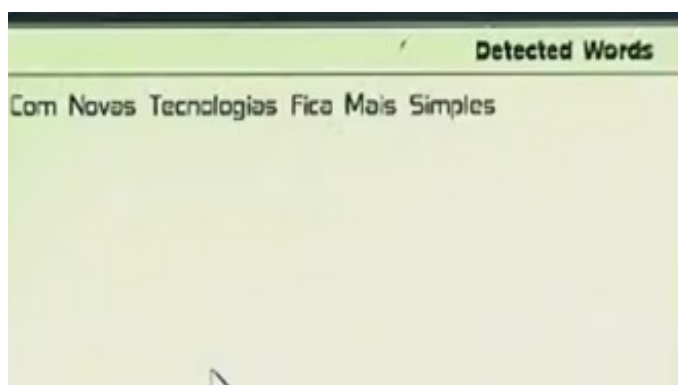
(b) gestos [42]

Figura 2.9: Exemplo de tradução de texto para gestos LGP do projeto *Virtual Sign*.

Para a tradução de gestos para texto os autores utilizaram o sensor *Microsoft Kinect* e duas luvas “5DT Data Glove” com 14 sensores (figura 2.10). A utilização do sensor *Microsoft Kinect* serviu para capturar os movimentos corporais e expressões faciais, enquanto que as luvas foram usadas para estimar a posição da mão e dos dedos.



(a) gestos [42]



(b) texto [42]

Figura 2.10: Exemplo de tradução de gestos LGP para texto do projeto *Virtual Sign*.

Os autores referem que os resultados deste sistema se revelam promissores. No entanto, o facto de não incluir a tradução diretamente a partir da fala é uma limitação à sua utilização.

Apresentam-se de seguida alguns tradutores automáticos desenvolvidos para outras línguas gestuais que não a portuguesa.

2.3.2 iCommunicator

Em 1999 a *Interactive Solutions, Inc.* começou o desenvolvimento do primeiro sistema automático de fala para ASL através de vídeos, o *iCommunicator*. Atualmente este sistema encontra-se na quinta versão e tem um custo de USD6,499 [25]. Este sistema permite às pessoas com grande dificuldade de audição a tradução em tempo real de:

- voz para texto;
- voz/texto para língua gestual;
- voz/texto para voz gerada pelo computador (síntese).

Este sistema [25] está a ser usado nas instituições de ensino, governamentais e de saúde e nas empresas dos Estados Unidos da América e Canadá. A utilização deste sistema requer um microfone ligado a um transmissor *wireless* e um receptor *wireless* ligado ao computador onde o sistema está instalado. Na figura 2.11 mostra-se este sistema em contexto de sala de aula e sua aparência no ecrã do computador.



(a) Uso em contexto de sala de aula [25]



(b) Aparência no ecrã do computador [25]

Figura 2.11: iCommunicator

O reconhecimento da fala para texto é realizado pelo *software Dragon Naturally Speaking* e depois a tradução para ASL é através da apresentação consecutiva dos vídeos que se encontram numa base de dados que o sistema possui [25, 43].

2.3.3 TESSA

Em 2002, surge um outro sistema experimental de tradução, o TESSA [31] com o objetivo de auxiliar a comunicação entre uma pessoa surda e o balconista dos correios através da tradução automática do discurso deste para língua gestual britânica (*British Sign Language (BSL)*) através de um avatar. O sistema apenas responde às falas do balconista

quando as frases proferidas se encontram na base de dados. Cada frase tem associada uma sequência de gestos animada por um avatar (ver figura 2.12).

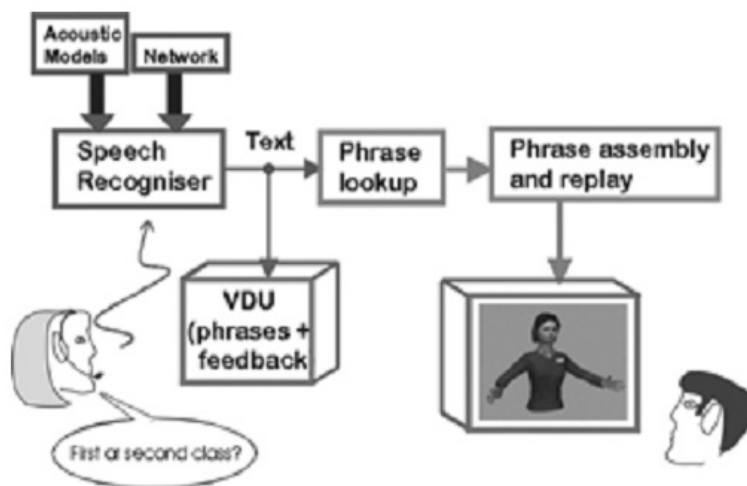


Figura 2.12: Diagrama do sistema TESSA [31]

O sistema utiliza um reconhecedor entrópico para converter fala para texto e usa um avatar para a *interface* gráfica [31].

A conversão de texto para gestos é feita através do método da procura da frase (em inglês designa-se por *phrase lookup*). O método da procura da frase consiste em a frase reconhecida ser procurada numa base de dados à qual está associada uma sequência de gestos, caso não exista correspondência o sistema não consegue fazer a tradução [31].

No reconhecimento dos gestos o sistema apresentou razoáveis resultados para um conjunto de frases, mas a avaliação global do sistema revelou-se muito baixa [31]. Como principais pontos a serem melhorados os surdos focam o melhoramento da aparência e das expressões faciais do avatar e a diminuição do tempo entre o fim da fala e o início da representação dos gestos [31].

2.3.4 Speech to sign language translation system for Spanish

Em 2008 foram apresentados os primeiros testes efetuados a um sistema automático de tradução para a língua gestual espanhola (*Lengue de Signos Española (LSE)*) através de um avatar [32]. Este sistema tem como objetivo ajudar os funcionários no atendimento a surdos quando estes iam solicitar ou renovar o seu Cartão de Cidadão. Na figura 2.13 apresenta-se o diagrama deste sistema.

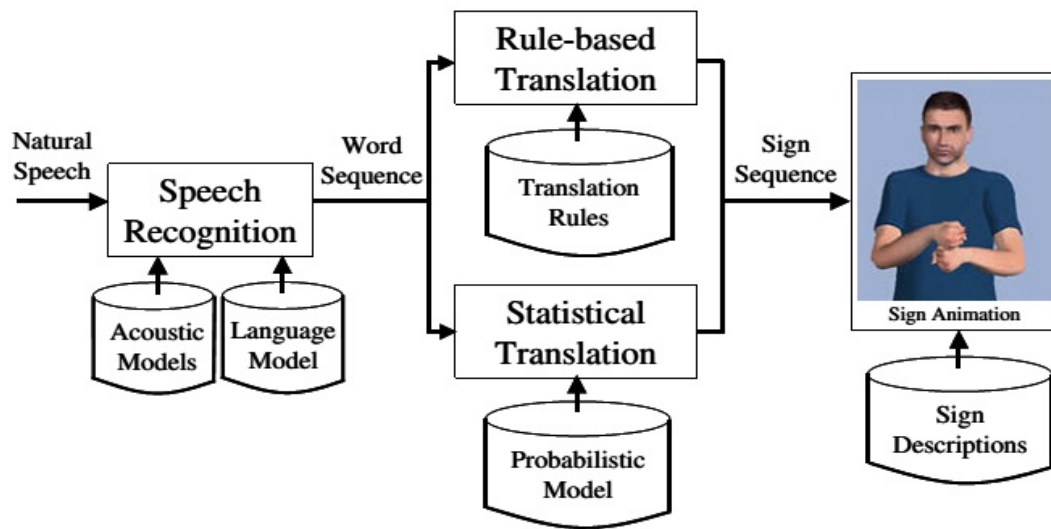


Figura 2.13: Diagrama do sistema “*Speech to sign language translation system for Spanish*” [32]

O sistema é constituído por 3 grandes blocos: um bloco de reconhecimento da fala (*speech recognition*), um bloco de tradução de texto em castelhano para a estrutura frásica da LSE e um bloco de animação.

O reconhecedor da fala foi desenvolvido pelo *Grupo de Tecnología del Habla da Universidad Politécnica de Madrid (GTH-UPM)* e utiliza modelos acústicos de linguagem de modo a encontrar o melhor texto corresponde à fala proferida [32].

No bloco de texto para estrutura frásica da LSE foram usadas dois métodos em paralelo: um baseado em regras e outro recorrendo a uma abordagem estatística (em inglês designam-se por *rule-based* e *statistical*, respetivamente).

O método por regras usa análises morfológica, sintática e semântica das línguas de origem e destino e é definido um conjunto de regras que permitem a tradução da estrutura da frase na língua origem para a língua destino. Neste caso, a língua origem é o castelhano e a língua destino é a LSE [32, 44].

O método por estatística baseia-se na teoria de decisão estatística e na aprendizagem estatística. Ao usar este método os sistemas têm que ser treinados usando um conjunto de pares de frases, ou seja, frases representativas da língua fonte e a suas frases correspondentes da língua destino [32, 44].

Comparando o desempenho de ambos os métodos (métodos por regras e por estatística) os autores concluíram que a utilização do método por regras garantia melhores resultados de tradução o que pode ser justificado pelo facto de ter sido usada uma pequena quantidade de dados para treinar os modelos estatísticos. No bloco de animação é usado um avatar para gerar os gestos.

O sistema evoluiu e em 2010 foi apresentado o *Advanced Speech Communication System*

for Deaf People [33] que é um sistema automático de renovação da carta de condução.

O diagrama de blocos é semelhante ao apresentado na figura 2.13. O bloco de tradução de texto é agora composto por três métodos: por exemplos (em inglês designa-se por *example-based*), por regras e por estatística. O método por exemplos baseia-se na analogia, ou seja, o sistema é treinado com frases representativas da língua fonte e suas correspondências na língua de destino e só conseguirá traduzir caso as frases a traduzir sejam semelhantes às frases representativas [33, 44].

Os autores definiram uma distância limite (*distance threshold*) para que o bloco de tradução de texto efetue ou não a tradução usando o método por exemplos. Se a distância mínima da sequência de palavras aos exemplos estiver abaixo do limiar definido pelos autores, o sistema realiza a tradução pelo método por exemplos, caso contrário a tradução será realizada por regras ou por estatística. A escolha entre regras e estatística está relacionada com um rácio ($\#glosses\#words$) entre o número de gestos (grados depois do processo de tradução) e o número de palavras da sequência. Caso esse rácio ($\#glosses\#words$) seja superior ou igual que um novo valor de referência (*threshold*) definido pelos autores, então a tradução é realizada pelo método por regras, caso contrário a tradução é pelo método por estatística. Esta descrição pode ser visualizada pela figura 2.14.

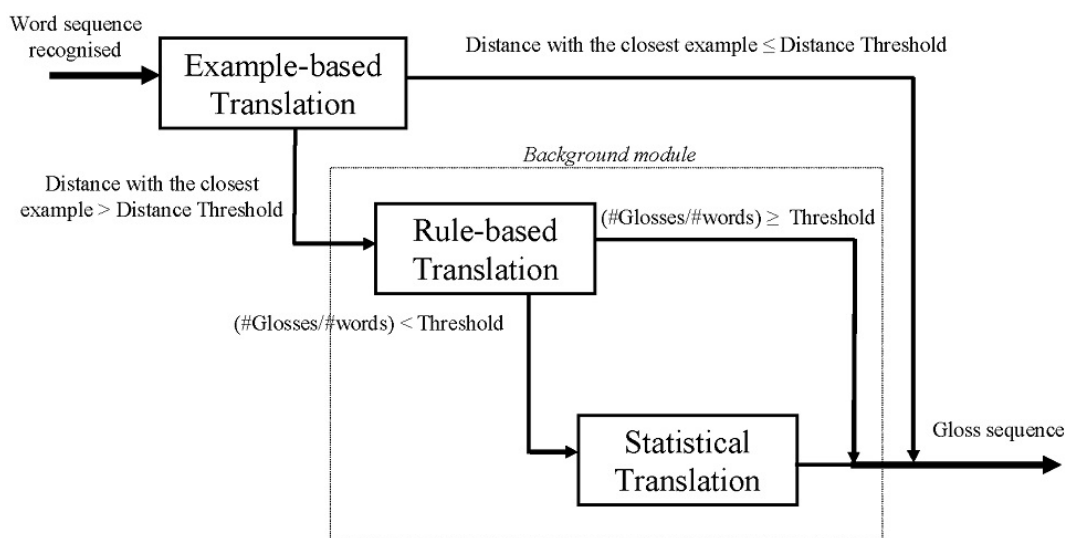


Figura 2.14: Diagrama da Máquina de Tradução do módulo tradução da fala para LSE de “Advanced Speech Communication System for Deaf People” [33]

Os autores verificaram que o sistema utiliza o método por exemplos em 95% dos casos, 4.1% das vezes usa o método por regras e apenas 0.9% das vezes é usado o método por estatística.

Os resultados dos questionários efetuados à avaliação do sistema foram pouco satisfatórios, apesar dos autores terem verificado que o sistema possui uma baixa taxa de erros

no reconhecimento de fala e a tradução de castelhano para a estrutura da **LSE** estar bem sucedida. Apontam como pontos a melhorar a naturalidade do avatar e aumentar o número de gestos padronizados da **LSE**, visto que as discrepâncias são interpretados pelos surdos como erros.

2.3.5 Speech Translation into Pakistan Sign Language

Em 2012 dois alunos da Escola de Computação do Instituto Politécnico de *Blekinge* da Suécia [34] apresentaram, na sua tese de mestrado, um sistema automático de tradução de fala/texto em inglês para língua gestual paquistanesa (*Pakistan Sign Language (PSL)*) através de vídeos. O diagrama da arquitetura deste sistema apresenta-se na figura 2.15.

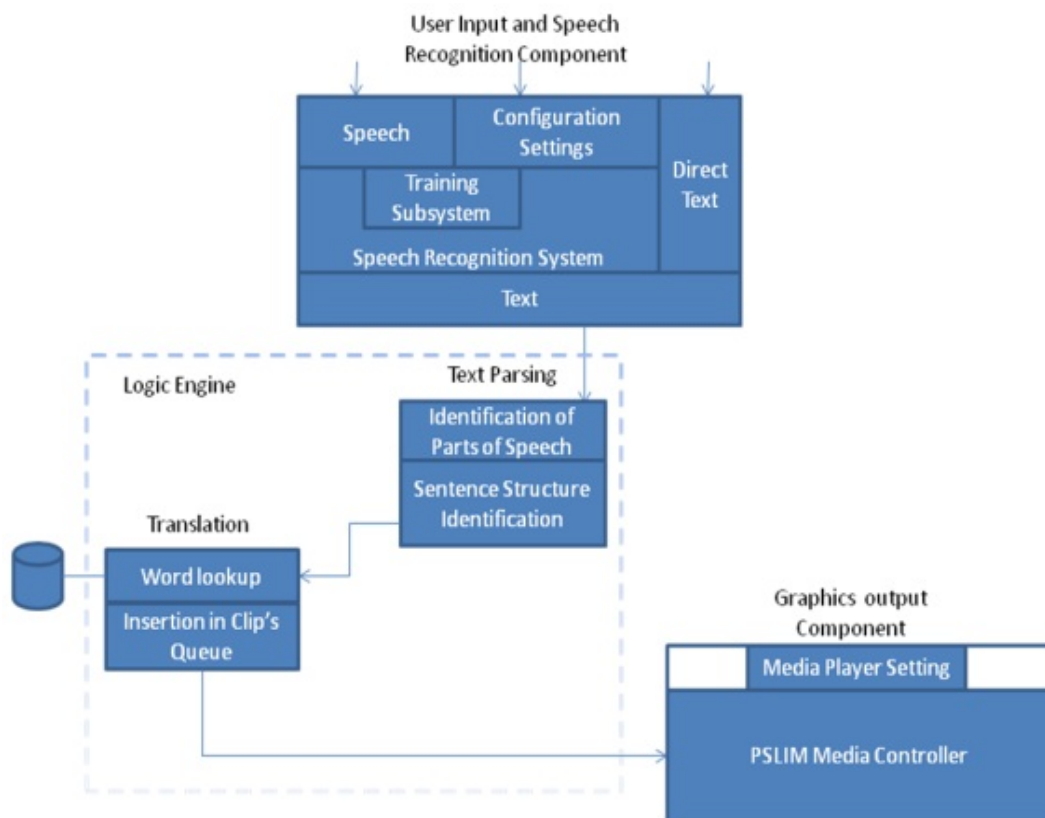


Figura 2.15: Diagrama do sistema “*Speech Translation into Pakistan Sign Language*” [34]

O sistema é constituído por 3 blocos principais: bloco de escrita e reconhecimento da fala (*User input and Speech Recognition Component*), bloco de lógica (*Logic Engine*) e bloco de apresentação dos vídeos (*Graphics output component*).

O primeiro bloco possibilita a escrita de uma mensagem ou então usar o reconhecimento de fala que faz a conversão do discurso para texto [34]. Em ambas as alternativas, o texto será enviado para bloco de lógica. Para o reconhecimento da fala os autores usaram a ferramenta *Microsoft Speech SDK v5.0*.

No bloco de lógica é onde é realizada a tradução de texto para estrutura da PSL [34]. O texto recebido é primeiramente estruturado de acordo com as regras da PSL (bloco *Text Parsing*) e depois enviado para o bloco de tradução (bloco *Translation*). No bloco de tradução, cada palavra do texto estruturado é usada para procurar o seu vídeo respetivo na base de dados do sistema [34]. Depois de o sistema encontrar a localização do vídeo na base de dados, o sistema insere o caminho da localização desse vídeo numa lista a ser enviada para o bloco de apresentação dos vídeos [34]. A lista está ordenada de acordo com o texto estruturado previamente e encontra-se na componente *Insert in Clip's Queue* do bloco de tradução [34].

No último bloco (*Graphics output component*) é onde os vídeos são reproduzidos de acordo com a lista previamente realizada.

Os autores verificaram que o sistema possui uma eficiência total de 61%, sem qualquer erro quando não se realiza o treino de fala no bloco *User input and Speech Recognition Component* [34]. Caso contrário o sistema garante uma eficiência total de 78%, sendo a restante percentagem erros no reconhecimento de fala (cerca de 5%) e no bloco de lógica (cerca de 17%) [34].

2.3.6 Falibras

Em 2002 começou o desenvolvimento deste sistema automático de fala em Português do Brasil para Língua Brasileira de Sinais (**LIBRAS**) através de vídeos. O objectivo deste sistema era facilitar a comunicação entre ouvintes e surdos e apoiar a formação de intérpretes [27]. Desde essa altura tem vindo a ser melhorado e em 2012 apresenta novas funcionalidades [27].

Na primeira versão era constituído por dois módulos principais: módulo interpretador e módulo de exibição. No módulo interpretador era possível fazer a conversão de fala em Português do Brasil para texto ou escrever texto diretamente a ser enviado para o módulo seguinte. Nessa versão, o módulo de interpretação fazia uma interpretação direta das palavras e pequenas expressões para **LIBRAS** e ignorava o contexto das frases. O módulo de exibição tinha como objetivo exibir os vídeos de acordo com a **LIBRAS** [27].

Entre os anos 2003 e 2004 foi melhorada a qualidade tradução. Para isso os autores adicionaram ao sistema a análise morfológica ao texto reconhecido [27]. A função da análise morfológica é classificar as palavras de acordo com a sua categoria gramatical (por exemplo: substantivo e verbo) e sua flexão (por exemplos: número, género e pessoa) [27].

Nos anos 2005 e 2007, os autores adicionaram ao sistema a análise sintática ao texto reconhecido. Nesta versão o sistema contava para a tradução com 4 módulos: analisador lexical, analisador sintático, analisador de contexto e gerador de tradução em **LIBRAS** [27].

O analisadores de léxico, sintático e contexto faziam as análises morfológica, sintática e verificação de concordância do gênero e número do texto reconhecido, respetivamente [27]. O gerador de tradução em **LIBRAS** fazia a conversão do texto para **LIBRAS** [27].

Nessa altura foi desenvolvida outra abordagem para a tradução. Como na versão anterior as regras sintáticas e de tradução estavam fixas e especificadas diretamente na aplicação, o que apresentava restrições práticas de desempenho, a tradução passou a ser realizada numa memória de tradução [27]. Utilizando essa abordagem, a tradução consulta uma base de dados que contém exemplos de traduções feitas por um tradutor humano e fará a tradução se os textos forem similares [27].

Atualmente, o sistema foi totalmente refatorado⁵ e apresenta diversas funcionalidades: animação 3D da **LIBRAS**, módulo de tradução mista, módulo de apoio a intérpretes e módulo de acessibilidade *web* [27].

A animação 3D da **LIBRAS** tem como objetivo a tradução de textos escritos e falados e sua representação através dum avatar [27].

O módulo de tradução mista tem como finalidade a tradução para **LIBRAS**. Usa duas abordagens: transferência sintática e memória de tradução. O texto é primeiramente avaliado se é uma exceção à regra. Se não for verifica se existem exemplos de tradução que possam servir de analogia para a tradução do texto na base de dados anteriormente descrita [27]. Nos casos em que se verifica a inexistência de exceções ou regras que interfiram na estrutura de tradução do texto de entrada, é utilizada a tradução por transferência sintática [27].

O módulo de apoio a intérpretes possui como objetivos: facilitar a preparação de intérpretes **LIBRAS** e aperfeiçoar a qualidade de tradução do sistema [27]. Este módulo permite ao intérprete carregar o texto a ser interpretado e verificar as regras sintáticas e as regras de tradução para esse texto [27]. Na ausência de regras, o intérprete pode adicionar novas regras sintáticas e/ou de tradução [27].

O módulo de acessibilidade *web* permite a conversão de conteúdos que se encontram na Internet em português para **LIBRAS**. Esta ferramenta apresenta-se como uma extensão (*add-on*) para o navegador *Firefox*. Esta extensão serve para 2 propósitos: traduzir textos da *web* de português/**LIBRAS** ou **LIBRAS**/português e aceder a dicionários; o outro propósito é incrementar o dicionário de **LIBRAS** e de português e fornecer novos exemplos de tradução para aperfeiçoar a qualidade do resultado [27].

Apesar destas funcionalidades que o sistema apresenta atualmente, os autores não apresentaram nenhuma avaliação do seu sistema e não referem qual o reconhecedor de fala utilizado.

⁵Refatoração: conjunto de alterações feitas na estrutura interna de um software de forma a torná-lo mais fácil de perceber sem modificar o seu comportamento observável [45]

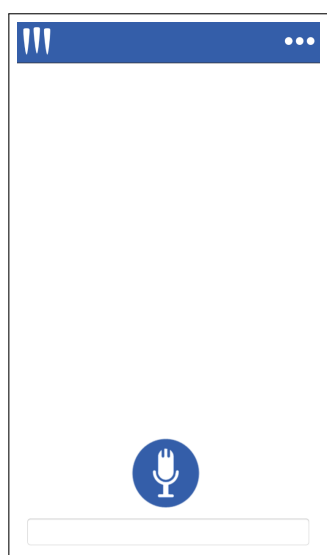
2.3.7 Tradutores Automáticos em Aplicações Móveis

Atualmente já se pode encontrar sistemas automáticos de fala para língua gestual em plataformas móveis. Essas aplicações são: *Mimix - Speech to sign language* [26], *ProDeaf* [28] e *Hand Talk* [29]. Todas elas receberam prêmios nacionais e internacionais.

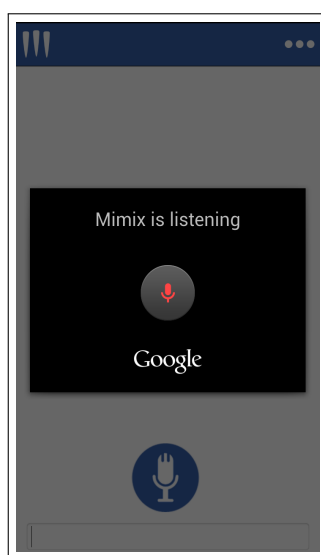
O *Mimix - Speech to sign language* [26], lançado no final de dezembro de 2010, tem como principal funcionalidade a tradução de fala/texto em inglês para ASL através dum avatar. Atualmente esta aplicação apenas se encontra na plataforma Android [26]. A usabilidade deste sistema é bastante simples pois o sistema utiliza apenas uma página (figura 2.16(a)) onde se realiza a tradução de fala/texto em inglês e sua representação gestual através do avatar .

Quando o utilizador quer efetuar uma tradução de fala para gestos o sistema necessita obrigatoriamente de estar conectado à internet pois o sistema utiliza o reconhecedor de fala da Google [26]. O reconhecedor de fala da Google faz a captura do sinal acústico, descodifica-o no servidor da Google e devolve para esta aplicação uma sequência de palavras (texto) que depois é representada através do avatar, os seus gestos correspondentes [26].

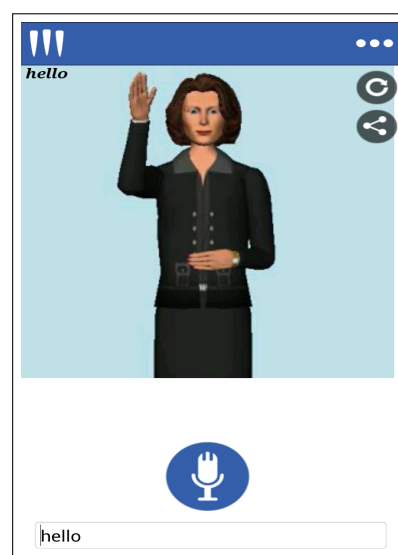
Para o sistema traduzir da fala para gestos o utilizador deve: carregar no botão microfone (figura 2.16(a)) para usar o reconhecedor da fala da Google, depois discursar (figura 2.16(b)) e por fim ver o avatar com a representação da sua mensagem em ASL (figura 2.16(c)).



(a) Mimix na primeira utilização.



(b) Mimix: reconhecimento da fala.



(c) Mimix: representação do gesto "Hello" ("Olá").

Figura 2.16: Mimix: primeira utilização, reconhecimento da fala e representação do gesto "Hello" ("Olá") previamente reconhecido.

O *ProDeaf* [28] foi criado em 2010 tendo como principais funcionalidades a tradução de fala/texto em Português do Brasil para **LIBRAS** através do avatar. Atualmente esta aplicação pode ser descarregada nas plataformas Android e iOS.

Este sistema foi testado pelo autor deste trabalho assim como o *Mimix*, considera-se o *ProDeaf* bastante superior uma vez que existe uma melhor qualidade gráfica nas expressões faciais e movimentos corporais do avatar e ainda permite a rotação do avatar em cerca de 180 graus, de forma a permitir que o utilizador veja os gestos em diferentes pontos de vista.

A usabilidade deste sistema é muito semelhante ao *Mimix* na medida que apenas utiliza uma única página para efetuar o reconhecimento da fala em Português do Brasil e depois apresenta a sua tradução em **LIBRAS** com o avatar.

À semelhança do *Mimix* foi utilizado o reconhecedor de fala da Google, o que significa que o sistema necessita obrigatoriamente da ligação à Internet para fazer a conversão de fala para texto [28].

Ao contrário do *Mimix* em que o avatar só aparecia depois do reconhecimento da fala ou do texto escrito pelo utilizador neste sistema o avatar está sempre presente.

Na figura 2.17 apresenta-se a representação do gesto “Olá” previamente reconhecido.



Figura 2.17: ProDeaf: representação do gesto “Olá”.

O *Hand Talk* foi lançado em julho de 2013, considerado pela ONU como a melhor aplicação social do mundo [29], que converte conteúdos escritos, falados ou fotografados em Português do Brasil para **LIBRAS**. Esta aplicação pode ser descarregada nas plataformas Android, iOS ou BlackBerry.

Foi descarregado este sistema, da plataforma Android, concluiu-se que o avatar é muito semelhante ao *ProDeaf*, pois apresenta o mesmo nível de cuidado nas expressões faciais e movimentos corporais e possui liberdade de rotação do avatar, em 360 graus, de forma a permitir que o utilizador veja os gestos em todos pontos de vista possíveis.

À semelhança do *Mimix* e do *ProDeaf*, este sistema apresenta também uma usabilidade muito simples, visto que utiliza uma única página para efetuar o reconhecimento da fala em Português do Brasil e depois apresenta a sua tradução em **LIBRAS** pelo avatar.

Tais como *Mimix* e *ProDeaf*, este sistema utiliza também o reconhecedor de fala da Google o que significa que este sistema requer sempre da ligação à Internet para fazer a conversão de fala em Português do Brasil para texto [29].

De igual modo ao *ProDeaf*, o avatar deste sistema está sempre presente, mesmo que não se realize nenhum reconhecimento de fala.

Na figura 2.18 apresenta-se a representação do gesto “Olá” previamente reconhecido.



Figura 2.18: Hand Talk: representação do gesto “Olá”.

2.4 Sumário

De acordo com os sistemas analisados na literatura pode-se concluir que, genericamente, um sistema de tradução automático de fala para língua gestual é constituído por 3 módulos principais:

- um reconhecedor de fala: parte do padrão acústico e fornece na sua saída a sequência de palavras reconhecidas;

- uma máquina de tradução (MT): parte do texto da língua oral e apresenta o mesmo texto segundo a estrutura frásica da língua gestual;
- uma *interface* gráfica: parte da sequência de palavras em língua gestual e faz a sua animação por gestos.

Na figura 2.19 ilustra-se a articulação entre os módulos descritos.



Figura 2.19: Sistema automático de fala para língua gestual genérico.

Em relação ao reconhecedor de fala, a revisão da literatura permite concluir que alguns trabalhos utilizam reconhecedores próprios [31–33], enquanto outros recorrem a motores de reconhecimento [25, 26, 28, 29, 34]. Esses motores, por exemplo o *Dragon naturally Speaking* [25], recebem o sinal acústico a reconhecer, descodificam-no e devolvem uma sequência de etiquetas (ex: palavras) que corresponde ao sinal reconhecido. Apesar da aparente facilidade, a utilização de um motor de reconhecimento impõe limitações, que vão desde a possibilidade de terem um custo associado, à dependência da língua entre outras. O desenvolvimento de um reconhecedor, por exemplo um reconhecedor entrópico [31], impõe outro tipo de exigências. O treino dos modelos acústicos obriga à existência de uma base de dados etiquetada que nem sempre se encontra disponível.

Por análise à literatura, verifica-se existirem diversos métodos para as máquinas de tradução (MT), tais como: procura da frase (*phrase lookup*) [31, 34], regras (*rule-based*) [32, 34, 44, 46], estatística (*statistical*) [32, 33, 44] e exemplos (*example-based*) [33, 44, 46].

O módulo de animação (*interface* gráfica) segue tipicamente uma de duas abordagens: ficheiros de vídeos (*videos clips*) ou avatar.

O uso da abordagem por ficheiros de vídeo consiste na utilização de uma base dados que contém vários vídeos com as representações dos gestos realizados por um gestuante humano, procede-se à concatenação desses vídeos formando uma representação das frases completas [23, 31, 34].

Avatar é uma personagem virtual, ou seja, um humano virtual [23, 31–33]. Para que essa personagem virtual tenha a capacidade de representar gestos é necessário a utilização de *software* específico para esse fim. Esse *software*, por exemplo o *MotionBuilder* [23],

é possível “dar vida” e definir o aspeto do avatar. Ao processo de “dar vida” ao avatar designa-se por animação do avatar [23].

O uso da abordagem por avatar consiste na criação de várias animações com as representações dos gestos pelo avatar, guardá-las numa base de dados e por fim concatená-las de modo a representar frases completas [23, 31, 32].

Para qualquer alternativa que se escolha existem vantagens e desvantagens que podem ser visualizadas na tabela 2.1.

Tabela 2.1: Comparação das *Interfaces Gráficas* [23, 31, 34, 41].

	Vídeos	Avatar
Vantagens	Tempos de gravações de vídeos são rápidos	Animações de gestos podem ser manipuladas e reutilizadas mais facilmente
	Surdos preferem vídeos em vez de avatares	Possibilita troca de intérpretes virtuais, troca de roupa e adereços
		Atualmente a captura de movimentos e expressões faciais para o avatar podem ser realizadas por tecnologia barata - Microsoft Kinect
Desvantagens	Sistema fica dependente do intérprete. A troca leva à refilmagem dos vídeos	Tempos de produções das animações elevados.
	Concatenação de vídeos são intrusivos, tornando a leitura aborrecida e difícil	Requer boas qualidades gráfica, animações faciais e corporais para ser aceite pelos surdos
		Se optar pelo uso da Microsoft Kinect, que não possui reconhecimento nas articulações dos dedos, requer o desenvolvimento de um bom algoritmo que faça essa funcionalidade ou então utilizar luvas de dados, como, 5DT Data Glove
		A concatenação de animações requer a utilização programas específicos que faça as transições de animações, por exemplo, MotionBuilder, caso contrário surgem erros nas transições de animações à semelhança do método pelos vídeos.

Capítulo 3

Desenvolvimento modular do sistema implementado

Este capítulo é constituído por 3 secções principais: reconhecimento da fala, máquina de tradução (MT) e *interface* gráfica.

Relativamente à secção do reconhecimento da fala serão abordadas as questões que levaram à escolha do motor de reconhecimento automático de fala, testes de desempenho do mesmo e conclusões. Esta secção constitui o módulo 1 deste trabalho e terá como finalidade a conversão de fala em LP para texto em LP (figura 1.2).

A secção da MT inicia-se com um estudo da LGP que permita a conversão de texto em LP para texto em LGP. Apresenta-se a forma de implementação da MT, os seus testes e conclusões. Esta secção constitui o módulo 2 deste trabalho e terá como finalidade a conversão do texto em LP proveniente do reconhecimento da fala para texto em LGP (figura 1.2).

A secção da *interface* gráfica inicia-se pela escolha do tipo de *interface* gráfica a ser usada, com base nessa escolha, apresenta-se todo o desenvolvimento, testes e conclusões da *interface* gráfica que constitui este sistema. Esta secção constitui o módulo 3 deste trabalho e terá como finalidade a geração de um vídeo final de acordo com texto LGP recebido da MT deste trabalho (figura 1.2).

3.1 Módulo 1: Reconhecimento de Fala

O reconhecimento automático de fala, em inglês designa-se por *Automatic Speech Recognition* [47], consiste em mapear um sinal acústico, capturado por um transdutor (usualmente um microfone) num conjunto de palavras, ou seja, transcrever uma sequência de palavras mais provável que corresponda a esse sinal acústico [47–49].

O campo do reconhecimento de fala recai em várias disciplinas para a construção dos seus modelos de reconhecimento e para a implementação das suas técnicas e algoritmos [47–49]. Estas disciplinas incluem entre outras, o processamento digital de sinal, o reconhecimento estatístico de padrões e a linguística [49].

Um modelo genérico de um reconhecimento de fala pode ser visualizado na figura 3.1.

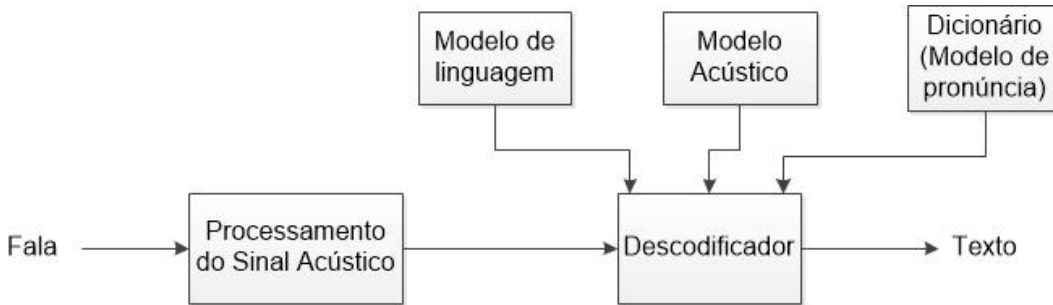


Figura 3.1: Diagrama de blocos genérico para um sistema de reconhecimento de fala [50].

Atendendo à figura anterior, a fala que é capturada pelo “Processamento do Sinal Acústico” vai ser convertida em vetores de observação que serão usados no treino e no reconhecimento [50]. O descodificador tem como funcionalidade em encontrar o melhor texto correspondente à fala proferida de acordo com os vetores de observação ou através do uso dos modelos acústicos e de linguagem [50].

Durante o reconhecimento da fala, os padrões temporais das palavras são variáveis [50]. As variações das palavras estão representadas no modelo acústico [50]. O modelo acústico tem que ser treinado com uma grande quantidade de dados de modo a modelar essa variabilidade e obter um modelo mais robusto possível [50].

É necessário a utilização de um modelo de linguagem de modo a representar a sequência de palavras proferidas [50]. Para tarefas que envolvam pouco vocabulário apenas é preciso a criação de uma rede de estados finitos de modo a procurar várias alternativas específicas para o vocabulário [50]. Já para tarefas que envolvam uma grande quantidade de vocabulário é preciso o uso de um texto de treinamento para criar o modelo de linguagem [50].

Certos sistemas de reconhecimento de fala utilizam também o dicionário (figura 3.1). O dicionário contém as pronúncias das palavras [50]. Esse modelo contém uma sequência de unidades de palavras com os modelos acústicos que abrangem cada palavra do vocabulário [50].

Como no presente trabalho não existe uma base de dados a fim de treinar os modelos de um sistema de reconhecimento de fala, optou-se por utilizar um motor de reconhecimento.

Um motor de reconhecimento é um reconhecedor de fala já implementado em que os utilizadores só precisam de falar e o sistema apresenta a sua conversão em texto. Nestes sistemas os utilizadores não precisam de treinar os modelos acústicos e de linguagem, sendo portanto uma vantagem em comparação a reconhecedores de fala criados de raiz. Apesar dessa vantagem, possui várias desvantagens: poderão não estar preparados para português europeu, as eficiências de conversão de fala para texto serem baixas e plataforma de utilização ser limitada.

Nesta secção serão abordadas as questões que levaram à escolha do motor de reconhecimento automático de fala, testes de desempenho do mesmo e conclusões. Esta secção constitui o módulo 1 deste trabalho e terá como finalidade a conversão de fala em LP para texto em LP (figura 1.2).

3.1.1 Escolha do Reconhecedor de Fala

Como a criação de um reconhecedor de fala requer vários conhecimentos técnicos o que não é trivial de realizar de raiz [48, 49], a escolha recaiu em motores de reconhecimento que permitissem a conversão de fala em LP para texto em LP.

Atualmente existem os seguintes motores de reconhecimento de fala: *Dragon Naturally Speaking* (secção 2.3.2), plataformas móveis (secção 2.3), *Simon* [51–53], *Audimus* [54, 55], *Julius* [56], *SpeechMagic* [57, 58], *Skype Translator* [59], *iSpeech* [60, 61], *Microsoft Speech Platform SDK 11* [62–64] e *Web Speech API* [5, 65–67].

A *Nuance* desenvolveu o *software Dragon Naturally Speaking* que utiliza o seu reconhecedor de fala de modo os utilizadores poderem ditar documentos, pesquisar na Internet, enviar *e-mails* e outras funcionalidades apenas por fala [68]. A empresa criou várias versões desse *software* de modo a ajustar-se às necessidades dos utilizadores mas todas as versões não são passíveis de ser utilizadas no âmbito deste trabalho [68, 69]. Qualquer versão é paga e suporta apenas os idiomas inglês, francês, alemão, espanhol, italiano e alemão [68, 70].

Cada plataforma móvel (*Android*, *iOS* e *Windows Phone*) tem o seu reconhecedor, logo, a escolha da plataforma definia o desenvolvimento do trabalho. Como o sistema não conseguiria albergar todos os surdos, visto que nem todos os surdos usam a mesma plataforma, existem também pessoas que não utilizam *smartphone* ou *tablets*, optou-se por não usar os reconhecedores das plataformas móveis.

O *Simon* é um programa de reconhecimento da fala livre [51]. Para o usar é necessário a utilização de modelos de fala [52]. Os modelos de fala caracterizam a fala do utilizador, isto é, para um conjunto de palavras e frases proferidas pelo utilizador é gerado um

modelo de fala [52]. Para criar este modelo de fala é necessário a criação dos modelos linguístico e acústico [52]. Enquanto que modelo linguístico descreve as palavras e frases gramaticalmente corretas, o modelo acústico descreve a sonoridade das palavras [52]. Atualmente o *Simon* só possui modelos de fala para inglês, alemão e Português do Brasil mas possibilita a criação de modelos da fala de qualquer idioma [53]. Para criar um modelo de fala para este trabalho é necessário utilizar algumas funções do *Hidden Markov Model Toolkit (HTK)* [53]. Como não existe modelo de fala para português europeu e a sua criação necessita uma base de dados em português europeu etiquetada que permita o treino dos modelos acústicos, optou-se por não utilizar o *Simon*.

O *Audimus* é um programa que transcreve qualquer conteúdo de áudio, fala e vídeo moldado no idioma português europeu [54, 55]. Contactando os responsáveis do *Audimus* foi dito que a sua utilização é paga, por esse motivo foi também descartado.

O *Julius* é um programa muito semelhante ao *Simon* pois permite a criação de diversas aplicações com o reconhecimento da fala de forma livre e a sua utilização requer uma preparação prévia dos modelos acústicos e linguísticos [56]. O *Julius* pode ser utilizado em *Linux* ou *Windows* e possui apenas modelos linguísticos e acústicos em inglês e japonês [56]. À semelhança do *Simon*, o *Julius* foi descartado por não existir modelos linguístico e acústico para português europeu e porque a sua criação necessita uma base de dados em português europeu etiquetada que permita o treino dos modelos acústicos.

O *SpeechMagic*, desenvolvido pela *Nuance*, tem como finalidade ajudar os clínicos nos relatórios dos doentes [57]. A receita médica é transcrita pelo reconhecedor da fala do *SpeechMagic* em vez do processo tradicional, que demora muito mais tempo, tem acesso a todo o histórico da paciente, aumentando assim a produtividade do clínico [57]. Apesar de suportar o português europeu foi também excluída por ser uma ferramenta paga [58].

O *Skype Translator*, pode ser utilizado em qualquer dispositivo compatível com o *Skype*, para fazer uma chamada de voz ou videochamada com pessoas que falem idioma diferente do emissor [59]. Atualmente apenas os idiomas inglês, espanhol, francês, alemão, italiano e mandarim são suportados, logo, esta ferramenta foi também descartada (não suporta português europeu) [59].

O *iSpeech* é uma ferramenta com reconhecimento de fala que permite diversas funcionalidades: desde comandos de voz, *email* e *SMS* por fala, pesquisa na Internet a notações por fala [60]. Esta ferramenta suporta o idioma português europeu mas tem custos associados [60, 61] e por isso optou-se por procurar outra alternativa.

Atualmente o *Microsoft Speech SDK v5.0* (secção 2.3.5) encontra-se na versão 11, designando-se agora por *Microsoft Speech Platform SDK 11* [62]. Para utilizá-lo com a função de reconhecimento de fala é necessário instalar: *Microsoft .NET Framework version 4.0*, *Speech Platform Runtime 11*, *Speech Platform SDK 11* e *Microsoft Speech Platform*

- *Runtime Languages* [62–64]. O *Microsoft Speech Platform SDK 11* é uma ferramenta sem custos, suporta o idioma português europeu mas as aplicações desenvolvidas só são executadas em *Windows* [63]. Atendendo a esse ponto negativo, pois teria-se um sistema executável apenas em computador, optou-se por avaliar outras alternativas.

O *Web Speech API*, encontra-se disponível aos programadores desde 2013 [67], possibilitando a criação de aplicações *web* que necessitem de síntese de fala ou reconhecimento de fala [65]. Para utilizá-lo os programadores devem desenvolver as suas aplicações na linguagem *Javascript* [5]. Relativamente à vertente de reconhecimento da fala, a *Google* disponibiliza várias funções de manipulação: funções de início e de paragem do reconhecimento de fala e função de recepção do reconhecimento [5]. A função de recepção do reconhecimento de fala é onde se encontra a transcrição em texto dos discursos proferidos pelo utilizador [5]. Todas estas funções são necessárias pois as análises dos sinais acústicos proferidos pelo utilizador são realizadas num servidor da *Google*, ou seja, a utilização do *Web Speech API* requer sempre de ligação à Internet e utilização do *browser Google Chrome* para que haja conversão de fala para texto [5, 65, 66, 71]. Esta ferramenta suporta 61 idiomas, em que está incluído o português europeu, mas tem um senão: os reconhecimentos da fala a serem analisados pelo servidor da *Google* não podem ter mais do que 60 segundos de duração [5, 66, 71].

Em suma, dos motores de fala analisados apenas 2 podem ser usados para este trabalho: *Microsoft Speech Platform SDK 11* da *Microsoft* ou *Web Speech API* da *Google*. Como este trabalho é de carácter inclusivo, pretende-se que seja utilizado por uma grande quantidade de pessoas, dessa forma escolheu-se o *Web Speech API*. Com o *Web Speech API* consegue-se maior usabilidade e portabilidade pois as análises dos reconhecimentos da fala requerem apenas qualquer dispositivo móvel com Internet e utilização do *browser Google Chrome* [5, 65, 66, 71]. O ponto fraco será que o sistema só conseguirá fazer reconhecimentos de fala até 60 segundos de duração [72].

3.1.2 Desenvolvimento, Testes e Conclusões do Reconhecedor

O reconhecedor da fala para este trabalho com o *Web Speech API* consistiu em criar uma página *web* com dois botões e uma caixa de texto e, através da utilização das funções de manipulação da *Google* (secção 3.1.1) conseguiu-se fazer a conversão de fala para texto em LP (figura 1.2).

A criação dessa página *web* foi realizada através das linguagens *Hyper Text Markup Language* (**HTML**) e *Javascript*. A linguagem **HTML** serviu para descrever o conteúdo da página [73] e a linguagem *Javascript* serviu para usar as funções de manipulação do *Web Speech API*.

Supondo que o utilizador pretende converter da fala para texto a frase “O meu carro é amarelo”. Para o fazer deve: carregar no botão “Start Reconhecimento” para executar a função de início de reconhecimento de fala (figura 3.2) [71], aceitar a utilização do microfone do seu dispositivo (figura 3.3) de modo a proferir a frase.



Figura 3.2: Testes ao módulo 1: início do reconhecimento da fala.



Figura 3.3: Testes ao módulo 1: utilização do microfone.

Quando o utilizador carregar em “Stop Reconhecimento” (figura 3.4) visualizará a frase reconhecida numa caixa de texto (figura 3.5).



Figura 3.4: Testes ao módulo 1: paragem do reconhecimento de fala.



Figura 3.5: Testes ao módulo 1: frase reconhecida.

Quando o utilizador carrega em “Stop Reconhecimento”, além do sistema apresentar a frase reconhecida, o sistema possibilita ao utilizador um botão (“Enviar”) para enviar o texto (frase reconhecida) para a MT deste trabalho (capítulo 3.2) ou então efetuar um novo reconhecimento (figura 3.5). Durante todo este processo surgem várias mensagens de como o utilizador deve proceder, tais como, como iniciar e terminar o reconhecimento da fala (figuras 3.2 e 3.4).

3.1.2.1 Avaliação da *Web Speech API*

O progresso da tecnologia do reconhecimento automático de fala tem permitido alcançar níveis de desempenho satisfatórios em muitas aplicações. No entanto a qualidade dos modelos acústicos, as condições de aquisição do sinal e a variabilidade imposta pelo locutor condiciona as taxas de reconhecimento [47–49].

Atendendo a que não são conhecidas taxas de reconhecimento da *Web Speech API* para nenhum idioma [5, 65–67], incluído o português europeu, entendeu-se fazer um pequeno teste a fim de obter uma ideia sobre a capacidade de reconhecimento desta ferramenta.

O teste consistiu na definição de um conjunto de frases e a sua gravação através de um microfone, posteriormente cada ficheiro de áudio foi enviado ao servidor *Google* que devolveu o resultado do reconhecimento.

A página *web* de teste da *Web Speech API* atualmente realizada obriga o utilizador a dar permissão de uso do microfone a cada frase que se pretende reconhecer [71]. Numa fase de testes, onde se pretende reconhecer um conjunto razoável de frases este procedimento não é exequível. Não disponibilizando a *Google* [5, 65–67] informação de como criar uma alternativa de introdução de áudio esta tarefa revelou-se um desafio.

Em [74] o autor refere que o servidor *Google* apenas consegue fazer a conversão de ficheiros de áudio do tipo “flac” e disponibiliza o seu algoritmo em *Python* que permite receber um ficheiro “flac”, enviar o seu conteúdo ao servidor da *Google* e depois a resposta é apresentada numa página *web*.

Para determinar a eficiência do reconhecedor (*Web Speech API*) para português europeu seguiram-se os seguintes passos:

1. Definição de frases de teste;
2. Gravação de ficheiros áudio do tipo “flac”;
3. Utilização do algoritmo do autor de [74] de forma a enviar os ficheiros áudio ao servidor da *Google* para apresentar os resultados de reconhecimento;
4. Avaliação dos resultados de reconhecimento.

Definiram-se 10 frases de teste que podem ser visualizadas na tabela 3.1. Estas frases foram escolhidas de forma a permitirem ao sistema a tradução de LP para LGP com diferentes regras gramaticais (tema abordado no capítulo 3.2). Foram também escolhidas de forma a incluírem várias temáticas, nomeadamente: temáticas do quotidiano, temáticas do curso de engenharia eletrotécnica, as cores, o feminino, o vestuário, os transportes, os

nomes próprios, a fruta, a numeração, os animais, os verbos regulares e os verbos SER e ESTAR (secção 3.2.1), tempos verbais, a interrogativa e a exclamativa. Essas frases foram também utilizadas no desenvolvimento da MT descrita no capítulo 3.2.

Tabela 3.1: Frases de Teste usadas na avaliação do *Web Speech API*

Frases de Teste	Descrição
Frase 1	Calcula o valor da resistência.
Frase 2	A mãe comeu duas maçãs.
Frase 3	Ela tem um cavalo e um coelho.
Frase 4	O Luís está no comboio.
Frase 5	A cor do gorro é roxo escuro.
Frase 6	Eu fui com ela ao cinema.
Frase 7	O meu carro é amarelo.
Frase 8	Quanto gastaste?
Frase 9	Ontem fui ao mercado.
Frase 10	Que dia bonito!

As gravações foram feitas por 6 voluntários: 3 femininos e 3 masculinos na faixa etária entre os 24 anos e os 60 anos. Cada locutor gravou as 10 frases proferidas a 3 ritmos diferentes (lento, normal e rápido) resultando em 30 ficheiros por locutor e perfazendo 180 no total.

Para cada gravação utilizou-se o gravador de áudio do *Windows* que grava em formato “wma” (anexo A.1). Fez-se a conversão para “flac” através do software “fre:ac” [74]. Cada gravação foi realizada numa sala com baixo nível ruído e evitando interferências.

Para obter os resultados de reconhecimento dos ficheiros de áudio foi necessário a utilização de outra ferramenta da *Google*, a *Google Speech API*. Com essa ferramenta o servidor da *Google* só permite efectuar até 50 reconhecimentos por dia (anexo A.1). Devido a essa limitação, não foi possível fazer tudo num só dia obter todos os resultados de reconhecimento.

A descrição detalhada do processo de envio de áudio para a *Google* para reconhecimento encontra-se no anexo A.1.

Para a avaliação foi usada a ferramenta *HResults* do software *HTK*.

“O *HTK* é uma ferramenta que permite construir e manipular os modelos escondidos Markov [75]”. Esta ferramenta possui diversas funcionalidades para as análises do discurso, treino *Hidden Markov Models* (*HMM*), teste e análise de resultados [75].

A ferramenta *HResults* necessita: de uma lista de palavras a reconhecer (*wordlist*);

de um ficheiro em formato “MLF”⁶ com informação original do conteúdo das frases ao nível da palavra e de um ficheiro em formato “MLF”⁶ com a informação do resultado de reconhecimento das frases ao nível da palavra.

Na tabela 3.2 ilustra-se parte desses dois últimos ficheiros para a frase 5 que se encontra na tabela 3.1.

Tabela 3.2: *HResults*: frase 5 (tabela 3.1) nos ficheiros em formato “MLF”

Informação Original	Informação do resultado de reconhecimento
#!MLF!# “*/f5.lab”	#!MLF!# “*/f5.rec”
a	a
cor	cor
do	do
gorro	burro
eh	eh
roxo	roxo
escuro	escuro
.	.

Uma vez que a mesma letra pode corresponder a mais do que um fonema o mais correto seria fazer a conversão grafema para fonema de cada palavra. Para essa conversão poder-se-ia ter utilizado o conversor de grafemas para fonemas para português europeu que se encontra em [76]. No entanto, como o vocabulário em questão é reduzido e o problema só se coloca na acentuação do /a/ em **está** e do /e/ em **é**, optou-se por usar um código para esses casos: **está** foi substituído por **estah** e **é** por **eh**. Todos os restantes acentos foram retirados.

A ferramenta *HResults* compara as duas frases e dá indicação do número de erros. Na tabela 3.3 mostra-se o alinhamento obtido em relação às frases apresentadas na tabela anterior. De acordo com esse alinhamento existiu apenas um erro. O sistema da *Google* reconheceu **burro** em vez de **gorro** (tabela 3.3).

Tabela 3.3: *HResults*: frase 5 (tabela 3.1) alinhamento

f5.rec: 1
Aligned transcription: f5.lab vs f5.rec
LAB: a cor do gorro eh roxo escuro
REC: a cor do burro eh roxo escuro

⁶É um formato do *software* **HTK**

Quando a frase reconhecida é diferente da frase original, como acontece na tabela 3.3, existe erros aos níveis da frase e da palavra. Uma avaliação ao nível da frase consiste em verificar se a frase reconhecida é exatamente igual à frase proferida [71]. Utilizando o alinhamento dessas frases, para cada palavra da frase reconhecida tem que se verificar se é a palavra correta ou se é um erro. Esse erro pode ser de inserção de palavras (sistema reconheceu palavras que não foram proferidas), eliminação de palavras (sistema eliminou palavras que foram proferidas) ou substituição de palavras (sistema confundiu as palavras que foram proferida) [71]. Depois dessa análise utiliza-se a equação 3.1 para fazer a avaliação ao nível da palavra, em percentagem, em que: W_{acc} corresponde à eficiência final ao nível da palavra; N é o número de palavras que constitui a frase original; D , S e I são os números de palavras eliminadas, substituídas e inseridas, respetivamente, aquando do alinhamento entre a frase reconhecida com a frase original [71].

$$W_{acc} = \frac{N - D - S - I}{N} \times 100\% \quad (3.1)$$

Tanto a avaliação ao nível da frase como a avaliação ao nível da palavra são realizadas pelo *HResults*.

Na tabela 3.4 apresentam-se os resultados das eficiências do *Web Speech API* totais e por género (M - masculino, F - feminino) para idioma português europeu, em percentagem, para os discursos lento, normal e rápido, aos níveis da frase e da palavra.

Tabela 3.4: Resultados das eficiências do *Web Speech API* para português europeu.

		Eficiência (%)					
		Fala Lenta		Fala Normal		Fala Rápida	
Nível	Frase	M:	40	M:	36,67	M:	23,33
		F:	60	F:	33,33	F:	20
		Total:	50	Total:	35	Total:	21,67
	Palavra	M:	69,39	M:	67,35	M:	54,42
		F:	85,71	F:	72,11	F:	60,54
		Total:	77,55	Total:	69,73	Total:	57,48

Em [71], o autor fez uma avaliação ao *Web Speech API*, em inglês, aos níveis da palavra e da frase.

Na tabela 3.5 apresentam-se os resultados das eficiências do *Web Speech API* totais e por género (M - masculino, F - feminino) para idioma inglês, em percentagem, aos níveis

da frase e da palavra pelo autor de [71].

Tabela 3.5: Resultados da eficiência do *Web Speech API* para inglês [71].

		Eficiência (%)	
		M:	F:
Nível	Frase	M:	21
		F:	20
		Total:	21
	Palavra	M:	74
		F:	72
		Total:	73

Comparando os resultados do idioma em inglês (tabela 3.5) com os resultados do idioma em português europeu (tabela 3.4) verificam-se as seguintes características no *Web Speech API*:

- no português, no seu melhor caso (fala lenta), garante eficiências totais ao nível da frase e da palavra em 50% e 77,55%, respetivamente. No inglês, essas eficiências são 21% e 73%, respetivamente.
- no português, quando se fala rápido, garante uma eficiência total ao nível da frase muito semelhante ao idioma inglês. Já a eficiência total ao nível da palavra, o idioma português apresenta pior desempenho em comparação ao inglês.
- no português, adotando fala lenta, os indivíduos do género masculino garantem o dobro da eficiência ao nível da frase em comparação ao idioma inglês. Já os indivíduos do género feminino garantem o triplo da eficiência em comparação ao idioma inglês. Mas esses valores caem, à medida que se aumenta a velocidade de fala, chegando a serem muito idênticos já a partir da fala normal.

Resumindo, o *Web Speech API* em português europeu oferece melhores resultados aos níveis da frase e da palavra quando se utiliza um discurso lento em comparação aos resultados obtidos em inglês; caso se adote um discurso normal, o sistema garante eficiências semelhantes ao nível da frase e da palavra em comparação ao inglês; caso se adote um discurso rápido o sistema oferece resultados semelhantes ao nível da frase em comparação ao idioma inglês e oferece piores resultados ao nível da palavra em comparação ao idioma inglês.

Assim, como o *Web Speech API* no seu melhor caso (fala lenta), oferece aos níveis da frase e da palavra eficiências totais de 50% e 77,55%, respetivamente, poderá querer

dizer que a sua escolha como reconhecedor de fala para este trabalho fique aquém do desejado. As razões dessa opinião devem-se: tempo máximo limitado de reconhecimento de fala em 60 segundos [72] e apenas 50% da fala proferida não sofrerá qualquer erro. Este desempenho tenderá ainda a piorar em ambientes reais. A solução, como trabalho futuro, passará por fazer avaliações aos níveis da frase e da palavra para o reconhecedor *Microsoft Speech Platform SDK 11*, pois foi descartado (ver secção 3.1.1). Caso nesse estudo se verificar semelhantes eficiências e restrições, será necessário criar um reconhecedor de fala de raiz a fim de garantir eficiências ao nível da frase e da palavra superiores e garantir que o tempo de reconhecimento de fala não seja tão limitado.

3.2 Módulo 2: Máquina de Tradução

As línguas gestuais não seguem a ordem e a estrutura frásicas das línguas orais. À semelhança das línguas orais, as línguas gestuais possuem uma gramática própria sendo estas totalmente independentes [3].

Nesta secção inicia-se com um estudo da **LGP** que permita a conversão de texto **LP** para texto **LGP**. Apresenta-se ainda, a forma de implementação, os seus testes e conclusões da **MT** deste trabalho. Esta secção constitui o módulo 2 deste trabalho e terá como finalidade a conversão do texto **LP** proveniente do reconhecimento da fala para texto em **LGP** (figura 1.2).

3.2.1 Estudo da Língua Gestual Portuguesa

A **LGP** é a língua gestual usada pela comunidade surda portuguesa que foi reconhecida como uma das línguas oficiais de Portugal em 1997 [2]. Esta língua permite a comunicação através dos movimentos das mãos, do corpo e expressões faciais. Possui vocabulário e gramática próprios à semelhança das línguas orais [3] (capítulo 1).

Na realização dos gestos o indivíduo tem sempre uma mão dominante (normalmente é a mão direita) e uma mão não dominante (mão esquerda) [77]. Os gestos são definidos de acordo com 5 parâmetros: configuração, orientação, local de articulação, movimento das mãos e componente não manual (expressões faciais e corporais) [2]. A alteração de algum parâmetro anterior configura um novo gesto ou o gesto em execução perde o sentido [2].

Em regra geral existem gestos para o vocabulário, mas quando não existem recorre-se à datilologia (soletração de palavras) [2]. O alfabeto datilológico é apenas um suplemento das línguas gestuais, muito comum na representação de nomes de pessoas, lugares e vocabulário técnico onde não existe tradução de **LP** para **LGP**.

Quer se use os gestos ou a datilologia, a comunicação está regida por uma estrutura

frásica definida. Essa estrutura é designada de GLOSA [2]. Em LP a estrutura frásica é normalmente SUJEITO-VERBO-OBJETO (SVO), já em GLOSA a estrutura é normalmente SUJEITO-OBJETO-VERBO (SOV) e em algumas situações poderá ser OBJETO-SUJEITO-VERBO (OSV) [78]. Na tabela 3.6 apresentam-se exemplos de traduções de frases em LP para LGP nas duas vertentes: SOV e OSV.

Tabela 3.6: GLOSA: frases em SOV e em OSV [78]

LP	LGP
O gato comeu o peixe. (SVO)	GATO+PEIXE+COMER// (SOV)
O João comprou o livro. (SVO)	LIVRO+J-O-Ã-O+COMPRAR// (OSV)

A tradução de LP para LGP palavra a palavra pode resultar em problemas graves de comunicação, nomeadamente em metáforas, provérbios e frases feitas [78]. Na tabela 3.7 são apresentados exemplos de algumas frases em LP e suas traduções para LGP de forma errada (seguindo a regra SOV) e correta.

Tabela 3.7: GLOSA: traduções erradas [78]

LP	LGP	
	Errado	Correto
Eu estou à espera de bebé.	EU+BEBÉ+ESPERAR//	EU+GRÁVIDA//
Eu dei com a cara na porta.	PORTA+CARA+BATER//	CASA+NINGUÉM//

Outros aspetos importantes para o desenvolvimento da MT são:

- **género:** a marcação só é realizada em seres animados (pessoas e animais). Geralmente a marcação do género só ocorre quando são seres femininos. Essa marcação é feita através da utilização do gesto MULHER precedido com o gesto pretendido, por exemplo, GATA=MULHER+GATO. Existem exceções: gestos entre masculino e feminino que são diferentes, por exemplo, PAI e MÃE e ainda diferenciação por sufixação, tais como, CABRA e BODE, em que BODE=CABRA+BARBICHA [2].
- **número:** a marcação do plural é realizada de diversas maneiras: pela repetição do gesto, por exemplo, COISA e COISAS; pelo redobro (realização do gesto por ambas as mãos), por exemplo, PESSOA e PESSOAS; utilização de números, por exemplo, ‘4 maçãs’=MAÇÃ+QUATRO ou através do determinativo para as quantidades não contáveis, por exemplo, ‘muitos papéis’=PAPEL+MUITO [77].
- **tipos de frase:** normalmente a negativa é realizada de forma simultânea da frase neutra com o movimento natural da cabeça referente ao gesto NÃO [78]. Em GLOSA

a estrutura é através do acréscimo do gesto NÃO no fim da estrutura de frase neutra, mas existem os verbos PODER, QUERER, CONSEGUIR e FALAR em que essa regra não se aplica [2]. Nesses verbos a regra é: NÃO+VERBO [2]. Veja-se a tabela 3.8 em que se mostra a GLOSA para frases negativas.

Tabela 3.8: GLOSA: frases negativas [2, 77, 78]

LP	LGP
O gato não comeu o peixe. (SVO)	GATO+PEIXE+COMER+NÃO// (SOV)
O João não comprou o livro. (SVO)	LIVRO+J-O-Ã-O+COMPRAR+NÃO// (OSV)
O Luís não está no comboio. (SVO)	L-U-Í-S+COMBOIO+NÃO// (SOV)
Eu não sou racista. (SVO)	EU+RACISTA+NÃO// (SOV)
Eu não posso. (SVO)	EU+NÃO+PODER// (SOV)

As duas primeiras linhas da tabela 3.8 são as frases negativas correspondentes às frases que estão na tabela 3.6. As 3ª e 4ª linhas apresentam-se as frases negativas quando em LP usam os verbos SER e ESTAR, que em GLOSA não são representados. Na última linha apresenta-se a frase negativa para o verbo PODER.

Nas frases interrogativas recorrem-se à expressão facial que pode ou não ser combinada com o uso de pronomes interrogativos [77]. Em GLOSA a estrutura é através do acréscimo do gesto do pronome interrogativo no fim da estrutura da frase neutra [2], veja a tabela 3.9 em que se mostra a GLOSA para frases interrogativas.

Tabela 3.9: GLOSA: frases interrogativas [2, 77, 78]

LP	LGP
Ele vai de comboio? (SVO)	ELE+COMBOIO+IR? (SOV)
Onde estás? (SVO)	TU+ONDE?// (SOV)
O teu pai vende cerejas? (SVO)	PAI+TEU+CEREJA+VENDER?// (SOV)

Analisando a tabela 3.9, na primeira linha apresenta-se uma frase interrogativa em GLOSA sem nenhum pronome interrogativo, já na segunda linha apresenta-se uma frase interrogativa, com o verbo ESTAR e com sujeito oculto [79]. Na última linha apresenta-se a interrogativa quando surge um pronome possessivo antes de um substantivo. Em qualquer tipo de frase (declarativa, interrogativa ou exclamativa) quando surge um pronome possessivo antes de um substantivo a regra é: SUBSTANTIVO+PRONOME POSSESSIVO [2].

Já nas frases exclamativas a sua realização é através do uso das expressões faciais e corporais [77].

- **tempo verbal:** na marcação do tempo verbal recorre-se a 3 espaços imaginários que envolvem o gestuante: o espaço atrás do ombro marca o passado, o espaço à frente do gestuante marca o presente e o espaço mais à frente do corpo do gestuante marca o futuro [78].

Em GLOSA a forma mais usual para marcar os tempos verbais é através da adição dos gestos referentes a tempos verbais no início das frases, exceção para o tempo presente que pode ou não utilizar gestos de marcação de tempo [2]. Apresentam-se alguns exemplos de marcação dos tempos verbais na tabela 3.10.

Tabela 3.10: GLOSA: tempos verbais [2, 77, 78]

LP	LGP
Ontem fui ao mercado. (SVO)	ONTEM+EU+MERCADO+IR// (SOV)
Amanhã telefonarei. (SVO)	AMANHÃ+EU+TELEFONAR// (SOV)
Agora não faço. (SVO)	AGORA+EU+FAZER+NÃO// (SOV)

- **Gestos compostos:** alguns gestos são realizados à custa da concatenação de outros gestos, por exemplo, AGRICULTOR e ALMOFADA [77]. O gesto AGRICULTOR=PESSOA+TRABALHAR+AGRICULTURA já o gesto ALMOFADA=DORMIR+FOFO [77].
- **Gestos variáveis:** existem gestos, tais com, PENSAR e ENSINAR que não sofrem variação de acordo com o contexto, mas existem outros que isso já não acontece por exemplo, ABRIR e RÁPIDO [78]. O gesto ABRIR executa-se de maneira diferente em função do contexto das ideias a transmitir [78]. O gesto RÁPIDO executa-se de maneira diferente caso se esteja a referir a seres animados (por exemplo: pessoas) ou não (por exemplo: carro) [78].

Existem ainda muitas mais particularidades da LGP que não foram aqui referidas, mas depois ter feito este breve estudo concluí-se que já se estava em posse de várias regras gramaticais para se inicializar o desenvolvimento da MT.

3.2.2 Escolha do Método para a Máquina de Tradução

A MT tem como finalidade a tradução automática de uma língua para outra através de algoritmos [80] (secção 2.3).

A primeira MT foi apresentada no ano de 1949 [80], com a capacidade de tradução de russo para inglês. Os autores desse projeto previam que as MT poderiam substituir os tradutores humanos dentro de 5 anos. Mas a tradução automática de igual eficiência em

comparação com os tradutores humanos revelou-se um enorme desafio que até hoje ainda perdura [80].

Desde essa altura vários métodos surgiram e foram utilizados a fim de atingir a melhor tradução possível. Esses métodos são: procura da frase (*phrase lookup*) [31, 34], regras (*rule-based*) [32, 34, 44, 46], estatística (*statistical*) [32, 33, 44], exemplos (*example-based*) [33, 44, 46], híbrido (*hibrid*) [46, 80, 81] ou multi-métodos (*multi-engine*) [46].

O método por procura da frase consiste em procurar numa base de dados a correspondência para gestos de cada frase reconhecida [31, 34]. Caso não exista na base de dados o sistema não consegue fazer a tradução [31].

A utilização do método por regras requer previamente uma análise morfológica, sintática e semântica das línguas de origem (língua oral) e de destino (língua gestual). Depois é definido um conjunto de regras que permitem a tradução da estrutura da frase da língua oral na estrutura da língua gestual [32, 34, 44, 46].

O método por estatística baseia-se na teoria de decisão estatística e na aprendizagem estatística. Ao usar este método os sistemas têm que ser treinados usando um conjunto de pares de frases, ou seja, frases representativas da língua fonte (língua oral) e a suas frases correspondentes da língua destino (língua gestual) [32, 33, 44].

O método por exemplos baseia-se na analogia. Este método utiliza frases representativas da língua oral e suas correspondências na língua gestual em que o sistema consegue fazer a tradução desde que as frases que chegam à MT sejam semelhantes às frases representativas da língua oral [33, 44, 46].

O método híbrido consiste na utilização dos métodos por regras e por estatística de forma complementar de modo a garantir uma melhor qualidade de tradução em comparação com a utilização dos métodos de forma independente [46, 80, 81]. Vários estudos revelam que a qualidade de tradução é semelhante quando se usam MT baseadas por regras e por estatística. Nesses estudos verificou-se que as MT por regras possuem como ponto forte a eficiência de tradução de palavras e como ponto fraco a falta de robustez nas falhas da análise das frases; já as MT por estatística os pontos forte e fraco são o oposto [46, 80, 81].

Em [46] o autor refere que qualquer que seja o método usado, regras, estatística, exemplos ou híbrido, a MT não garante os melhores resultados de tradução. Assim nos dias de hoje existem diversos trabalhos em que usam os vários métodos de tradução combinados na MT com o objetivo de fornecer uma melhor qualidade de tradução em comparação com os métodos anteriores usados isoladamente. Esse tipo de máquinas é denominado MT de multi-métodos [46].

O método por exemplos apenas consegue fazer a tradução caso as frases que chegam

à **MT** sejam semelhantes às frases representativas da língua fonte [33, 44, 46]. Assim este método foi excluído pois pretende-se desenvolver uma **MT** em que não se sabe a estrutura das frases que chegam à **MT**.

Foi também excluído o método por procura da frase pois o sistema só conseguia fazer a tradução caso a frase que entra na **MT** seja igual à frase que esteja na base de dados onde existe a frase correspondente para a língua gestual [31].

A **MT** que usa o método por estatística necessita de ser treinada através da utilização de um conjunto de pares de frases, ou seja, frases representativas da língua fonte (língua oral) e as suas frases correspondentes da língua destino (língua gestual), que os autores chamam *corpora* [82]. Uma vez que, no âmbito deste trabalho não se dispõe de uma base de dados que permita o treino de um método estatístico, não foi possível a implementação deste método.

O método híbrido foi descartado por inerência à ausência de um treino estatístico. Este método concilia os resultados dos métodos por regras e por estatística. Estes dois métodos geram erros complementares. A conjugação dos seus resultados faz com que os erros sejam mascarados e se obtenha uma boa qualidade de tradução [46, 80, 81].

Apesar do método por multi-métodos ser apresentado como o método que oferece melhor qualidade de tradução [46], não foi usado nesta fase do trabalho por obrigar à conjugação de diversos métodos.

Assim o método escolhido para implementar a **MT** foi o método por regras. A definição das regras partiu do estudo da **LGP** (secção 3.2.1) previamente realizado.

3.2.3 Desenvolvimento da Máquina de Tradução

A **MT** deste trabalho está dividida em dois módulos (figura 3.6): o módulo A tem como finalidade fazer as análises sintática e morfológica da sequência de palavras proveniente do módulo 1 (reconhecimento da fala) e no módulo B estão implementadas as regras da **LGP** de forma a permitir a tradução para GLOSA que será enviada para o módulo 3 (*interface* gráfica).

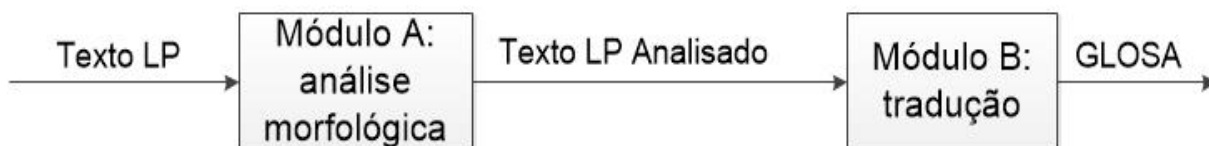


Figura 3.6: Módulo 2 do sistema automático de fala LP para LGP.

Para o módulo A recorreu-se à ferramenta *online* chamada de *Lemmatization, PoS and Parsing* [6] que faz a análise morfossintática ou lematização para vários idiomas, incluindo, o português [83]. Esta ferramenta possui duas versões: análise morfossintática

(*morphosyntactical analysis*) ou modo lematização (*lemmatization*) [6]. Enquanto que a primeira versão tem como finalidade apresentar uma versão simplificada estruturada das palavras que compõem a árvore morfosintática [6], a segunda versão tem como finalidade apresentar uma versão simplificada de lemas⁶ [6]. Ambas as versões poderiam ser integradas neste trabalho através do uso de um *script PHP: Hypertext Preprocessor (PHP)*, Java, Python ou Visual Basic [6]. Qualquer um desses *scripts* permite fazer a ligação com um servidor onde se realiza a análise e depois receber os resultados do servidor de forma aos programadores desenvolverem as suas aplicações [6].

Como o módulo de reconhecimento de fala foi realizado através de uma página *web* com as linguagens **HTML** e *Javascript* (secção 3.1.2), foram escolhidas as versões (análise morfosintática e modo lematização) na linguagem **PHP**. A justificação é que a linguagem **PHP** é facilmente inserida em código **HTML** [84]. Desta forma garante-se a ligação do texto **LP** proveniente do reconhecedor de fala com o início da **MT** deste trabalho (figura 1.2).

Como as duas versões têm finalidades diferentes, realizou-se um pequeno teste de modo a saber qual das duas é mais útil para este trabalho. O teste consistiu no envio da frase “Ontem fui ao mercado” (tabela 3.1) às duas ferramentas e verificar as suas respostas pelo servidor.

Com a primeira versão (modo lematização), a resposta do servidor vem com 2 parâmetros para cada palavra que constitui a frase. Esses parâmetros são: a palavra de entrada (*token*) e a palavra lema (*lemma*). Com o exemplo anterior os 2 parâmetros são iguais para todas as palavras excepto para o verbo. No verbo, o primeiro parâmetro está ‘fui’ e no outro parâmetro está ‘ir’.

Com a segunda versão (análise morfosintática), a resposta do servidor vem com 3 parâmetros para cada palavra que constitui a frase. Esses parâmetros são: a palavra de entrada (*token*), categoria gramatical (*tag*) [85] e a palavra lema (*lemma*). Na tabela 3.11 apresenta-se o resultado simplificado desta versão para a frase “Ontem fui ao mercado”.

Tabela 3.11: Resultado simplificado do servidor da *Lemmatization, PoS and Parsing* na versão análise morfosintática: “Ontem fui ao mercado”

Token	Ontem	fui	ao	mercado	.
Tag	ENTPN-	VI-1ASA-N-	X-N-	NCMSN-	1D
Lemma	ontem	ser	ao	mercado	.

Como em **LGP** a estrutura frásica é **SOV**, opondo-se à **LP** que **SVO** (secção 3.2.1), o desenvolvimento do algoritmo de tradução consistirá em criar a GLOSA com as palavras

⁶Lema: Palavra que figura como entrada num dicionário ou num vocabulário [45]

lemas de forma sequencial e quando surgir ou existir um verbo deslocar a palavra lema do verbo associado para o fim das outras palavras lemas. Assim, foi escolhida a *Lemmatization, PoS and Parsing* na versão análise morfossintática pois era a única que possibilitava o desenvolvimento das regras de acordo com **LGP**.

Foram usadas as frases da tabela 3.1 para desenvolver as regras no módulo B pois incluem várias temáticas e regras da **LGP**, nomeadamente: temáticas do quotidiano, temáticas do curso de engenharia eletrotécnica, as cores, o feminino, o vestuário, os transportes, os nomes próprios, a fruta, a numeração, os animais, os verbos regulares e os verbos SER e ESTAR, tempos verbais, a interrogativa e a exclamativa (secção 3.2.1). As suas correspondências em **LGP**, na vertente **SOV** (secção 3.2.1), encontram-se na tabela 3.12.

Tabela 3.12: GLOSA das Frases de Teste usadas na avaliação do *Web Speech API*

Frases de Teste	GLOSA
Frase 1	TU+NÚMERO+RESISTÊNCIA+CALCULAR
Frase 2	MÃE+MAÇÃ+2+COMER
Frase 3	MULHER+ELE+CAVALO+TAMBÉM+COELHO+TER
Frase 4	L-U-Í-S+COMBOIO
Frase 5	COR+GORRO+ROXO+ESCURO
Frase 6	EU+MULHER+ELE+CINEMA+IR
Frase 7	CARRO+MEU+AMARELO
Frase 8	TU+GASTAR+QUANTO
Frase 9	ONTEM+EU+MERCADO+IR
Frase 10	DIA+BONITO

O resultado do servidor com os 3 parâmetros para todas as palavras que constituem a frase foi colocado numa matriz que foi enviada para o módulo B. A matriz possui sempre 3 colunas e o número de linhas é variável de acordo com o tamanho da frase, isto é, o sistema conta o número de palavras de entrada que constituem a frase de modo a criar as linhas necessárias na matriz. Depois cada uma dessas linhas serão criadas 3 colunas: em cada coluna terá apenas um parâmetro. Depois em cada linha irá ser preenchida com os 3 parâmetros. Como exemplo, na frase “Ontem fui ao mercado.” (tabela 3.11), o sistema criará uma matriz com 5 linhas (o ‘.’ é contado como uma palavra) e em cada linha existirá 3 colunas. Na primeira linha da matriz o sistema colocará na primeira coluna o “Ontem”, na segunda coluna o “ENTPN-” e na terceira coluna o “ontem”; na segunda linha da matriz o sistema colocará na primeira coluna o “fui”, na segunda coluna o “VI-1ASA-N-” e na terceira coluna o “ser”; nas restantes linhas da matriz é realizado o mesmo processo.

Estando a matriz anterior completamente preenchida foi desenvolvido o algoritmo de

tradução.

O algoritmo de tradução consistiu: em cada linha da matriz é verificada a classe gramatical da palavra de entrada em análise. De acordo com essa classe, será colocada a palavra lema respectiva num outro vetor. Nesse vetor terá a GLOSA do texto em LP que entrou nesta MT (figura 3.6). Nas próximas linhas da matriz, o sistema analisa novamente a classe gramatical, verifica se existe um verbo, um número ou um pronome interrogativo no vetor da GLOSA, de modo a adicionar as novas palavras lemas ao vetor da GLOSA de forma correta.

Só quando a GLOSA à saída do módulo B for igual à GLOSA que se encontra na tabela 3.12, para todas as frases, é que se pode afirmar que as regras estão bem implementadas e a “Máquina de Tradução” desta dissertação estará finalizada (figura 1.2).

O diagrama de blocos do algoritmo de tradução desenvolvido encontra-se na figura 3.7.

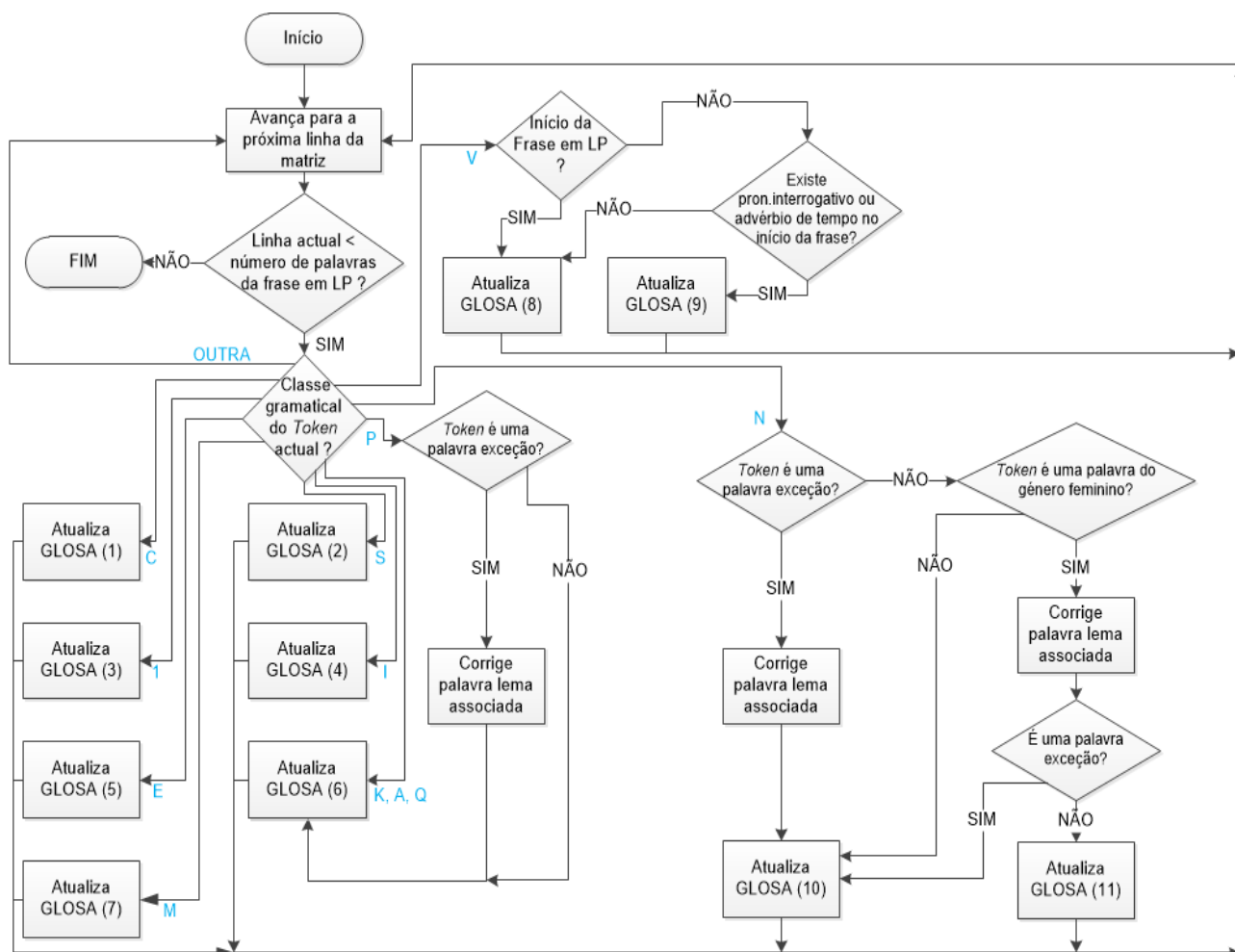


Figura 3.7: Diagrama do funcionamento do algoritmo de tradução desenvolvido.

Para o desenvolvimento das regras da **LGP** foi necessário consultar os códigos [85] para: pronomes pessoais ('P'), verbos ('V'), quantidades ('Q'), nomes ('N'), pontuação ('1'), advérbio ('E'), número ('M'), pronomes possessivos ('S'), conjunções ('C'), adjetivos ('A'), desconhecidos ('K') e interrogativa ('I') (figura 3.7). As principais regras da **LGP** que foram implementadas encontram-se nos blocos “Atualiza GLOSA (1)” a “Atualiza GLOSA (11)”. Na tabela 3.13 apresenta-se a descrição de funcionamento desses blocos.

Tabela 3.13: Descrição dos blocos com as principais regras da **LGP** implementadas

Atualiza GLOSA	Descrição de funcionamento
1	Substituição da palavra lema ‘e’ por ‘também’. Adiciona também ao vetor da GLOSA de igual funcionamento do bloco “Atualiza GLOSA (6)”
2	Adiciona o palavra lema do possessivo no fim do vetor da GLOSA. Faz também o tratamento do ‘dela’ e ‘delas’, colocando no vetor MULHER+DELE ou MULHER+DELES. Ativa variável associada aos possessivos que será útil nos outros blocos
3	Reajuste final da GLOSA com as uma das regras da LGP : VERBO+NÃO+PRONOME INTERROGATIVO, VERBO+NÃO, VERBO+PRONOME INTERROGATIVO ou apenas VERBO e ainda o acréscimo de “\” no fim do vetor com a GLOSA
4	Guarda o pronome interrogativo e ativa uma variável associada à interrogativa que será útil nos outros blocos
5	Igual funcionamento do bloco “Atualiza GLOSA (6)” e por vezes ativa variáveis associadas à negação e ao advérbio de tempo, que serão úteis nos outros blocos
6	Verifica se vetor GLOSA já possui uma palavra lema associada ao verbo. Se existir, então adiciona a palavra lema atual atrás da palavra lema do verbo no vetor com a GLOSA. Caso contrário, adiciona a palavra lema atual no final do vetor com a GLOSA
7	Igual funcionamento do bloco “Atualiza GLOSA (6)” e a ativa variável associada aos números que será útil nos outros blocos
8	Adiciona no vetor com a GLOSA com a regra da LGP : PRONOME PESSOAL + VERBO. Se for SER ou ESTAR só adiciona o pronome pessoal no vetor com a GLOSA
9	Adiciona a palavra lema do verbo no fim do vetor com a GLOSA. Se o verbo for SER ou ESTAR não modifica o vetor. Se a primeira palavra da frase foi um pronome interrogativo ou um advérbio de tempo, acrescenta ao vetor da GLOSA de acordo com a seguinte regra da LGP : PRONOME PESSOAL+VERBO. Se o verbo for SER ou ESTAR então acrescenta apenas ao vetor o PRONOME PESSOAL
10	Adiciona a palavra lema de acordo com as regras da LGP : SUBSTANTIVO(nome)+PRONOME POSSESSIVO+VERBO ou SUBSTANTIVO+NÚMERO+VERBO
11	Adiciona MULHER+palavra lema com as mesmas regras do “Atualiza GLOSA 10”

As frases de teste usadas (tabela 3.1) no desenvolvimento do algoritmo de tradução possuem todas a estrutura sintática **SVO** na forma direta [79], isto é, o sujeito é precedido do predicado. A marcação do predicado é realizado pelo surgimento de um verbo [79]. Basicamente, o algoritmo desenvolvido faz a conversão para **LGP** através do deslocamento da palavra lema do verbo para o final do vetor que contem a GLOSA da frase recebida. Deste modo, está a ser aplicada a vertente **SOV** da **LGP** (secção 3.2.1).

Supondo que se pretende traduzir a frase “Ontem fui ao mercado.” (tabelas 3.1 e 3.11) para GLOSA. Inicialmente essa frase é analisada no módulo A e devolve o resultado (tabela 3.11) num vetor. Depois o algoritmo conta o número de palavras de entrada que esse vetor possui. Dessa forma, o sistema está a preparar o número de linhas necessárias para criar a matriz com os 3 parâmetros (*token*, *tag* e *lemma*) para realizar a tradução. Criada e preenchida a matriz (as palavras que preenchem a matriz podem ser visualizadas na tabela 3.11) inicia-se o algoritmo de tradução (figura 3.7).

Ao iniciar-se o algoritmo de tradução, o vetor que contem a GLOSA é inicializado a nulo. A palavra “Ontem” como é um advérbio de tempo sofrerá as regras do bloco “Atualiza GLOSA (5)” (figura 3.7 e tabela 3.13) e o vetor da GLOSA passará a ter ONTEM. Na próxima palavra, que é “fui”, serão aplicadas as regras do bloco “Atualiza GLOSA (9)”, pois a palavra anterior era um advérbio de tempo (figura 3.7 e tabela 3.13). Deste modo, o vetor GLOSA atual será ONTEM+EU+IR. A próxima palavra (“ao”) é excluída pois não é usada em **LGP** (figura 3.7 e tabela 3.13). Assim, o vetor com a GLOSA permanece inalterado e avança-se para o próxima palavra a ser analisada. “Mercado” é um nome e terá que ser colocado atrás do verbo. Ele sofrerá as regras do bloco “Atualiza GLOSA (10)” sem passar nas correções das palavras lemas (figura 3.7 e tabela 3.13). Desta vez, o vetor com a GLOSA estará com ONTEM+EU+MERCADO+IR. A última palavra é a pontuação (‘.’). Nesse caso o vetor com a GLOSA ficará com ONTEM+EU+MERCADO+IR+\\, ou seja, foram as aplicadas as regras do bloco “Atualiza GLOSA (3)” (figura 3.7 e tabela 3.13).

As palavras de exceção são: mim, ti, contigo, pequeno-almoço e alfa-pendular. As suas palavras lemas corrigidas para **LGP** são: eu, tu, tu, pequeno-almoço e alfa-pendular, respetivamente. Algumas palavras do género feminino são: mãe, tia, cadela e galinha. As respetivas palavras lemas corrigidas são: mãe (exceção), tio, cão e galinha (exceção) (secção 3.2.1). Para corrigir as palavras de exceção e as palavras do género feminino de forma rápida e eficiente foi criada uma base de dados com duas tabelas: uma tabela para as palavras de exceção e uma outra com as palavras do género feminino.

3.2.4 Testes e Conclusões

Na tabela 3.14 apresenta-se a página *web* de teste deste módulo com a tradução de “Olá!” para GLOSA. A frase proveniente do módulo 1 (Texto LP) irá ser analisada pelo servidor da *Lemmatization, PoS and Parsing* na versão análise morfossintática, onde se realiza a análise morfológica. O resultado dessa análise morfológica (Texto LP analisado) será enviado ao módulo B onde se irá fazer a tradução para GLOSA.

Tabela 3.14: Testes ao módulo 2: “olá!”.

```

Texto LP: olá !
=====
Texto LP analisado:

#Olá|NCMSN-|olá#!|1D|!##
=====
Módulo B: MATRIX

Array ( [0] => Array ( [token] => Olá [tag] => NCMSN- [lemma] => olá )
[1] => Array ( [token] => ! [tag] => 1D [lemma] => ! ) )
=====
GLOSA:
OLÁ+//

```

Para cada palavra ou símbolo que esteja presente no texto **LP** encontra-se caracterizado por 3 parâmetros: palavra de entrada (por exemplo: Olá), categoria gramatical [85] associada à palavra de entrada (por exemplo: NCMSN-) e palavra lema (por exemplo: olá). A distinção das caracterizações das palavras é com o recurso do #; a distinção de cada parâmetro, para cada palavra que constitui o texto **LP**, é através de |.

Quer seja “olá” ou “!” os parâmetros palavra de entrada e de palavra lema são muito semelhante ou praticamente iguais, na categoria gramatical isso já não acontece.

Relativamente à categoria gramatical de “Olá” o código significa [85]: ‘N’ - nome (categoria gramatical); ‘C’ - nome comum (subcategoria dos substantivos); ‘M’ - género masculino; S - singular; ‘N’ - palavra normal; ‘-’ - não especificada a frequência de palavras no corpus geral.

Relativamente à categoria gramatical de “!” o código significa [85]: ‘1’ - pontuação; ‘D’ - outro tipo de pontuação.

Para o desenvolvimento das regras da **LGP** no módulo B foi necessário consultar os códigos [85] para: pronomes pessoais (‘P’), verbos (‘V’), quantidades (‘Q’), nomes (‘N’),

pontuação ('1'), advérbio ('E'), número ('M'), pronomes possessivos ('S'), conjunções ('C'), adjetivos ('A'), desconhecidos ('K') e interrogativa ('I').

Antes da criação das regras criou-se uma matriz (Módulo B: MATRIX) que separa cada parâmetro (palavra de entrada, categoria gramatical e palavra lema) para cada palavra do texto LP, de forma a criar o algoritmo de tradução de acordo com a categoria gramatical (secção 3.2.3).

Depois de todo texto LP ter sido analisado e estruturado de acordo com as regras da LGP, foi enviada a GLOSA para o módulo 3 deste trabalho (figura 1.2).

Na tabela 3.15 apresenta-se na página *web* de teste a tradução da frase “O meu carro é amarelo” (tabela 3.1) para GLOSA.

Tabela 3.15: Testes ao módulo 2: “o meu carro é amarelo” (tabela 3.1)

<p>Texto LP: o meu carro é amarelo . =====</p> <p>Texto LP analisado:</p> <pre>##o TDMSN- o#meu SDMS1SN- meu#carro NCMSN- carro##é VI-3PSA-N- ser##amarelo APMSN- amarelo##.1D .##</pre> <p>=====</p> <p>Módulo B: MATRIX</p> <pre>Array ([0] => Array ([token] => o [tag] => TDMSN- [lemma] => o) [1] => Array ([token] => meu [tag] => SDMS1SN- [lemma] => meu) [2] => Array ([token] => carro [tag] => NCMSN- [lemma] => carro) [3] => Array ([token] => é [tag] => VI-3PSA- N- [lemma] => ser) [4] => Array ([token] => amarelo [tag] => APMSN- [lemma] => amarelo) [5] => Array ([token] => . [tag] => 1D [lemma] => .))</pre> <p>=====</p> <p>GLOSA: CARRO+MEU+AMARELO+//</p>

Com a figura anterior é possível verificar que: nos verbos [85] os parâmetros de palavra de entrada ('é') e palavra lema ('ser') são diferentes; as regras no módulo B estão bem implementadas, pois: como a frase original (Texto LP) possui um pronome possessivo (meu) antes de um substantivo (carro) e possui o verbo SER, a GLOSA será: SUBSTANTIVO+PRONOME POSSESSIVO+OBJECTO (secção 3.2.1).

Atualmente consegue-se traduzir para GLOSA as frases que sejam semelhantes às frases de teste (tabela 3.1) mas, quando não existe gesto para alguma palavra em LGP o recurso é a datilologia (secção 3.2.1). Optou-se por não implementar a datilologia neste módulo por não se saber à partida que vídeos dos gestos estariam disponíveis para representar as GLOSAS, ou seja, se não existir vídeo para representar CAVALO e se não existir datilologia no módulo 3, a frase 3 da tabela 3.12 não será visualizada. Como a implementação do algoritmo da datilologia nos 2 módulos (MT e Interface Gráfica) só traria complexidade em ambos os algoritmos para obter um mesmo resultado final, optou-se implementar apenas a datilologia no módulo 3.

Assim, a GLOSA da frase 4 da tabela 3.12 em vez de ser L-U-Í-S+COMBOIO está LUÍS+COMBOIO.

Atualmente este módulo possui diversas limitações:

- **frases negativas:** atualmente o sistema só realiza a negativa para verbos regulares. Para verbos irregulares, tal como, poder, o sistema não efetua a GLOSA corretamente. Na frase “Eu não posso” (tabela 3.8) a tradução seria EU+NÃO+PODER mas infelizmente o sistema apresenta EU+PODER+NÃO. A solução passará por adicionar ao algoritmo o tratamento das frases com os verbos irregulares como trabalho futuro.
- **género feminino:** por vezes a ferramenta *Lemmatization, PoS and Parsing* [83] falha nas palavras “lemma” por isso resolveu-se corrigir algumas delas. Para o tratamento criou-se uma tabela com 30 palavras com género feminino, tais como: tia, avó, urso e galinha. Como existe uma enorme quantidade de palavras na LP com género feminino será necessário aumentar esta tabela em trabalho futuro a fim de eliminar este problema.
- **distinção de gestos variáveis:** existem gestos com o mesmo nome mas que se realizam de forma diferente (secção 3.2.1), tal como, partir. O gesto PARTIR pode ter dois significados: iniciar trajeto ou quebrar coisas [77]. Este sistema ainda não possui essa capacidade de distinção e que terá que ser implementada em trabalho futuro.
- **GLOSA na vertente OSV:** na frase “O João comprou o livro” (tabela 3.6) a GLOSA deveria ser LIVRO+J-O-Ã-O+COMPRAR infelizmente o módulo de tradução (figura 3.6) apenas realiza para a vertente SOV ficando J-O-Ã-O+LIVRO+COMPRAR. Assim, será necessário um acompanhamento dos linguistas da LGP para que o sistema tenha a capacidade de distinção da GLOSA em SOV ou GLOSA em OSV em função da frase LP que entra nesta MT.

- **traduções erradas:** na frase “Eu dei com a cara na porta” (tabela 3.7) a GLOSA deveria ser CASA+NINGUÉM mas este módulo traduz palavra a palavra e sua GLOSA deu EU+CARA+PORTA+DAR. Este problema não foi solucionado pois só com o conhecimento do contexto da frase pelo o sistema é possível fazer a conversão para GLOSA corretamente. Este problema está associado às limitações do método por regras aplicada a esta MT, que tem como ponto fraco as análises às frases (secção 3.2.2). A solução é aplicar um método que envolva o treino estatístico na MT, de forma que este problema não ocorra tão frequentemente (secção 3.2.2).
- **erros devidos à ferramenta *Lemmazation, PoS and Parsing* na versão análise morfossintática** (secção 3.2.3): por vezes a análise das frases por esta ferramenta falha e a conversão para GLOSA fica comprometida. Por exemplo: “Ontem fui ao mercado” (tabela 3.1) a GLOSA deveria ONTEM+EU+MERCADO+IR (tabela 3.12), mas devido à falha da análise da frase de forma completa no verbo no parâmetro número [85] a GLOSA final fica ONTEM+MERCADO+IR. Como não foi encontrada nenhuma ferramenta que fizesse a análise morfossintática, com semelhantes características à que foi usada para português, daí o sua utilização (secção 3.2.3).
- **caraterísticas das frases:** atualmente o sistema só está preparado para frases que sejam idênticas às frases da tabela 3.1. Será necessário a utilização de frases de testes com maior nível de complexidade de forma a desenvolver ainda mais as regras já implementadas na MT.

São demasiadas limitações o que evidência a grande dificuldade em desenvolver uma MT como se pode verificar também nas secções 2.3 e 3.2.2.

3.3 Módulo 3: *Interface* Gráfica

Esta secção inicia-se pela escolha do tipo de *interface* gráfica a ser usada, com base nessa escolha, apresenta-se todo o desenvolvimento, testes e conclusões da *interface* gráfica que constitui este sistema. Esta secção constitui o módulo 3 deste trabalho e terá como finalidade a geração de um vídeo final de acordo com texto LGP recebido da MT deste trabalho (figura 1.2).

3.3.1 Escolha da *Interface* Gráfica

Para escolher qual o método a ser usado para a *interface* gráfica fez-se uma análise da tabela 2.1 onde se evidencia as vantagens e desvantagens dos métodos possíveis que podem

ser utilizados.

Foi escolhida a *interface* gráfica baseada por vídeos pois oferecia maior aceitabilidade pela comunidade surda e os tempos de gravações dos vídeos serem rápidos (tabela 2.1). Em contrapartida, a *interface* gráfica ficará dependente do intérprete e as transições entre os diferentes vídeos concatenados num vídeo final serão incomodativas para os surdos (tabela 2.1).

3.3.2 Desenvolvimento da *Interface* Gráfica

Para realizar a *interface* gráfica fez-se:

1. descarregamento de vídeos temáticos de **LGP** do YouTube [86], cada vídeo continha uma temática da **LGP**, por exemplo: verbos, cores e meios de transporte;
2. recortes nos vídeos temáticos para múltiplos vídeos com um único gesto, por exemplo, comer, amarelo e carro; cada um desses vídeos foram guardados no servidor com tipo “webm”;
3. criação de uma página *web* de teste em **PHP** que fizesse a conversão de GLOSA (secção 3.2) para vídeo e sua visualização no *Google Chrome* (secção 3.1.1).

Para os recortes dos vídeos temáticos e a guarda dos vídeos recortados em tipo “webm” utilizou-se o *software* AVS Video Editor 7.1 [87]. Ao todo geraram-se cerca de 300 vídeos em formato “webm”. Todos os vídeos guardados possuem as seguintes especificações de vídeo (tabela 3.16):

Tabela 3.16: Especificações dos vídeos LGP guardados.

Caraterística	Descrição
Tamanho	640x480 pixels
Bitrate	1024 kbps
Taxa de frames	25 frames/segundo
Compressão de vídeo	VP8

Escolheu-se o formato “webm” para os vídeos pois é o único formato presente no *software* AVS Video Editor 7.1 que fosse possível de ser visualizado numa página *web* no *Google Chrome* [88].

Recebendo a GLOSA do módulo 2 (secção 3.2) criou-se um *script* em **PHP** e com a utilização do *software* FFMPEG [7, 8, 89] foi possível concatenar os vídeos recortados de forma apresentar um vídeo final que pudesse ser visualizado no *Google Chrome*. Caso não existam os vídeos da GLOSA este *script* faz a datilologia.

O algoritmo desenvolvido que fizesse a conversão de GLOSA para um vídeo final consistiu: para cada palavra da GLOSA é verificada se existe o seu vídeo correspondente numa **BD** (figura 1.3). Essa **BD** contém uma tabela de todos os vídeos guardados que existem no servidor da aplicação. Se a palavra existir na **BD**, então o algoritmo cria um ficheiro de texto com a localização do vídeo correspondente para essa palavra; se não existir na **BD**, o algoritmo parte a palavra letra a letra. Cada letra dessa palavra é enviada à **BD** a fim de encontrar o seu vídeo respetivo e, coloca a localização do vídeo correspondente para essa letra no ficheiro de texto. Esse ficheiro de texto fica guardado no servidor. Deste modo, está a ser usada a datilologia (secção 3.2.1). Se não existir vídeo para essa letra, o sistema não coloca nenhuma localização do vídeo no ficheiro de texto e continua o algoritmo, pois não existem vídeos das letras acentuadas, tais como, ‘é’, ‘ç’ e ‘â’. Estando criado o ficheiro de texto com todas as localizações de todos os vídeos respetivos à GLOSA utilizou-se o *software* FFMPEG [7, 8, 89]. Esse *software* lê linha a linha do ficheiro de texto e faz a concatenação dos vídeos e gera um vídeo final que represente a GLOSA. Esse vídeo final é depois mostrado numa página *web* desta aplicação para o surdo a ver.

3.3.3 Testes e Conclusões

Na tabela 3.17 apresenta-se um exemplo do conteúdo no ficheiro de texto que permitirá a concatenação dos vídeos pelo FFMPEG num vídeo final. Observando novamente a figura 1.2, o sistema irá concatenar de acordo com a GLOSA para a frase “Ela viaja no domingo”.

Tabela 3.17: Concatenar vídeos com o FFMPEG [8]: frase “Ela viaja no domingo”.

file ‘mulher.webm’
file ‘ele.webm’
file ‘domingo.webm’
file ‘viajar.webm’

Se não existir vídeos para as palavras da GLOSA, o sistema fará a datilologia (secção 3.3.2). Nesse caso aparecerá algo do género como na tabela 3.18. Na tabela 3.18 apresenta-se o conteúdo do mesmo ficheiro de texto mas agora preparado para gerar um vídeo para a frase “A cor do gorro é roxo escuro” (tabela 3.1).

Tabela 3.18: Concatenar vídeos com o FFMPEG [8]: frase “A cor do gorro é roxo escuro”.

```
file 'c.webm'  
file 'o.webm'  
file 'r.webm'  
file 'gorro.webm'  
file 'roxo.webm'  
file 'escuro.webm'
```

Na tabela anterior verifica-se que o sistema teve que recorrer à datilologia para a palavra “cor”. O sistema recorreu à datilologia por não existir o vídeo respetivo para essa palavra (secção 3.3.2).

Nas figuras 3.8(a) a 3.8(c) apresentam-se os gestos CARRO, MEU e AMARELO concatenados e apresentados num vídeo final numa página *web* de teste no *Google Chrome*, constituindo desta forma a passagem de GLOSA da frase “O meu carro é amarelo” (tabela 3.1) para vídeo em LGP.

Atualmente este módulo possui algumas limitações:

- **gestos multi-palavras:** o sistema só consegue apresentar os gestos que contêm uma única palavra, por exemplo, CARRO. Gestos, tal como, FIGUEIRA DA FOZ [77] em que existe gesto, este sistema apresentará consecutivamente os gestos FIGUEIRA e FOZ, o que dará outro sentido em LGP. A solução passará modificar este módulo ou o módulo anterior (secção 3.2) em trabalho futuro.
- **poucos vídeos:** apesar de se dispor de cerca de 300 vídeos guardados, gestos tal como, BONITO (que tem gesto em LGP) não se encontra guardado, e consequentemente, o sistema fará datilologia. Como trabalho futuro é necessário aumentar o número de vídeos guardados.
- **frases exclamativas e negativas:** o sistema realiza a GLOSA para frases exclamativas e negativas (secção 3.2) mas sua apresentação em vídeo em LGP possui um grande problema. O problema é que nas frases exclamativas, tal como, “Que dia bonito” (tabela 3.1) a expressão facial de admiração não é apresentada, apenas se mostra no vídeo a concatenação dos gestos com uma expressão facial neutra; em frases negativas, tal como, “O João não comprou o livro” (tabela 3.8) a negação deverá que ser apresentada durante a duração do vídeo o que o sistema não consegue realizar. Como a construção dos vídeos é por concatenação é impossível representar as frases negativas e interrogativas em vídeo, pois elas precisam de expressões faciais ao longo do vídeo final. Uma possível solução passará por gravar em vídeo as frases

interrogativas e negativas de forma completa. Desse modo, ter-se-á que ajustar os módulos da **MT** e/ou *interface* gráfica a fim do sistema conseguir utilizar os vídeos dessas frases.

- **datilologia:** as palavras que não têm a sua correspondência em vídeo com o gesto irão para um algoritmo de datilologia, (secção 3.3.2). Por exemplo, a palavra “LUÍS” é uma dessas palavras. Tomando como exemplo, o sistema fará a concatenação corretamente dos vídeos “L” e “U” mas, a partir de “Í” o FFMPEG já não consegue concatenar. A razão deve-se por não existir vídeos de gestos dos caracteres acentuados e com cedilha, tais como: ‘í’, ‘ç’ e ‘ã’. Este problema limita a utilização de frases no âmbito da engenharia eletrotécnica, pois existe muito vocabulário acentuado, tais como: osciloscópio, resistência, tensão elétrica, máquina elétrica e lâmpada, em que não se encontram vídeos desses vocábulos na aplicação. Assim o sistema irá recorrer à datilologia e ocorrerá o problema anterior descrito. Assim como trabalho futuro será necessário fazer a gravação dos vídeos das letras e palavras acentuadas.
- **qualidade visual mediana e irregular:** observando a figura 3.8 constata-se que a qualidade visual não é constante. Se na figura 3.8(a) apresenta-se mais escura, na figura 3.8(b) apresenta-se mais branca. Além desse aspeto verifica-se um sombreamento no lado esquerdo em todas as figuras, que afeta a visualidade das expressões faciais. Verifica-se também que o vestuário do gestuante não é sempre o mesmo (figuras 3.8(a) e 3.8(c)). Na figura 3.8(c) observa-se pouca qualidade nos dedos o que dificulta a leitura do gesto por parte do surdo. Assim como trabalho futuro passará por fazer a gravação dos vídeos num estúdio onde se possa garantir boas condições de luminosidade a um único gestuante com a mesma roupa. Depois guardar os vídeos com uma boa qualidade visual no formato “webm” (secção 3.3.2).

Teste Módulo3

GLOSA: CARRO MEU AMARELO //



(a) Gesto: CARRO

Teste Módulo3

GLOSA: CARRO MEU AMARELO //



(b) Gesto: MEU

Teste Módulo3

GLOSA: CARRO MEU AMARELO //



(c) Gesto: AMARELO

Figura 3.8: Testes ao módulo 3.

Capítulo 4

Funcionamento Global do Sistema

Neste capítulo apresenta-se a integração dos 3 módulos num *website*. Depois da integração apresentou-se o sistema final a alguns surdos e intérpretes da **ASAE**, a alguns especialistas em áudio e/ou vídeo da **ESTG** e a algumas pessoas sem qualquer especialização nas áreas de vídeo, áudio ou **LGP**, para avaliarem o sistema e obter conclusões sobre o trabalho realizado.

4.1 Integração dos 3 Módulos

O *website* onde é feita a integração dos 3 módulos foi realizado através da ferramenta *bootstrap* [90]. A *bootstrap* é uma *framework* que permite desenvolver *websites* e aplicações, com um simples código base, que se autoajusta (*responsive*) qualquer que seja o tamanho do ecrã dos dispositivos, nomeadamente, computadores, *smartphones* e *tablets* [90].

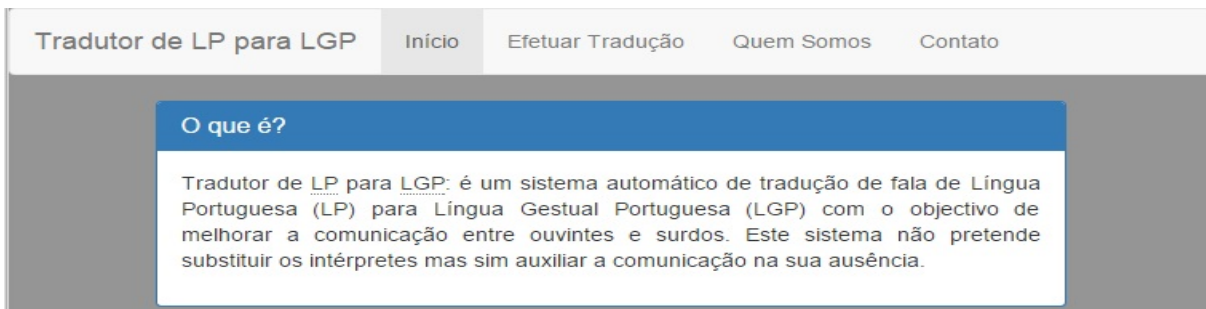
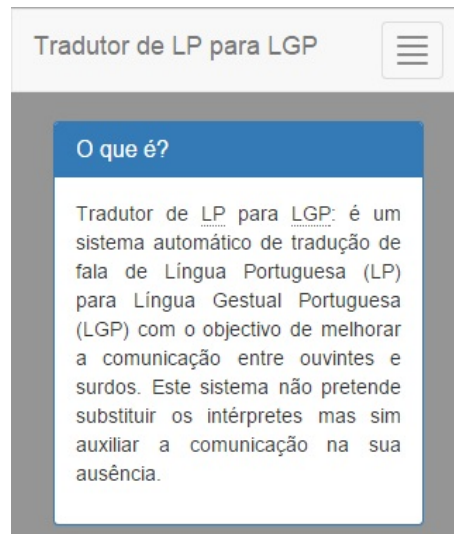
Uma ferramenta *framework* pode ser vista como uma aplicação quase completa em que o programador apenas tem que acrescentar os pedaços que são específicos para a sua aplicação [91].

Fazendo o *download* do *bootstrap* e utilizando a vária documentação existente em [90] realizou-se o *website* para este projeto.

Na figura 4.1 é possível ver a página *web* inicial para computadores e para *tablets*, verificando o ajuste da informação de acordo com o tamanho do ecrã do dispositivo usado.

Como se pode ver na figura 4.1(a) este *website* possui apenas um menu para 4 páginas distintas: “Início”, “Efetuar Tradução”, “Quem Somos” e “Contato”.

Quando o utilizador acede a este *website* será a página “Início” por defeito que será mostrada. Esta página possui diversas informações, nomeadamente, o que é o tradutor de **LP** para **LGP**, como usar esse tradutor e algumas curiosidades da **LGP** e dicas de como

(a) Página Início em *Desktop* (computador)(b) Página Início em *Tablets*Figura 4.1: *Website* com integração dos 3 módulos.

comunicar com surdos sem o conhecimento da **LGP**.

Na página “Quem Somos” apresentam-se os membros e parceiros constituintes que permitiram o desenvolvimento deste trabalho.

Na página “Contato” apresenta-se o contato que o utilizador pode usar para qualquer esclarecimento ou sugestão sobre este trabalho.

A página onde se realiza a tradução é na “Efetuar Tradução”. É nesta página onde os 3 módulos estão integrados (secções 3.1, 3.2 e 3.3).

Quando o utilizador se encontra na “Efetuar Tradução” a fim de realizar a tradução para **LGP** deve seguir os seguintes passos:

1. Carregar no botão “START Reconhecimento”;
2. Aceitar o uso do microfone;
3. Falar;
4. Carregar no botão “STOP Reconhecimento”;

5. Carregar no botão “Traduzir”.

Quando o utilizador faz o passo 4 aparecerá uma caixa de texto com o reconhecimento que poderá ser modificado ou corrigido caso o utilizador assim o pretenda, ou então realizar um novo reconhecimento (voltando para o passo 1) ou continuar para o passo 5 (figura 4.2).



Figura 4.2: *Website* com integração dos 3 módulos: frase LP reconhecida.

Quando o utilizador faz o passo 5 o sistema realiza os módulos 2 e 3 e apresenta em vídeo a tradução para LGP ao utilizador. Neste momento o utilizador pode rever o vídeo ou realizar um novo reconhecimento (figura 4.3). Ainda na figura 4.3 verificam-se as seguintes características:

- ajuste do *website* às dimensões do ecrã de um *smartphone* (secção 4.1);
- o utilizador vê a frase LP reconhecida e sua respetiva GLOSA (secção 3.2);
- o sistema verifica se existe vídeo cada palavra da GLOSA, o que corresponde à sequência de gestos a serem concatenados num único vídeo. Caso não exista vídeo para a palavra da GLOSA em análise pelo sistema, o sistema fará a datilologia para essa palavra (secção 3.3). Neste caso existem todos os vídeos para a GLOSA;
- o vídeo possui dimensões de tela reduzidas devido ao facto do uso do *bootstrap* (secção 4.1), mas caso o utilizador queira utilizar todo o ecrã do seu dispositivo o *Google Chrome* permite fazê-lo [88].

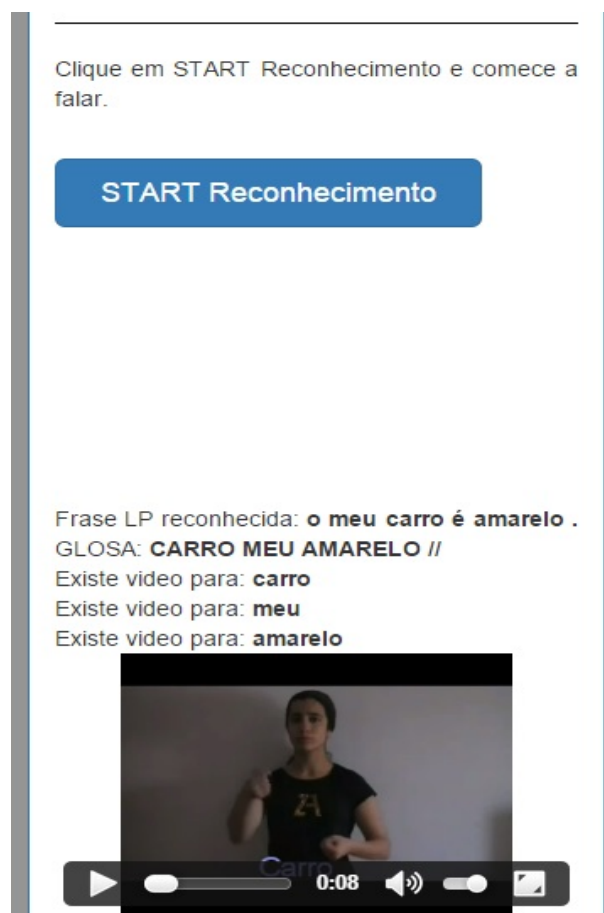


Figura 4.3: *Website* com integração dos 3 módulos: vídeo correspondente à frase LP reconhecida.

Deste modo encontra-se concluída a integração dos 3 módulos num *website*.

Atualmente o *website* criado não se encontra disponível na Internet. De forma a poder utilizá-lo é necessário instalá-lo previamente num computador com ligação à Internet. Esse tutorial de instalação e utilização do sistema desenvolvido encontra-se disponível no anexo [A.2](#).

4.2 Avaliação por alguns surdos e intérpretes da ASAE, alguns Especialistas e Público-Geral

Concluída a integração dos 3 módulos num *website* foi-se ao encontro de alguns surdos e intérpretes da [ASAE](#), também de alguns especialistas nas áreas de áudio e/ou vídeo na [ESTG](#) e de algumas pessoas ouvintes sem nenhuma especialização de modo que analisassem e dessem a sua opinião, através de um questionário, sobre o funcionamento global do sistema.

4.2.1 Objetivos

A avaliação do sistema com alguns surdos e intérpretes da ASAE teve como principais objetivos:

- confirmação das informações inseridas na página “Início” do *website* (figura 4.1(a));
- avaliação do reconhecedor de fala;
- confirmação/avaliação da tradução das frases em LP para GLOSA (sequência de palavras em LGP) de forma correta;
- avaliação da *interface* gráfica: qualidade visual e qualidade de apresentação dos gestos, isto é, verificação se os gestos são os corretos;
- utilidade do sistema desenvolvido.

A avaliação do sistema com alguns especialistas em áudio e/ou vídeo da ESTG, bem como algumas pessoas ouvintes comuns, pessoas que não possuem qualquer especialização em áudio, vídeo e/ou LGP, teve como principais objetivos:

- avaliação do reconhecedor de fala;
- avaliação da *interface* gráfica: qualidade visual;
- utilidade do sistema desenvolvido.

4.2.2 Método de Apresentação do Sistema

Na página “Efetuar Tradução” encontram-se frases em LP e suas respectivas GLOSA (figura 4.4).

Exemplos GLOSA

LP	LGP
Determina o tempo de descarga do condensador.	TU+TEMPO+DESCARGA+CONDENSADOR+DETERMINAR
A mãe comeu duas bananas.	MÃE+BANANA+2+COMER
Ela viajou no domingo.	MULHER+ELE+DOMINGO+VIAJAR
A Daniela está no comboio.	D-A-N-I-E-L-A+COMBOIO
A cor do gorro é roxo escuro.	COR+GORRO+ROXO+ESCURO
Eu fui com ela ao cinema.	EU+MULHER+ELE+CINEMA+IR
O meu carro é amarelo.	CARRO+MEU+AMARELO
Ontem fui ao mercado.	ONTEM+EU+MERCADO+IR
Quanto gastaste?	TU+GASTAR+QUANTO?
Que dia bonito!	DIA+BONITO

Figura 4.4: *Website* com integração dos 3 módulos: frases LP e respectivas GLOSA.

As frases da figura anterior foram proferidas e os surdos e intérpretes avaliaram a GLOSA e os vídeos finais respetivos. Foram proferidas frases que não estivessem na figura anterior a fim de verificarem que o sistema não está só preparado para aquelas frases.

Como os surdos presentes possuíam ausência de fala, outros faziam-no com muita dificuldade, pediu-se aos intérpretes que proferissem algumas frases e que efetuassem a tradução de forma a poderem dar uma opinião sobre o sistema com maior rigor.

No fim alguns surdos e intérpretes preencheram um questionário de avaliação (secção 4.2.3) sobre este sistema.

Foi-se ao encontro de alguns especialistas e pessoas comuns de modo a darem a sua opinião sobre o sistema. Pediu-se a essas pessoas que proferissem algumas frases da figura 4.4 e outras frases que não existissem nessa figura de forma a poderem opinar sobre o sistema, mais concretamente, o reconhecimento de fala e os vídeos finais das frases proferidas. No fim essas pessoas preencheram um questionário de avaliação (secção 4.2.3) sobre este sistema, à semelhança dos surdos e intérpretes como foi descrito anteriormente.

Na posse de todos os resultados dos questionários de avaliação obtiveram-se as conclusões finais do sistema desenvolvido presentes na secção 4.2.4.

4.2.3 Questionário de Avaliação ao Sistema

O questionário de avaliação que os inquiridos preencheram acerca do funcionamento do sistema desenvolvido nesta dissertação encontra-se no anexo B.

As perguntas são de duas índoles: sobre a pessoa que está a preencher o questionário e sobre a avaliação do sistema. Relativamente às perguntas sobre a pessoa são: a idade, a profissão, o conhecimento das LGP e da LP e que área de conhecimento possui de acordo com profissão. Em relação às perguntas sobre a avaliação do sistema estas são niveladas, excluindo a última pergunta que é de resposta aberta.

As perguntas niveladas e as suas escalas de avaliação acerca do sistema desenvolvido encontram-se na tabela 4.1.

Tabela 4.1: Questionário: perguntas sobre o sistema.

PERGUNTA	ESCALA
A aplicação é útil?	1 - Pouco Útil 5 - Muito Útil
Como classifica a experiência/facilidade de usar a aplicação?	1 - Muito Má 5 - Muito Boa
A aplicação tem visual atrativo?	1 - Nada Atrativo 5 - Muito Atrativo
Avaliação do Reconhedor de Fala. Palavras reconhecidas?	1 - Nenhuma Palavra 10 - Todas Palavras
Avaliação da conversão para GLOSA. Palavras convertidas?	1 - Nenhuma Palavra 10 - Todas Palavras
Avaliação dos vídeos em LGP. Palavras GLOSA visualizadas?	1 - Nenhuma 10 - Todas
Avaliação dos vídeos em LGP: tamanho da janela de vídeo.	Pequeno / Grande / Bom
Avaliação dos vídeos em LGP: qualidade visual.	1 - Muito Má 10 - Muito Boa

4.2.4 Resultados e Conclusões

Ao todo foram inquiridas 12 pessoas: 6 femininas e 6 masculinas na faixa etária entre os 22 anos e os 60 anos, de diversas profissões: desde estudantes, a professores na área de áudio e/ou vídeo a operários fabris que responderam ao questionário.

Neste universo de 12 pessoas apenas duas conheciam de igual modo as LP e a LGP, que são os intérpretes; as restantes pessoas só conheciam apenas a LP, que são as pessoas

ouvintes comuns, ou então só conheciam a **LGP**, que são os surdos.

Na questão da utilidade do sistema desenvolvido (tabela 4.1) cerca de 60 % dos inquiridos acha que o sistema é útil (figura 4.5). Não obteve maior nível de utilidade devido às diversas limitações que o sistema possui (secções 3.1, 3.2 e 3.3) e de não permitir uma comunicação bidirecional.

A aplicação é útil?

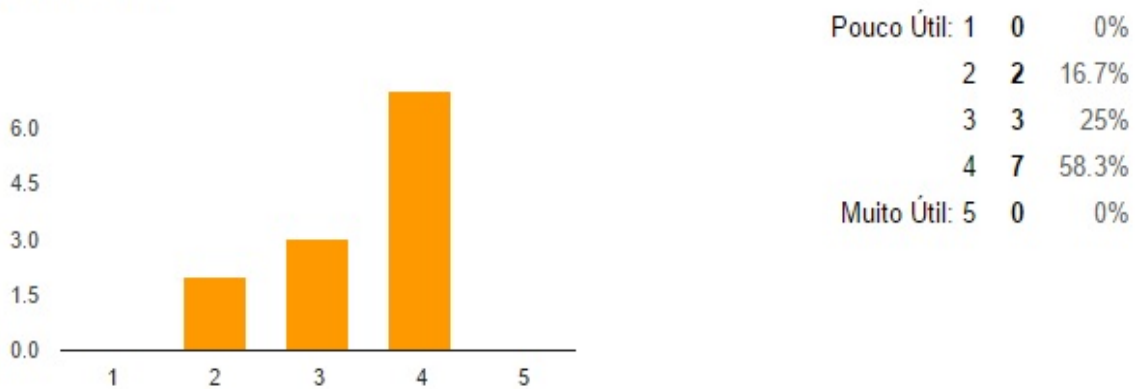


Figura 4.5: Resultados questionário: questão sobre utilidade do sistema.

Na questão da experiência/facilidade de usar o sistema (tabela 4.1) os resultados foram bons, pois obtiveram-se 50% com nota 4 e 41,7% com nota 3 (figura 4.6). Assim será necessário melhorar um pouco mais a facilidade de uso do sistema em trabalho futuro.

Como classifica a experiência/facilidade de usar a aplicação?

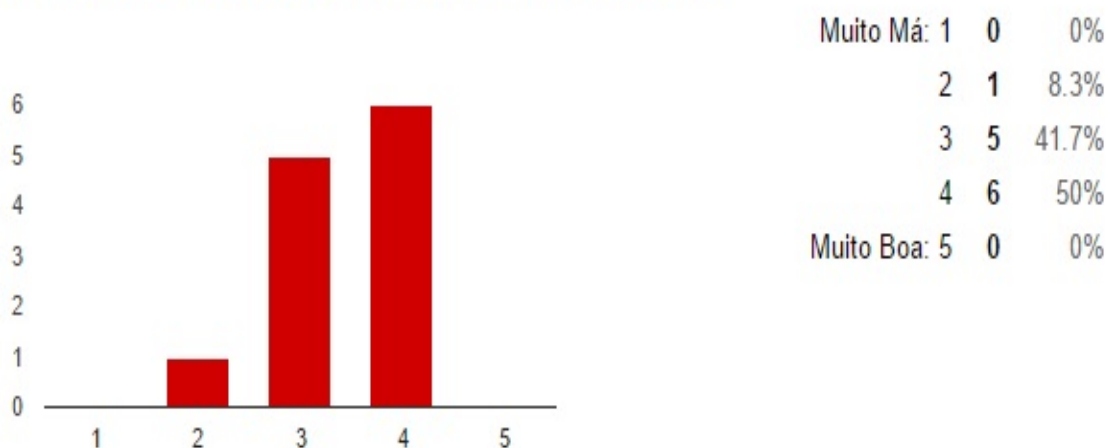


Figura 4.6: Resultados questionário: questão sobre experiência/facilidade de usar o sistema.

Na questão sobre o visual atrativo (tabela 4.1) obtiveram-se resultados tripartidos: 25% nas notas 2, 3 e 4 (figura 4.7) significando que o sistema precisa melhoramentos a nível visual em trabalho futuro.

A aplicação tem visual atrativo?

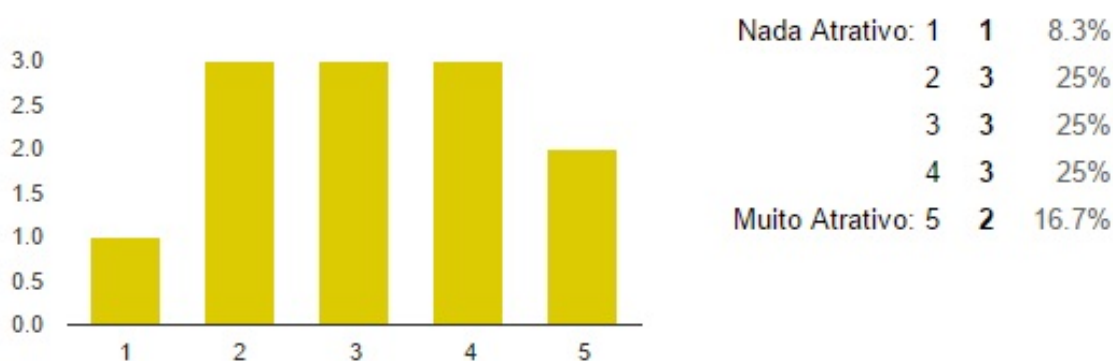


Figura 4.7: Resultados questionário: questão sobre o visual atrativo.

Na questão de avaliação do reconhecedor de fala (tabela 4.1) os resultados foram baixos, pois obtiveram-se 20% nas notas 3 e 7, 26,7% na nota 5 e 13,3% nas notas 6 e 9 (figura 4.8). O reconhecedor de fala não é possível melhorar pois é reconhecedor desenvolvido pela *Google* (secção 3.1), a solução passará por fazer testes a outros reconhedores e verificar qual deles é melhor que o atual e utilizá-lo. Se não se conseguirem melhores resultados será necessário criar um reconhecedor de fala de forma a ter uma melhor aceitação pelos utilizadores.

Avaliação do Reconhecedor de Fala.

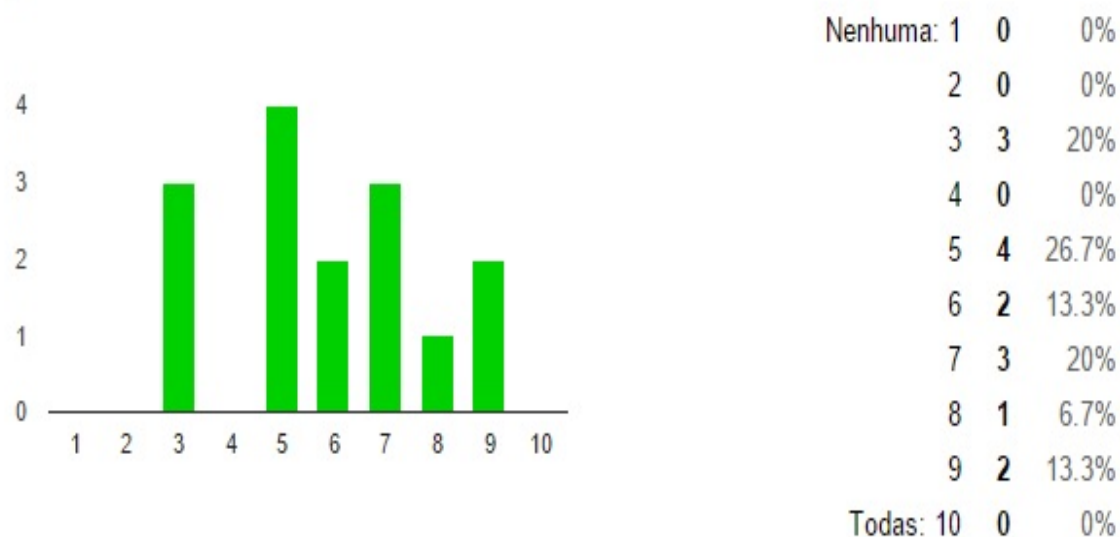


Figura 4.8: Resultados questionário: questão de avaliação do reconhecedor de fala.

Na questão de avaliação para GLOSA (tabela 4.1) os resultados foram mediantemente bons, pois obtiveram-se 33,3% nas notas 5 e 7 e 16,7% na nota 9 (figura 4.9) significando que o sistema precisa de uns melhoramentos nas regras da *LGP*, visto a que a ferramenta

Lemmatization, PoS and Parsing na versão análise morfossintática já é bastante robusta (secção 3.2).

Avaliação da conversão para GLOSA.

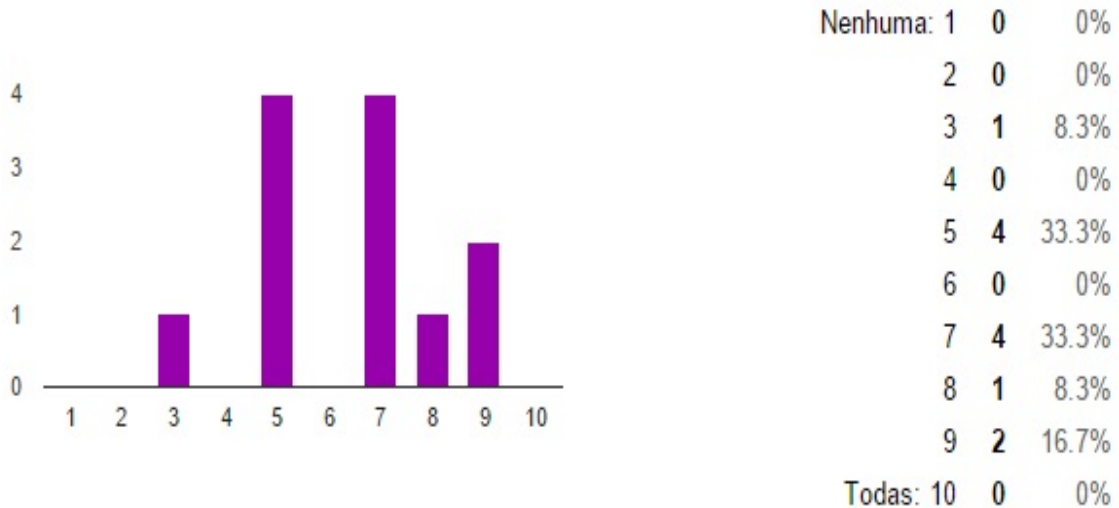


Figura 4.9: Resultados questionário: questão sobre a avaliação da conversão para GLOSA.

Na questão de avaliação de todos os gestos reproduzidos no vídeo (tabela 4.1) os resultados foram bons, pois obtiveram-se 16,7% na nota 5, 33,3% na nota 7 e 25% na nota 8 (figura 4.10) significando que o sistema reproduz a maior parte dos vídeos corretamente e que o sistema necessita de uns pequenos melhoramentos nesta área (secção 3.3).

Avaliação dos vídeos em LGP: tradução para LGP.

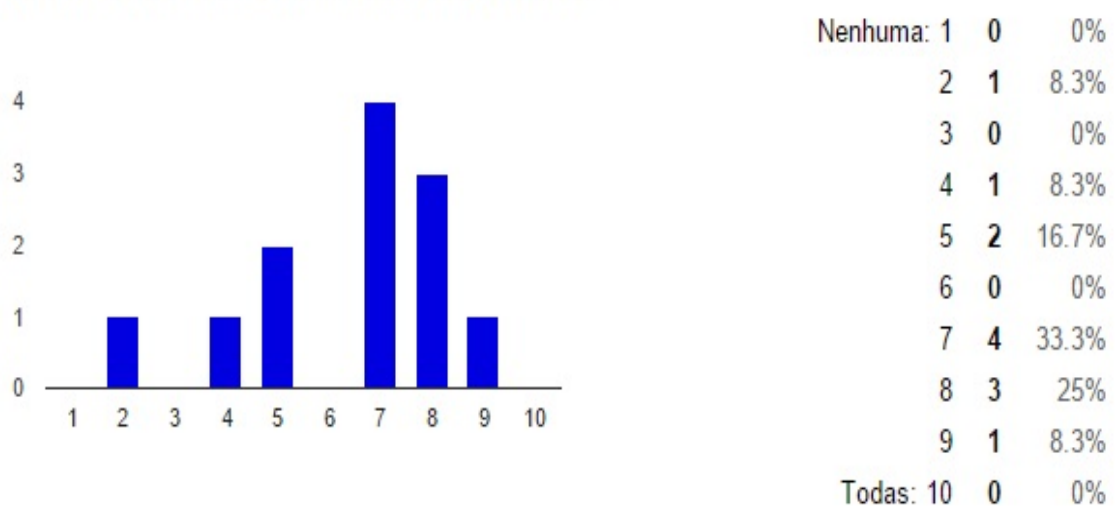
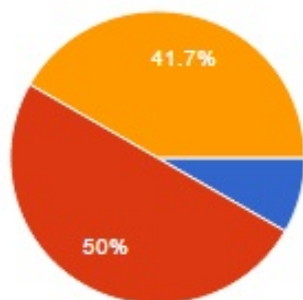


Figura 4.10: Resultados questionário: questão sobre a avaliação dos vídeos em LGP.

Na questão do tamanho da janela de vídeo (tabela 4.1) obtiveram-se resultados bipartidos: 50% acham que a janela é grande e 41,7% acha que o tamanho está bom (figura 4.11)

significando que se pode diminuir um pouco a janela, visto que em caso de necessidade poder-se-á aumentar pois o *Google Chrome* possibilita essa opção (secção 3.3).

Avaliação dos vídeos em LGP: tamanho da janela de vídeo.



Pequeno	1	8.3%
Grande	6	50%
Bom	5	41.7%

Figura 4.11: Resultados questionário: questão sobre o tamanho da janela de vídeo.

Na questão da qualidade visual dos vídeos (tabela 4.1) os resultados foram baixos, tiveram-se: 8,3% nas notas 2, 6, 7 e 10 e 25% nas notas 4 e 5 (figura 4.12). A solução para aumentar a qualidade visual passará por gravar novos vídeos dos gestos com um gestuante experiente na LGP de forma a garantir expressões corporais e faciais que sejam bem aceites pelos surdos e garantir iluminação simétrica; melhorar a fluidez e desenvolver um algoritmo de transições de vídeos de forma mais suave como trabalho futuro.

Avaliação dos vídeos em LGP: qualidade visual.



Muito Má: 1	0	0%
2	1	8.3%
3	0	0%
4	3	25%
5	3	25%
6	1	8.3%
7	1	8.3%
8	2	16.7%
9	0	0%
Muito Boa: 10	1	8.3%

Figura 4.12: Resultados questionário: questão sobre a qualidade visual dos vídeos.

De acordo com as opiniões dos surdos e intérpretes, o sistema é difícil utilização em contexto real. Baseiam a sua opinião em duas principais razões: o tempo entre a frase reconhecida e a sua representação em LGP em vídeo é elevado e, o sistema não realiza uma comunicação bidirecional.

Em relação à primeira razão, eles referem que esse tempo atrasa muito a recepção da informação proveniente da pessoa ouvinte que em ambiente real é muito problemático. Um exemplo real é o contexto de sala: o professor é obrigado a esperar pela tradução do sistema o que iria atrasar a sua aula, visto que o sistema só permite reconhecimentos até 60 segundos de duração e depois serão convertidos para LGP em vídeo (secção 3.1). Assim, os surdos referem que preferem os intérpretes. Será necessário em trabalho futuro: aumentar o tempo de reconhecimento com outro reconhecedor de fala, visto que o reconhecedor usado possui esta limitação; otimizar os algoritmos da MT e da *interface* gráfica a fim de diminuir este tempo.

Em relação à segunda razão, como o sistema só está preparado para realizar uma comunicação entre um ouvinte e surdo, levando mais uma vez os surdos a preferirem os intérpretes, pois com eles podem efetuar uma comunicação bidirecional.

De acordo com as opiniões dos inquiridos, pode-se concluir que estes consideram o sistema com limitações mas, acham-no interessante como meio de divulgação à comunidade ouvinte em geral e como ponto de partida para o desenvolvimento de um sistema mais robusto e evoluído. Eles consideram o sistema interessante pois com ele qualquer ouvinte inexperiente em LGP pode aprender várias temáticas e regras da LGP, nomeadamente: algumas temáticas do quotidiano e do curso de engenharia eletrotécnica, as cores, a fruta, a numeração, o feminino, o vestuário, o alfabeto, os animais, os verbos regulares e os verbos SER e ESTAR e tempos verbais.

Em suma, será necessário melhorar cada módulo (secções 3.1, 3.2 e 3.3), bem como o *website* de interligação dos módulos (secção 4), a fim de obter melhor usabilidade e satisfação dos utilizadores.

Capítulo 5

Conclusões e trabalho futuro

Com este trabalho propôs-se desenvolver um sistema automático de fala em **LP** para **LGP** em vídeo de forma a possibilitar a comunicação entre ouvintes e surdos na ausência de intérpretes sobre assuntos do quotidiano e conteúdos do curso de engenharia eletrotécnica. Este trabalho está assente em 3 módulos distintos: reconhecimento de fala, máquina de tradução e *interface* gráfica.

Relativamente ao reconhecedor de fala foi usado o *Web Speech API*. É uma ferramenta da *Google* que permite a ligação do sinal acústico ao servidor da *Google* e este devolve os resultados dos reconhecimentos em texto na aplicação desenvolvida. Esta ferramenta só permite reconhecimentos com duração máxima de 60 segundos e para o português europeu apresenta eficiências totais aos níveis da frase e da palavra de 50% e 77,55%, respetivamente, no seu melhor caso - fala lenta. A avaliação, feita por um conjunto restrito de pessoas, apresentou resultados baixos neste módulo. Assim, como trabalho futuro, será necessário a utilização de um reconhecedor de fala em português europeu que não ofereça uma limitação tão baixa de tempo de reconhecimento e que ofereça eficiências totais aos níveis da frase e da palavra superiores.

Relativamente à **MT** esta faz a conversão do texto do módulo anterior para texto em **LGP**. A **MT** está preparada para frases de várias temáticas e regras da **LGP**, nomeadamente: as cores, o feminino, o vestuário, os transportes, os nomes próprios, a fruta, a numeração, os animais, os verbos regulares e os verbos SER e ESTAR, tempos verbais, a interrogativa e a exclamativa. Na avaliação pelos inquiridos, este módulo apresenta resultados razoáveis. Assim, como trabalho futuro, será necessário acrescentar mais regras da **LGP** e otimizar o algoritmo deste módulo a fim de possibilitar a conversão de texto mais complexo em **LP**.

No módulo da *interface* gráfica faz-se a conversão do texto em **LGP** (proveniente da **MT**) para um vídeo final. Na avaliação apresenta uma boa representação em vídeo das mensagens mas, é necessário melhorar a qualidade visual e o tamanho de janela de

vídeo é considerado grande. Assim, como trabalho futuro, será necessário fazer gravações dos vídeos dos gestos a um gestuante experiente em **LGP** num estúdio, que garanta boa luminosidade e captura das expressões faciais e corporais muito necessárias em **LGP**, a fim de garantir melhor aceitabilidade dos utilizadores.

No *website* faz-se a interligação dos módulos anteriores e obtiveram-se resultados bons na usabilidade e razoáveis no seu visual. Em trabalho futuro será necessário melhorar a usabilidade e o seu visual para ser mais bem aceite pelos utilizadores.

Na avaliação global do sistema desenvolvido, os inquiridos dizem que o sistema é difícil aplicação em contexto real mas interessante no ensino. Põem em causa a sua aplicação em contexto real, pois o sistema apresenta: limitação de tempo até 60 segundos nos reconhecimentos; muito tempo entre a frase proferida e sua representação em **LGP** e de não possibilitar uma comunicação bidirecional, não permitindo uma simples conversa quotidiana. Acham-no interessante no ensino, pois o sistema consegue mostrar várias temáticas e regras da **LGP**. Deste modo, se reforça a necessidade de melhoramento em todos os módulos a fim de permitir a sua utilização em ambiente real.

Não foi um trabalho trivial, exigiu muita pesquisa e estudo de várias tecnologias então desconhecidas: linguagens de programação *javascript*, **PHP** e **HTML**, *bootstrap*, **FFMPEG** e estudos das línguas **LP** e **LGP**, ou seja, não foi um caminho linear. Apesar das diversas limitações que o sistema apresenta estão lançadas as bases para que este sistema seja melhorado e que depois seja disponibilizado e utilizado pela população em geral. O grande ponto forte deste sistema é a sua arquitetura modular, ou seja, o melhoramento de um módulo não põe em causa o funcionamento dos outros dois módulos, permitindo assim que várias pessoas possam trabalhar e melhorar este trabalho.

Como este sistema só permite a comunicação de fala para gestos em **LGP**, será muito interessante desenvolver um sistema que seja complementar a este, de forma a possibilitar uma comunicação bidirecional. Partindo da captura dos gestos por câmaras e depois fazer a conversão para fala em **LP** é extremamente difícil (capítulo 2), a solução mais viável é desenvolver um sistema que faça a conversão de GLOSA para fala em **LP**.

No âmbito deste trabalho foi submetido um artigo a uma conferência nacional.

Bibliografia

- [1] W. H. Organization. (Março 2015) Deafness and hearing loss. Acedido: 10-03-2015. [Online]. Disponível: <http://www.who.int/mediacentre/factsheets/fs300/en/>
- [2] M. Pacheco, “Formação de língua gestual portuguesa - nível i,” 2014, curso de Iniciação de Língua Gestual Portuguesa.
- [3] D. Silva. Comunica com as mãos! Acedido: 07-10-2014. [Online]. Disponível: http://partilhascruzadas.files.wordpress.com/2011/07/lgp_diana.pdf
- [4] A. S. L. University. American sign language - fingerspelling and numbers: Introduction. Acedido: 07-10-2014. [Online]. Disponível: <http://www.lifeprint.com/ASL101/fingerspelling/fingerspelling.htm>
- [5] G. Shires. Voice driven web apps: Introduction to the web speech api. Acedido: 05-10-2014. [Online]. Disponível: <http://updates.html5rocks.com/2013/01/Voice-Driven-Web-Apps-Introduction-to-the-Web-Speech-API>
- [6] Meaning. Lemmatization, pos and parsing sdks. Acedido: 15-02-2015. [Online]. Disponível: <https://www.meaningcloud.com/developer/lemmatization-pos-parsing/sdks>
- [7] FFmpeg. Ffmpeg. Acedido: 27-05-2015. [Online]. Disponível: <https://www.ffmpeg.org/>
- [8] ——. Concatenating media files. Acedido: 27-05-2015. [Online]. Disponível: <https://trac.ffmpeg.org/wiki/Concatenate>
- [9] C. Veloso, “Tsf fez emissão de rádio adaptada para surdos,” 2005, acedido: 27-09-2014. [Online]. Disponível: <http://www.publico.pt/media/jornal/tsf-fez-emissao-de-radio-adaptada-para-surdos-14762>
- [10] Programa da manhã da antena1 com língua gestual. Antena1. Acedido: 27-09-2014. [Online]. Disponível: <http://www.rtp.pt/extra/index.php?article=258&visual=4>

- [11] L. Bonixe, *Emissão para surdos na rádio pública*, Novembro 2013, acessido: 27-09-2014. [Online]. Disponível: <http://radioejornalismo.blogspot.pt/2013/11/emissao-para-surdos-na-radio-publica.html>
- [12] C. youtube do Grupo First. Serviin no benficatv - entrevista com nuno carvalho. Acessido: 27-09-2014. [Online]. Disponível: <https://www.youtube.com/watch?v=rEURyGqpp0Y>
- [13] P. do Cidadão Surdo. Notícias. Acessido: 27-09-2014. [Online]. Disponível: <http://www.portaldocidadaosurdo.pt/Not%C3%ADcias.aspx>
- [14] Rtp legenda noticiários a partir de sexta-feira numa iniciativa inédita a nível mundial. LUSA - Agência de Notícias de Portugal, S.A. Acessido: 03-10-2014. [Online]. Disponível: https://www.l2f.inesc-id.pt/wiki/index.php/RTP_legenda_notici%C3%A1rios_a_partir_de_sexta-feira_numa_iniciativa_in%C3%A9dita_a_n%C3%ADvel_mundial
- [15] Língua gestual. Acessido: 03-10-2014. [Online]. Disponível: <http://www.rtp.pt/wportal/acessibilidades/gestual/>
- [16] Rtp na linha da frente. Acessido: 03-10-2014. [Online]. Disponível: <http://magazineconsigo.blogspot.pt/2011/12/rtp-na-linha-da-frente.html>
- [17] V. Portugal. (2012) Vodafone portugal lança primeiro serviço de apoio a clientes em língua gestual portuguesa. Acessido: 29-09-2014. [Online]. Disponível: <http://press.vodafone.pt/2012/04/27/responsabilidade-social/>
- [18] L. Simões, “19 de maio - peregrinação de surdos a fátima,” acessido: 30-09-2014. [Online]. Disponível: <http://www.santuario-fatima.pt/portal/index.php?id=63736>
- [19] Missas em fátima vão ter língua gestual. Rádio Renascença. Acessido: 30-09-2014. [Online]. Disponível: http://rr.sapo.pt/informacao_detalhe.aspx?fid=29&did=107218
- [20] A. C. Silva, C. M. Coelho, F. Cabedo, M. Sousa, “Os novos públicos da rádio – relato de um projeto de uma *webradio* para surdos,” 105-111 *Média e Inclusão*, Outubro 2013.
- [21] Spread the sign. União Europeia. Acessido: 07-10-2014. [Online]. Disponível: <http://www.spreadthesign.com/pt/>
- [22] L. D. Deusdado, ” “Ensino da língua gestual assistido por personagens 3d virtuais, Universidade do Minho, 2002, dissertação de Mestrado.

- [23] J. C. M. Bento, ” “Avatares em língua gestual portuguesa, Universidade de Lisboa - Faculdade de Ciências, 2013, dissertação de Mestrado.
- [24] S. Babu K., R. Kumar K.K., S. M. Idicula, ” *in the book Intelligent Computing Models* “Generating american sign language asl gestures from speech for assisting deaf-mute people. New Delhi: Narosa Publications, 2009.
- [25] Informação do icommunicator. Interactive Solutions, Inc. Acedido: 30-09-2014. [Online]. Disponível: <http://www.icommunicator.com/index.html>
- [26] Mimix. Acedido: 14-10-2014. [Online]. Disponível: <http://www.mimix.me/>
- [27] N. M. e. C. L. C. Brito, Patrick Henrique Silva e Franco, “Falibras: Uma ferramenta flexível para promover acessibilidade de pessoas surdas,” 8 *TISE–Nuevas Ideas em Informatica Educativa*, 2012.
- [28] ProDeaf. Acedido: 12-10-2014. [Online]. Disponível: <http://www.prodeaf.net/>
- [29] Hand Talk. Acedido: 14-10-2014. [Online]. Disponível: <http://www.handtalk.me/>
- [30] Poli-libras. Acedido: 13-10-2014. [Online]. Disponível: <http://www.polilibras.com.br/>
- [31] S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, S. Abbott, ” *Proceedings of the fifth international ACM conference on Assistive technologies* “Tessa, a system to aid communication with deaf people. pp. 205–212 ACM, 2002.
- [32] R. San-Segundo, R. Barra, R. Córdoba, L. D’Haro, F. Fernández, J. Ferreiros, J. M. Lucas, J. Macías-Guarasa, J. M. Montero, J. M. Pardo, “Speech to sign language translation system for spanish,” vol. 50 *Speech Communication*.
- [33] R. San Segundo, V. López, R. Martín, S. L. Lutfi, J. Ferreiros, R. de Córdoba, J. M. Pardo, ” *INTERSPEECH* “Advanced speech communication system for deaf people, 250–253 2010.
- [34] A. A. Haseeb A. Ilyas, ” “Speech translation into pakistan sign language, School of Computing at Blekinge Institute of Technology, Sweden, 2012, dissertação de Mestrado.
- [35] J. Bungeroth H. Ney, ” *Gesture in Human-Computer Interaction and Simulation* “Automatic generation of german sign language glosses from german words, pp. 49–52 2006.

- [36] E. Efthimiou, G. Sapountzaki, K. Karpouzis, S.-E. Fotinea, ” *Computers Helping People with Special Needs* “Developing an e-learning platform for the greek sign language, pp. 1107-1113 vol. 3118 2004.
- [37] V. Lombardo, F. Nunnari, R. Damiano, ” *Intelligent Virtual Agents* “A virtual interpreter for the italian sign language, pp. 201–207 2010.
- [38] K. K. El-Darymli, O. O. Khalifa, H. Enemosah, ” *the IEEE International Conference of Computer and Communication Engineering (ICCCE'06)* “Speech to sign language interpreter system (sslis), Kuala Lumpur, Malaysia, 2006.
- [39] D. Song, X. Wang, X. Xu, “Chinese sign language synthesis system on mobile device,” 986–992, vol. 29, *Procedia Engineering, 2012 International Workshop on Information and Electronics Engineering*, 2012.
- [40] T. Kuroda, K. Sato, K. Chihara, ” *Proceedings of 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech. Skövde, Sweden* “S-tel: An avatar based sign language telecommunication system, 1998.
- [41] N. e. R. R. e. B. M. e. B. J. Escudeiro, Paula e Escudeiro P. e. L. J. Baltasar, Ana Bela e Rodrigues, “Virtual sign supporting inclusive learning and equity,” *Computers and Technology in Modern Education*, 2014.
- [42] Isep apresenta virtual sign. Acedido: 15-10-2014. [Online]. Disponível: https://www.facebook.com/video.php?v=671978426217114&comment_id=693181464096810&offset=0&total_comments=9/
- [43] A. Leibs. The icommunicator translates speech into text or sign language. Assistive Technology Expert. Acedido: 30-09-2014. [Online]. Disponível: <http://assistivetechology.about.com/od/DHHSC1/p/The-Icommunicator-Translates-Speech-Into-Text-Or-Sign-Language.htm>
- [44] S. Tripathi J. K. Sarkhel, “Approaches to machine translation,” 388–393 *57 International journal of Annals of Library and Information Studies*, 2010.
- [45] D. Editora, *Dicionário da Língua Portuguesa 2003*. Porto Editora, 2003.
- [46] F. Sánchez Martínez, “Choosing the best machine translation system to translate a sentence by using only source-language information,” 2011.
- [47] J.-C. Junqua, *Robust Speech Recognition in Embedded Systems and PC Applications*. Kluwer Academic Publishers, 2000.

- [48] L. P. Becchetti, Claudio e Ricotti, *Speech Recognition Theory and C++ Implementation*. John Wiley & Sons, 1999.
- [49] B.-H. Rabiner, Lawrence e Juang, *Fundamentals of Speech Recognition*. Prentice-Hall International, Inc, 1993.
- [50] R. A. Bates, ” “Speaker dynamics as a source of pronunciation variability for continuous speech recognition models, University of Washington, 2003, dissertação de Doutorado.
- [51] Simon. KDE UserBase Wiki. Acedido: 12-10-2014. [Online]. Disponível: <https://userbase.kde.org/Simon>
- [52] Simon/what do i need? KDE UserBase Wiki. Acedido: 12-10-2014. [Online]. Disponível: https://userbase.kde.org/Simon/What_do_I_need%3F
- [53] Simon/base models. KDE UserBase Wiki. Acedido: 12-10-2014. [Online]. Disponível: https://userbase.kde.org/Simon/Base_models
- [54] Audimus. INESC - Laboratório de sistemas de Língua Falada. Acedido: 11-10-2014. [Online]. Disponível: <https://www.l2f.inesc-id.pt/wiki/index.php/AUDIMUS>
- [55] Audimus server. INESC - Laboratório de sistemas de Língua Falada. Acedido: 11-10-2014. [Online]. Disponível: http://www.voiceinteraction.pt/?page_id=374
- [56] Julius. Julius Team. Acedido: 13-10-2014. [Online]. Disponível: http://julius.osdn.jp/en_index.php?q=index-en.html#about_models
- [57] 360 | speechmagic sdk: Deeply embed advanced speech recognition capabilities. Nuance. Acedido: 14-10-2014. [Online]. Disponível: http://www.nuance.com/ucmprod/groups/healthcare/@web-enuk/documents/collateral/nc_031043.pdf
- [58] Deeply embed advanced speech recognition capabilities. Nuance. Acedido: 14-10-2014. [Online]. Disponível: <http://www.nuance.com/products/speechmagic/index.htm>
- [59] Skype. Microsoft. Acedido: 17-10-2014. [Online]. Disponível: <http://www.skype.com/pt/translator-preview/>
- [60] ispeech. iSpeech,Inc. Acedido: 18-10-2014. [Online]. Disponível: <https://www.ispeech.org/developers>
- [61] Class speechrecognizer. iSpeech,Inc. Acedido: 18-10-2014. [Online]. Disponível: <http://www.ispeech.org/instructions/sampleprojects/android/Documentation/index.html>

- [62] Microsoft speech platform. Microsoft. Acedido: 08-10-2014. [Online]. Disponível: [https://msdn.microsoft.com/en-us/library/hh361572\(v=office.14\).aspx](https://msdn.microsoft.com/en-us/library/hh361572(v=office.14).aspx)
- [63] Microsoft speech platform sdk 11 requirements and installation. Microsoft. Acedido: 08-10-2014. [Online]. Disponível: [https://msdn.microsoft.com/en-us/library/hh362873\(v=office.14\).aspx](https://msdn.microsoft.com/en-us/library/hh362873(v=office.14).aspx)
- [64] Microsoft speech platform - runtime languages (version 11). Microsoft. Acedido: 08-10-2014. [Online]. Disponível: <https://www.microsoft.com/en-us/download/details.aspx?id=27224>
- [65] Web speech api specification. W3C. Acedido: 05-10-2014. [Online]. Disponível: <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>
- [66] G. Shires. webspeechdemo.html. Acedido: 05-10-2014. [Online]. Disponível: <https://github.com/GoogleChrome/webplatform-samples/blob/master/webspeechdemo/webspeechdemo.html>
- [67] ——. Web speech api - creates interactive experiences. Google Developers. Acedido: 05-10-2014. [Online]. Disponível: <http://commondatastorage.googleapis.com/io-2013/presentations/4057%20Web%20Speech%20API%20creates%20Interactive%20Experiences%20-%20Google%20I-O%202013%20Codelab.pdf>
- [68] Find your dragon for pc. Nuance. Acedido: 08-10-2014. [Online]. Disponível: <http://www.nuance.com/for-individuals/by-product/dragon-for-pc/index.htm>
- [69] *Nuance Dragon NaturallySpeaking 13 comparison by product*, Nuance, 2014, acedido: 08-10-2014. [Online]. Disponível: http://www.nuance.com/ucmprod/groups/corporate/@web-enus/documents/collateral/dns13_pro-pre-homefeatures.pdf
- [70] Dragon. Nuance. Acedido: 08-10-2014. [Online]. Disponível: http://www.nuance.com/ucmprod/groups/dragon/@web-enus/documents/collateral/nc_017505.pdf
- [71] J. Adorf, “Web speech api,” KTH Royal Institute of Technology, Stockholm, Sweden, 2013, relatório Técnico.
- [72] Google chrome: How to use the web speech api. StiltSoft. Acedido: 05-10-2014. [Online]. Disponível: <http://stiltsoft.com/blog/2013/05/google-chrome-how-to-use-the-web-speech-api/>
- [73] J. Duckett, *Beginning web programming with HTML, XHTML, and CSS*. John Wiley & Sons, 2008.

- [74] T. Payton. Python google speech to text api implementation. Acedido: 01-05-2015. [Online]. Disponível: <http://codeabitswiser.com/2014/09/python-google-speech-api/#comment-223>
- [75] HTK. Acedido: 26-06-2015. [Online]. Disponível: <http://htk.eng.cam.ac.uk/>
- [76] S. P. L. do Instituto de Telecomunicações. Conversor de grafemas para fonemas para português europeu. Acedido: 26-06-2014. [Online]. Disponível: <http://www.co.it.pt/~labfala/g2p/>
- [77] A. B. Baltazar, *Dicionário da Língua Gestual Portuguesa*. Porto Editora, 2010.
- [78] S. S. Isabel Mesquita, *Guia Prático de Língua Gestual Portuguesa - Ouvir o Silêncio*. Editora Nova Educação, 2008.
- [79] L. C. Celso Cunha, *Nova Gramática do Português Contemporâneo*. Edições João Sá da Costa, 2000.
- [80] AsiaOnline. Learn about machine translation (a.k.a. automated translation). Acedido: 17-01-2015. [Online]. Disponível: <http://www.asiaonline.net/EN/MachineTranslation/default.aspx?QID=1#QID1>
- [81] G. Thurmair, “Comparing different architectures of hybrid machine translation systems,” 2009.
- [82] J. Bungeroth H. Ney, ” 4 *Workshop on representation and processing of sign languages, LREC* “Statistical sign language translation, 2004.
- [83] Meaning. Lemmatization, pos and parsing. Acedido: 15-02-2015. [Online]. Disponível: <https://www.meaningcloud.com/developer/lemmatization-pos-parsing>
- [84] Php. Acedido: 08-02-2015. [Online]. Disponível: <https://secure.php.net/>
- [85] Meaning. Tagset for portuguese. Acedido: 15-02-2015. [Online]. Disponível: <https://www.meaningcloud.com/developer/documentation/morphosyntactic-tagsets>
- [86] Lgp. Acedido: 21-05-2015. [Online]. Disponível: <https://www.youtube.com/playlist?list=PL8858D9CEBEBB89E4>
- [87] Avs video editor 7.1. Acedido: 21-05-2015. [Online]. Disponível: <http://www.avs4you.com/avs-video-editor.aspx>
- [88] W3SCHOOLS. Html5 video. Acedido: 21-05-2015. [Online]. Disponível: http://www.w3schools.com/html/html5_video.asp

- [89] FFmpeg. Zerano ffmpeg. Acedido: 27-05-2015. [Online]. Disponível: <http://ffmpeg.zerano.com/builds/>
- [90] Bootstrap. Acedido: 20-03-2015. [Online]. Disponível: <http://getbootstrap.com/>
- [91] Frameworks. Acedido: 20-03-2015. [Online]. Disponível: <http://www.dsc.ufcg.edu.br/~jacques/cursos/map/html/frame/oque.htm>

Anexo A

Tutoriais

A.1 Tutorial Google Speech API

Este tutorial descreve como utilizar a Google Speech API a fim de analisar os ficheiros áudio no idioma português de Portugal para o servidor da Google e este devolve os resultados numa *string*.

Tutorial: Analisar ficheiros de áudio em Português de Portugal com Google Speech API

O Google Speech API possui as seguintes características:

- Requer ser membro do “Chromium-dev Group” – **Passo 1**
- Requer uma “server key” – **Passo 1**
- Apenas permite fazer 50 requisitos por dia ao servidor Google – **Passo 1**
- Requer ficheiros de áudio do tipo flac – **Passo 2**
- Os scripts de comunicação entre ficheiros áudio e o Google Speech API devem estar alojados num webserver ou local webserver – **Passo 3**

1. Habilitar “Speech API” no “Google Developers Console”

1.1. Ser membro do “Chromium-dev Group”.

1.1.1. Ver Imagem 1. Disponível:

<https://groups.google.com/a/chromium.org/forum/?fromgroups#!forum/chromium-dev>

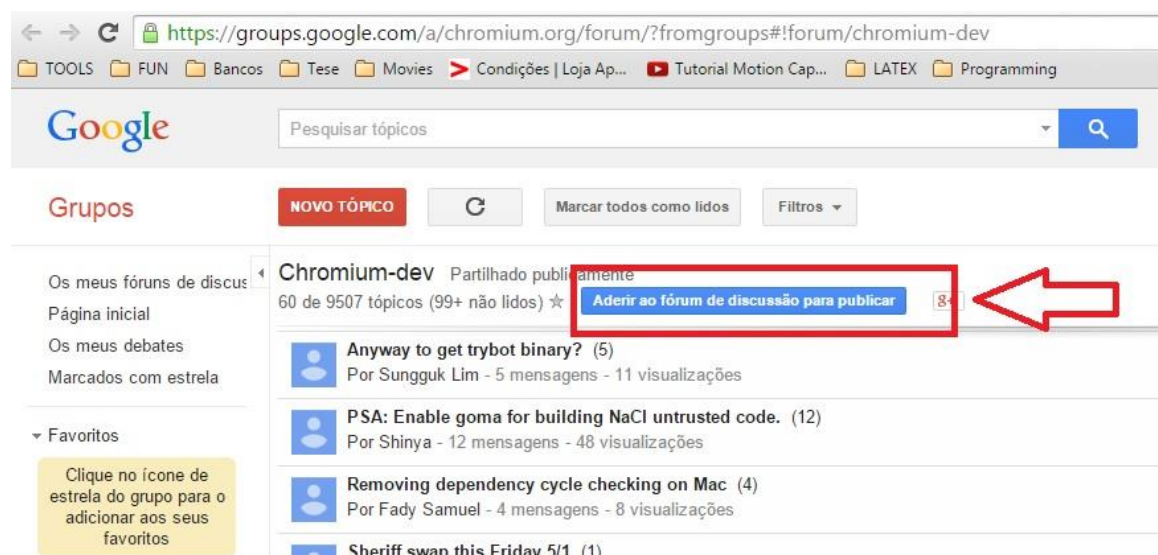


Imagem 1 - Ser membro do "chromium-dev discussion list"

1.2. Criar um novo projecto. Ver Imagem 2. Disponível:

<https://console.developers.google.com/project>

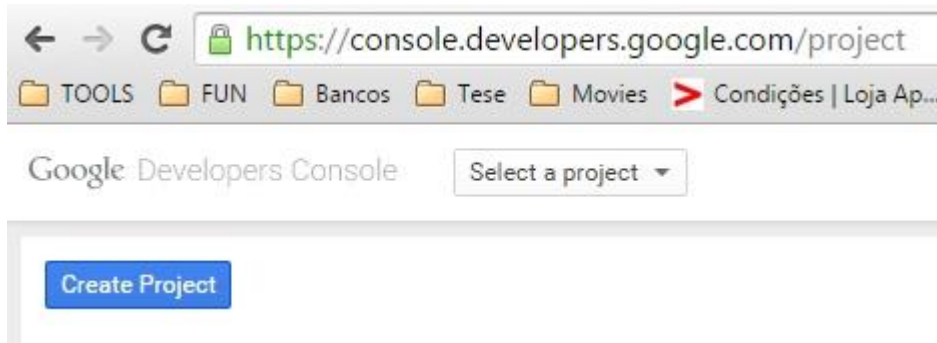


Imagem 2 - Criar novo projecto

1.3. Dar um nome ao projecto e criar. Pode alterar o ID ao projecto. Ver Imagem 3.

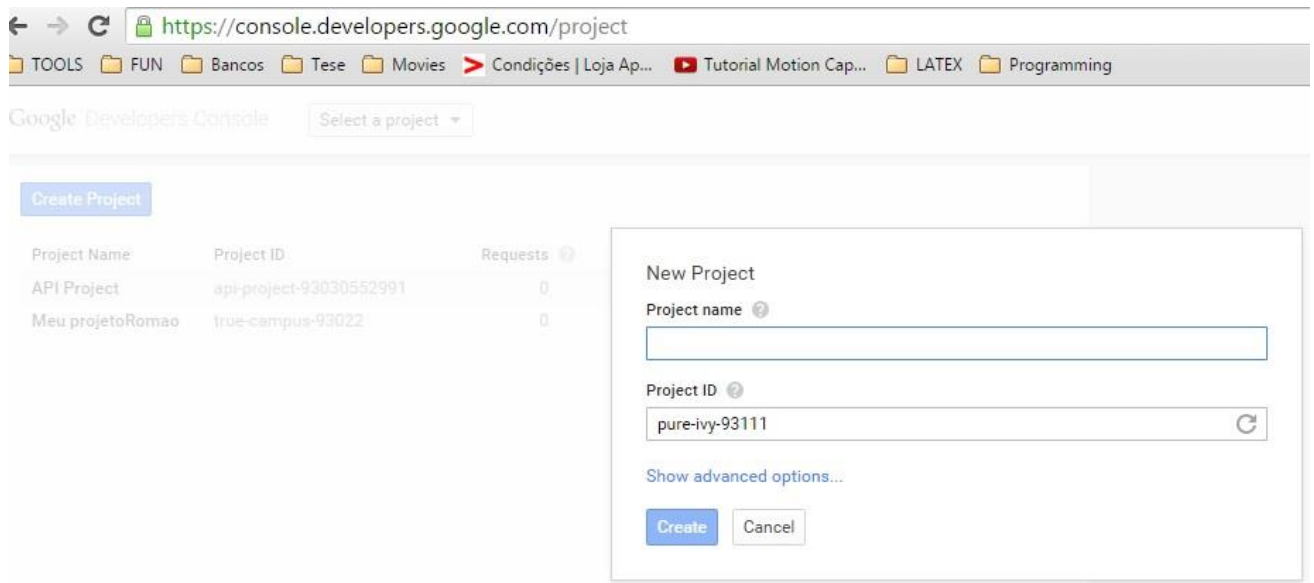


Imagem 3 - Nome ao projecto e projecto criado

1.4. Seleccionar o projecto recém criado

1.5. Procurar Speech API. Ver Imagem 4.

1.5.1. Primeiro carregar em APIs (caixa 1) e depois escrever “Google Speech” na caixa 2

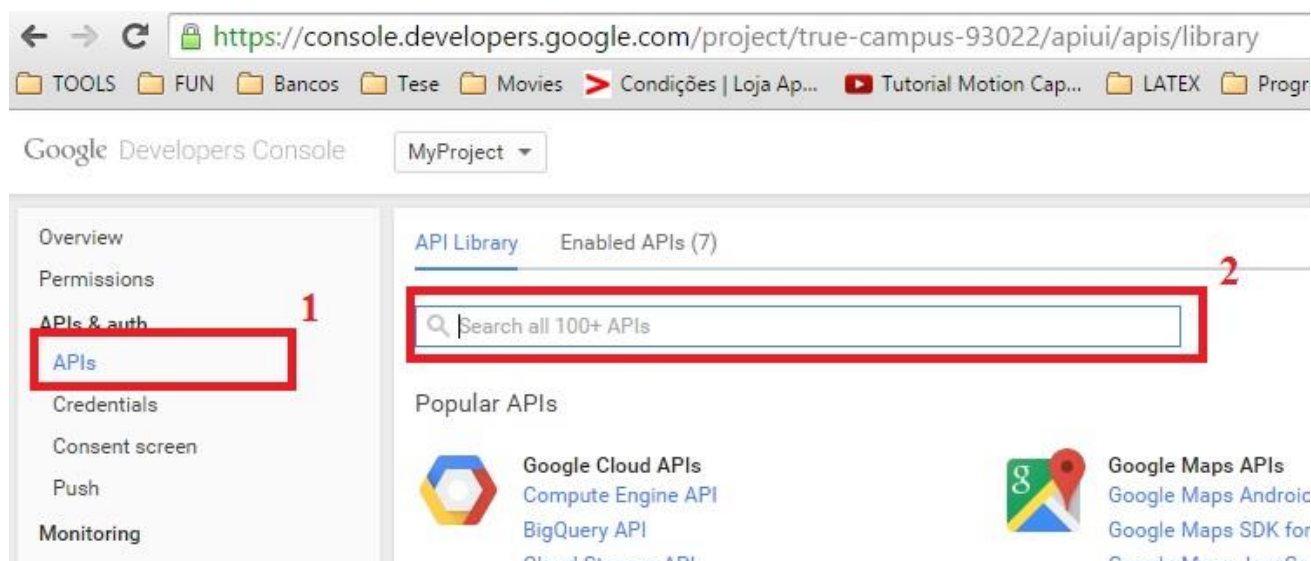


Imagem 4 – Procurar Google Speech

1.5.2. Habilitar “Speech API”. Ver Imagem 5.

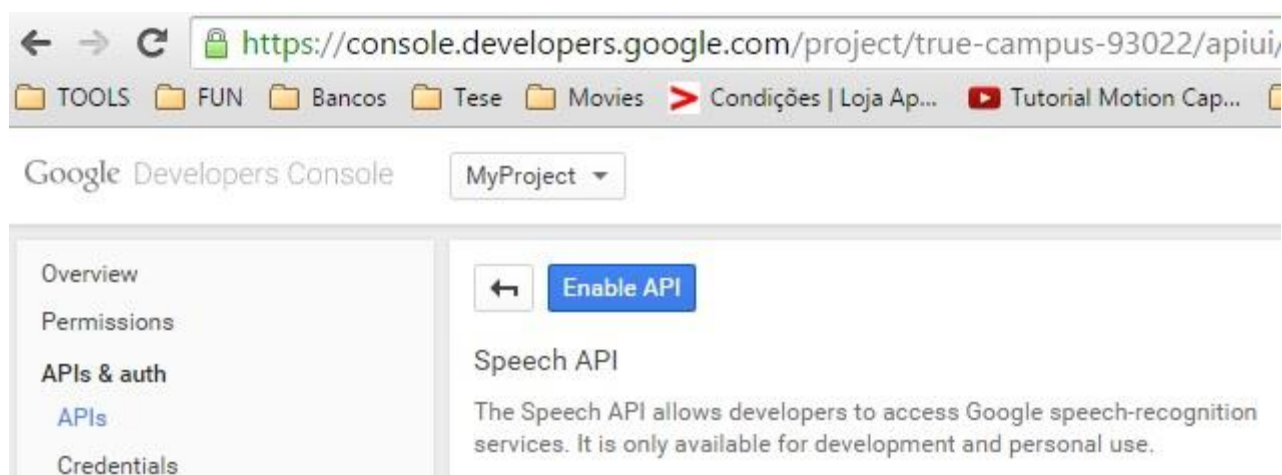
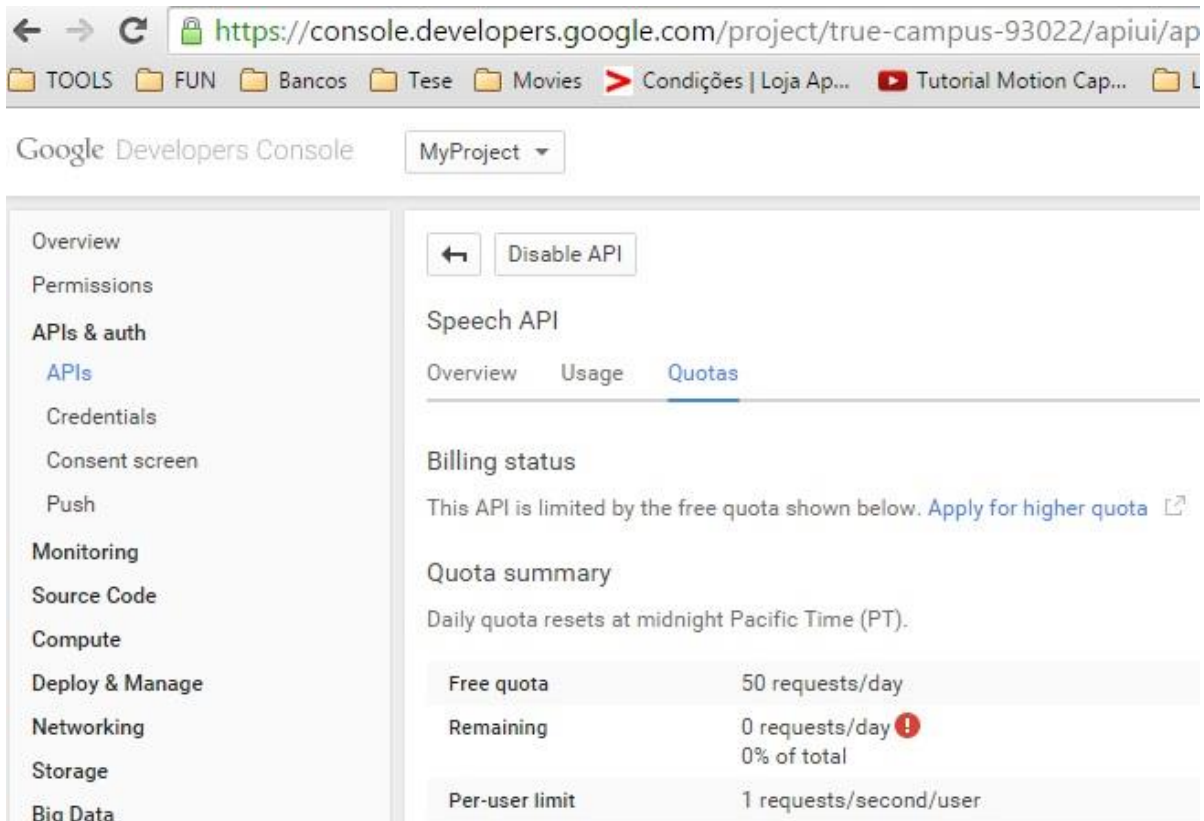


Imagem 5 – Habilitar Speech API

1.5.3. O “Speech API” só permite fazer 50 pedidos por dia ao servidor da Google. Ver Imagem 6.



The screenshot shows the Google Developers Console interface for the Speech API. The left sidebar contains a navigation menu with categories like Overview, Permissions, APIs & auth, Monitoring, and Compute. The main content area is titled "Speech API" and has tabs for Overview, Usage, and Quotas. Under the "Quotas" tab, there is a "Billing status" section with a warning that the API is limited by the free quota. Below that is a "Quota summary" section with a table showing the following details:

Category	Value
Free quota	50 requests/day
Remaining	0 requests/day ! 0% of total
Per-user limit	1 requests/second/user

Imagem 6 – Limite de pedidos por dia do Google Speech API

1.5.4. Para utilizar Google Speech API precisa-se obrigatoriamente possuir uma chave – “server key”. Ver Imagem 7 a Imagem 10.

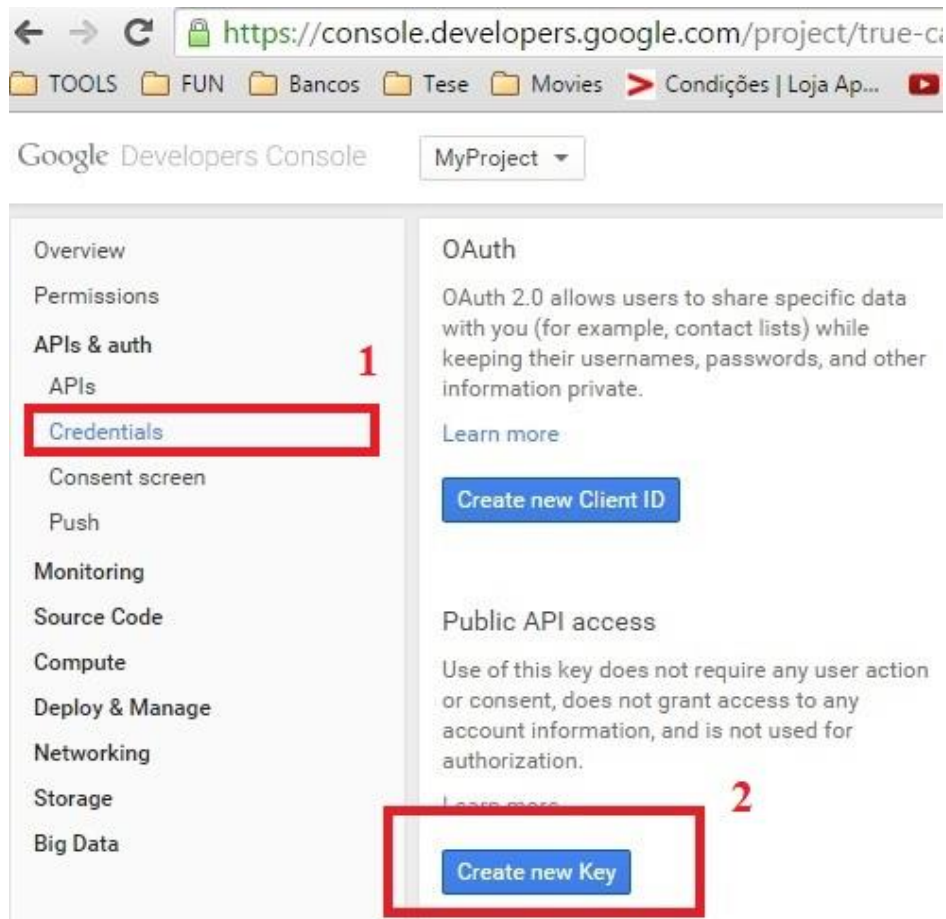


Imagem 7 – Criar uma chave para Google Speech API. Primeiro fazer caixa 1 e depois caixa 2.

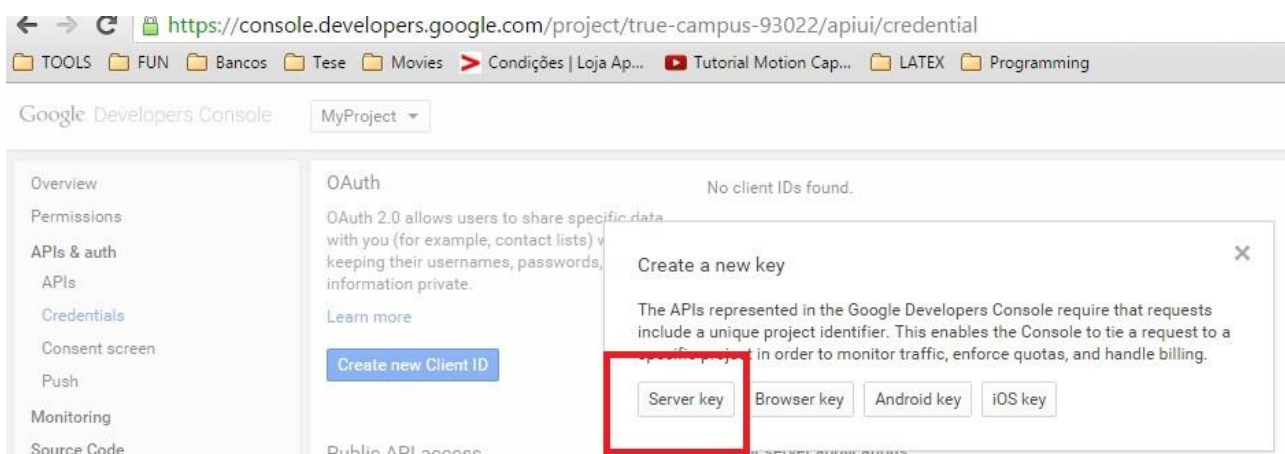


Imagem 8 - A chave a ser criada será do tipo "server"

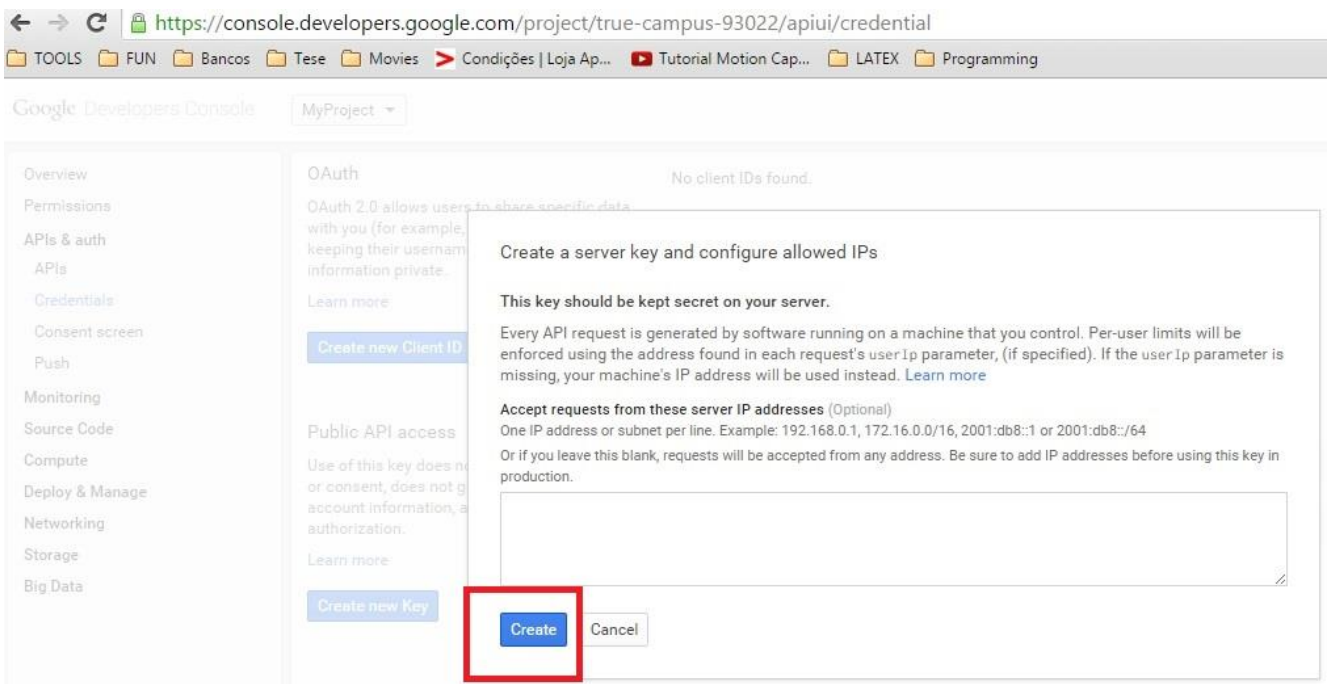


Imagem 9 - Criar chave

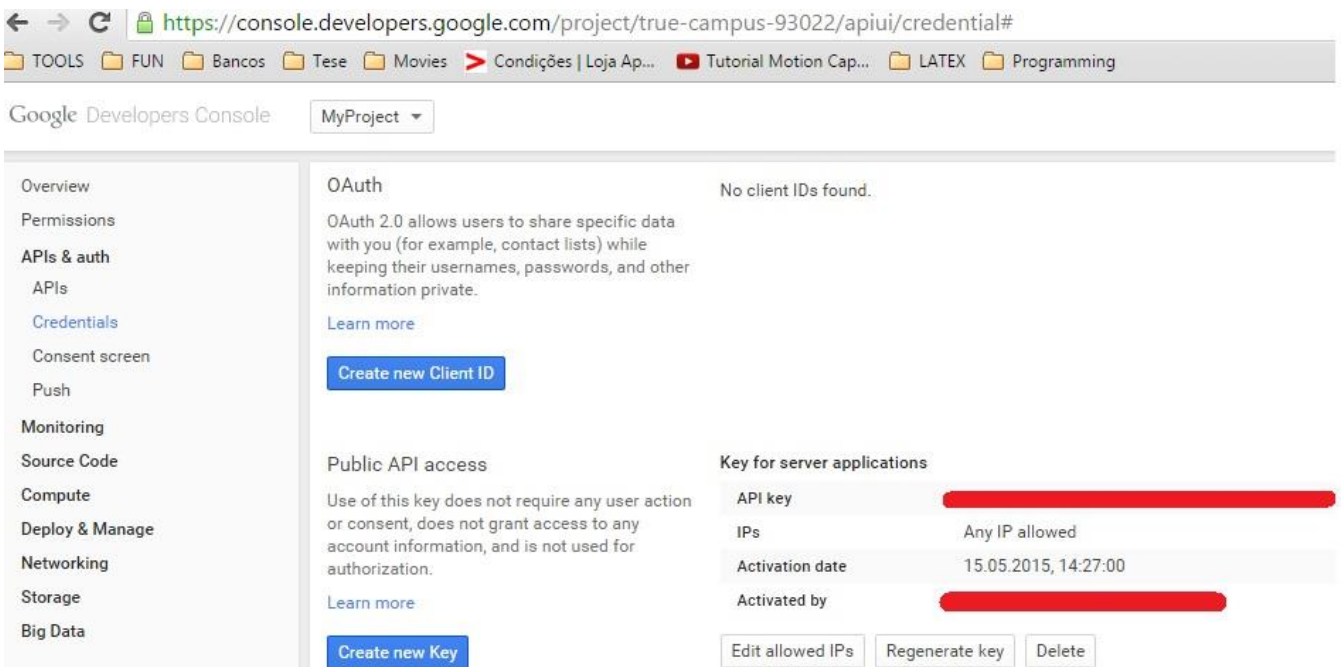


Imagem 10 - Chave criada (API Key). Utilizar chave (API Key) em Google Speech API

2. Criar ficheiros flac para Google Speech API

O Windows vem por defeito com gravador de voz. O gravador de voz requer o uso do microfone. Quando o utilizador faz a utilização do gravador de voz os ficheiros de áudio gerados terão a extensão wma, não possibilitando outras extensões, por exemplo, flac. Como o Google Speech API apenas funciona com ficheiros de áudio em formato flac é necessário recorrer a um programa que faça a conversão de wma para flac.

2.1. Gravar ficheiros wma com gravador do Windows. Utilizou-se o Windows 8.1.
Ver Imagem 11 a Imagem 14.



Imagem 11 – Procurar gravador de áudio no Windows 8.1. Escrever “gravador de áudio” em procurar

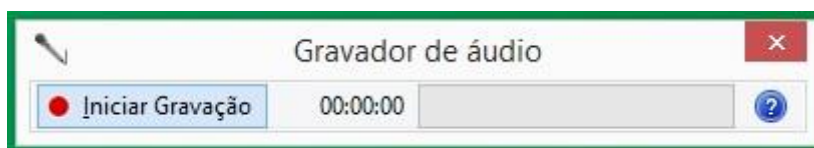


Imagem 12 – Iniciar gravação.

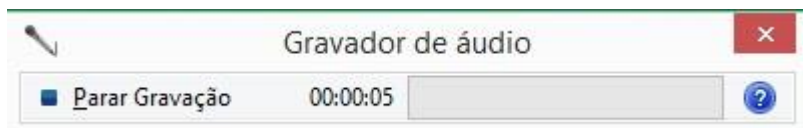


Imagem 13 – Parar gravação.

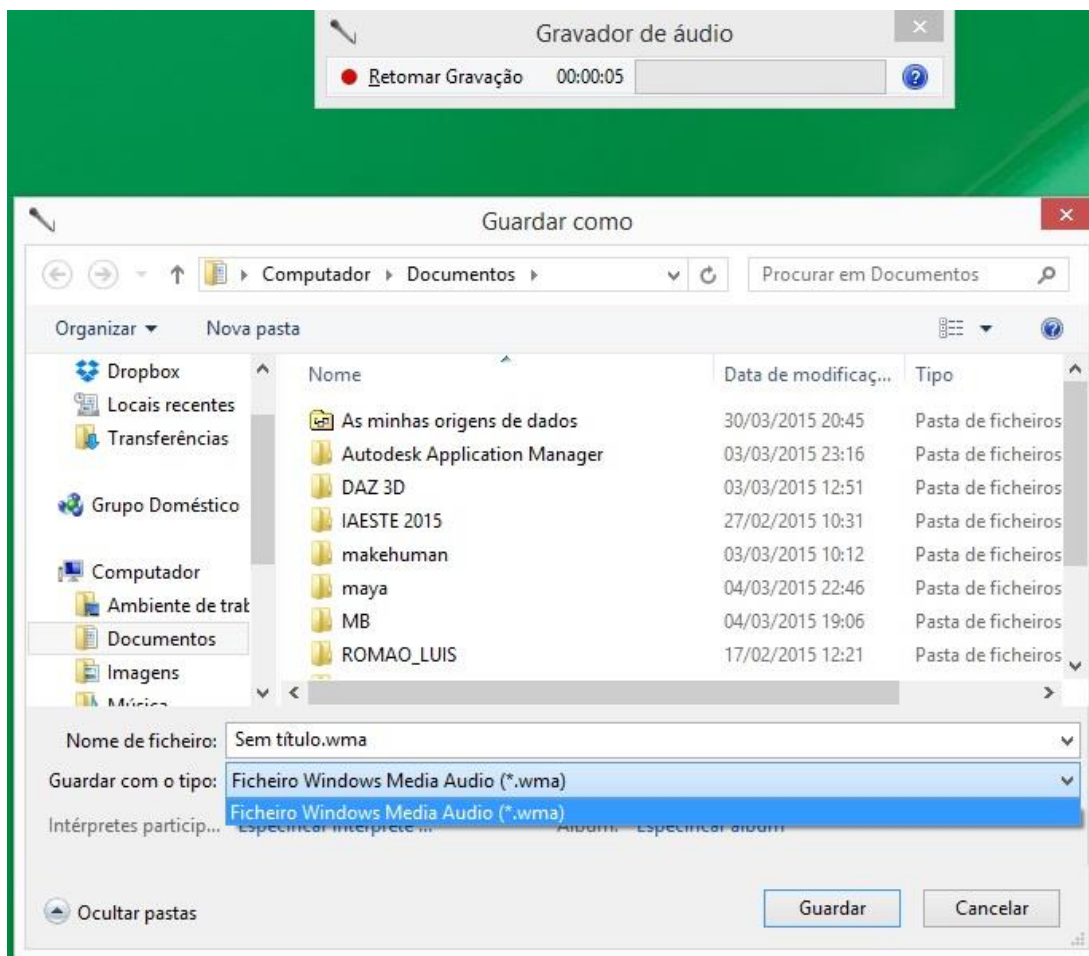


Imagem 14 – Guardar gravação.

Caso não se guarde a gravação, o gravador de voz continuará a captura de voz não possibilitando uma nova gravação. **Só guardando a gravação é possível efetuar uma nova captura.** Como se pode ver pela Imagem 14 o gravador apenas permite gravar em formato wma.

2.2. Conversão dos ficheiros wma para ficheiros flac

2.2.1. Software utilizado: “fre:ac”. Disponível: <http://www.freac.org/>

Características do “fre:ac”:

- Conversor de áudio grátis e CD ripper que suporta vários formatos e codificadores. Atualmente converte entre MP3, MP4/M4A, WMA, Ogg Vorbis, FLAC, AAC, WAV e formato Bonk.
- Interface de utilizador multilingue disponível em 40 idiomas
- Download para Windows, OS X, Linux e FreeBSD. Disponível: <http://www.freac.org/index.php/en/downloads-mainmenu-33/202-freac-1023>

2.2.2. Conversão com “fre:ac”. Ver Imagem 15 a Imagem 21

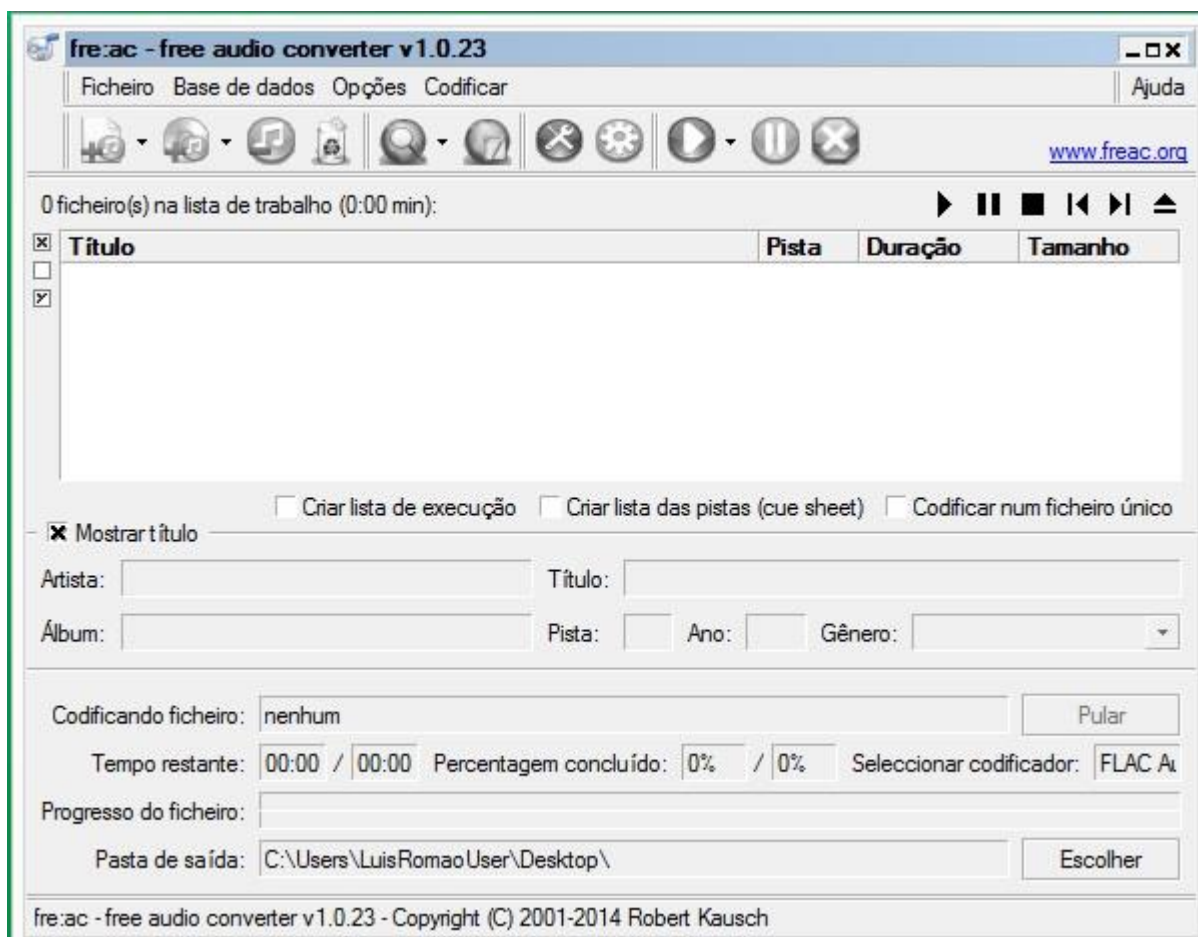


Imagem 15 – Início do software “fre:ac”

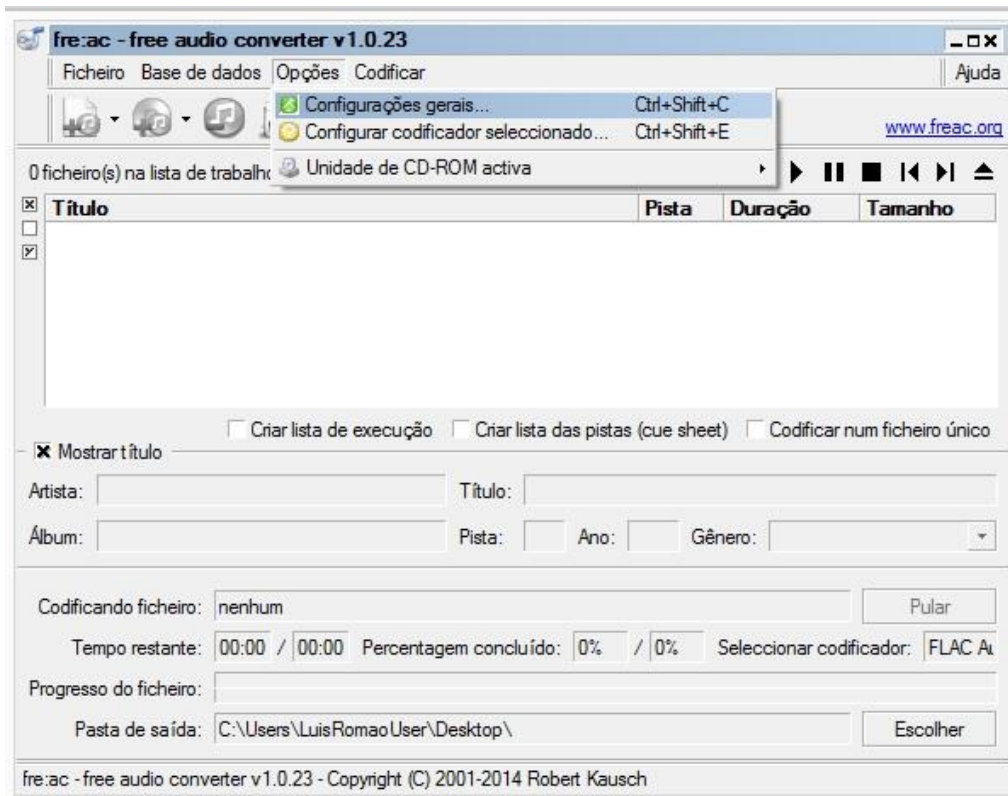


Imagem 16 – Configurações gerais.

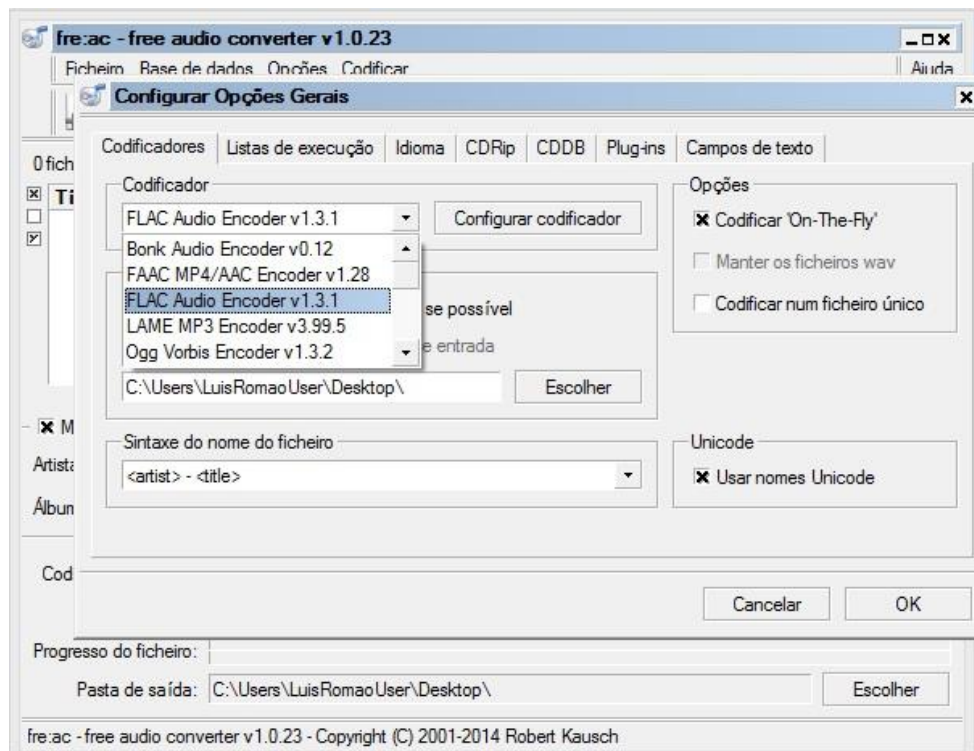


Imagem 17 – Configurações gerais: seleccionar para Codificador FLAC

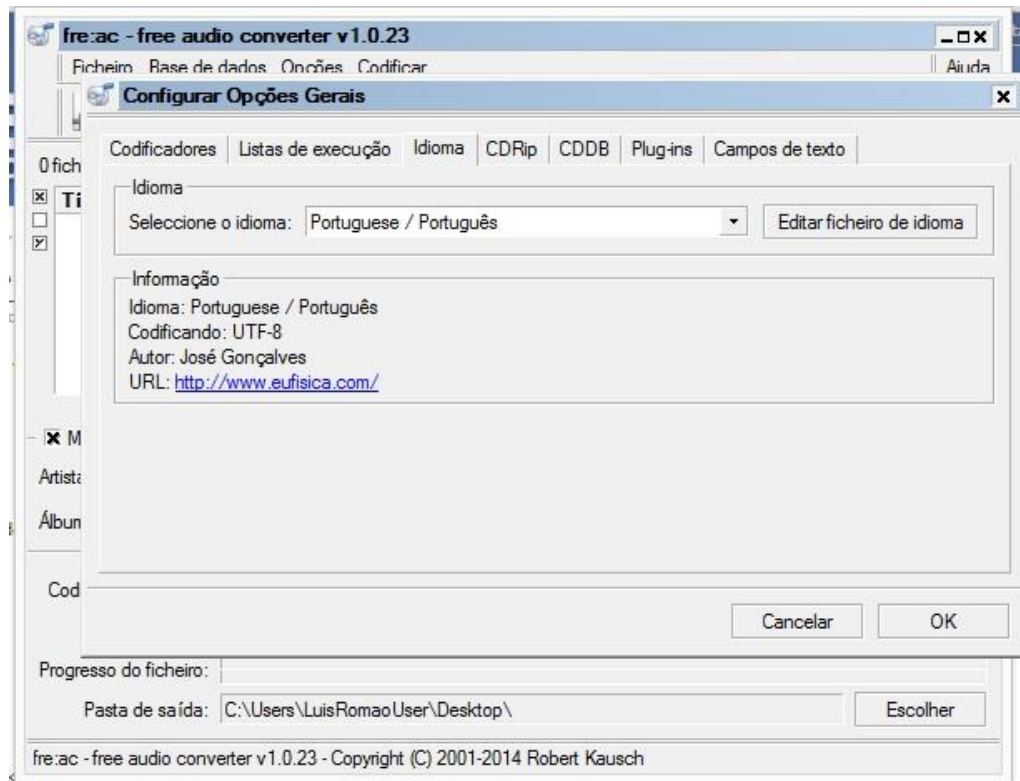


Imagem 18 - Configurações gerais: seleccionar idioma para Português de Portugal

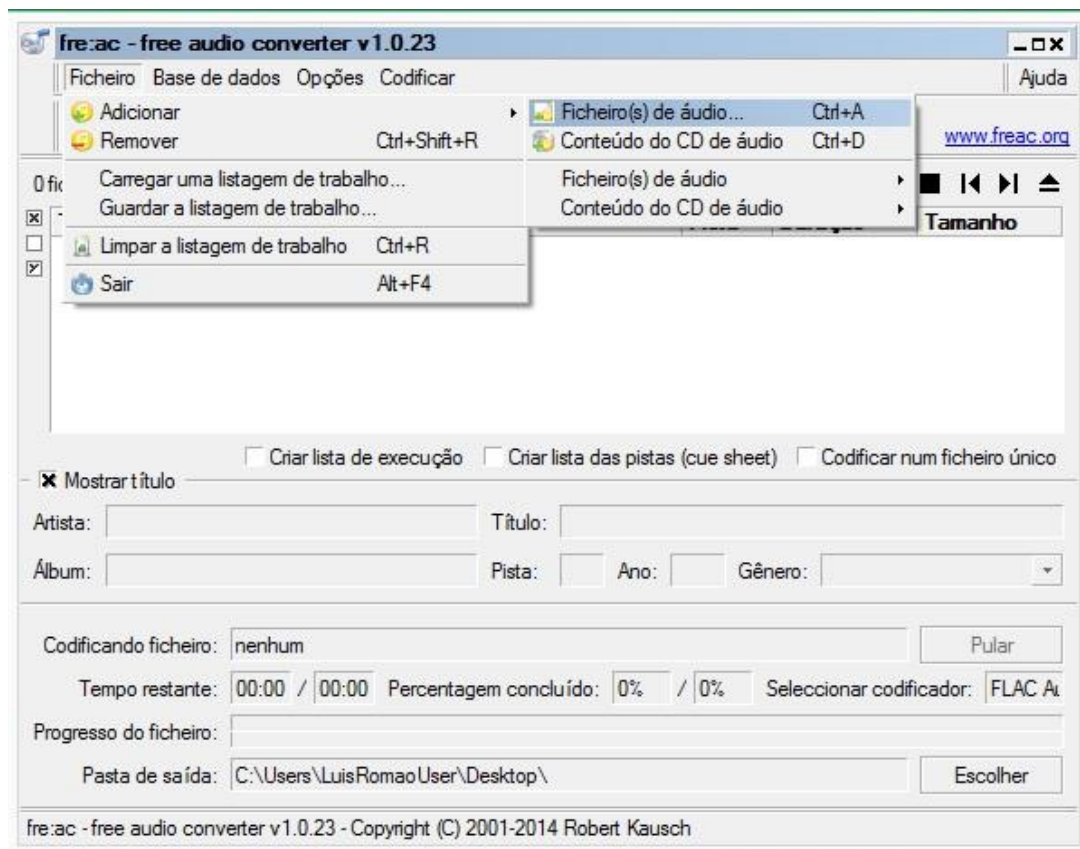


Imagem 19 – Adicionar ficheiros de áudio a serem convertidos pelo “fre:ac”

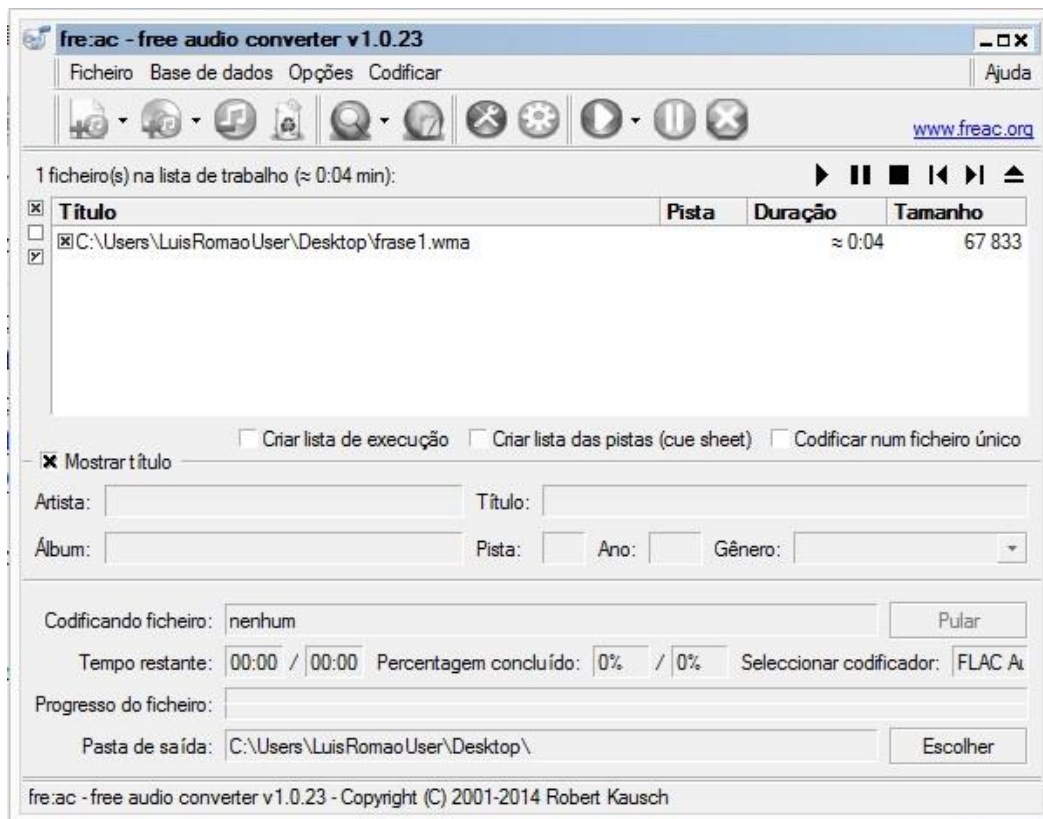


Imagem 20 – Ficheiros adicionados com êxito ao “fre:ac” que serão convertidos

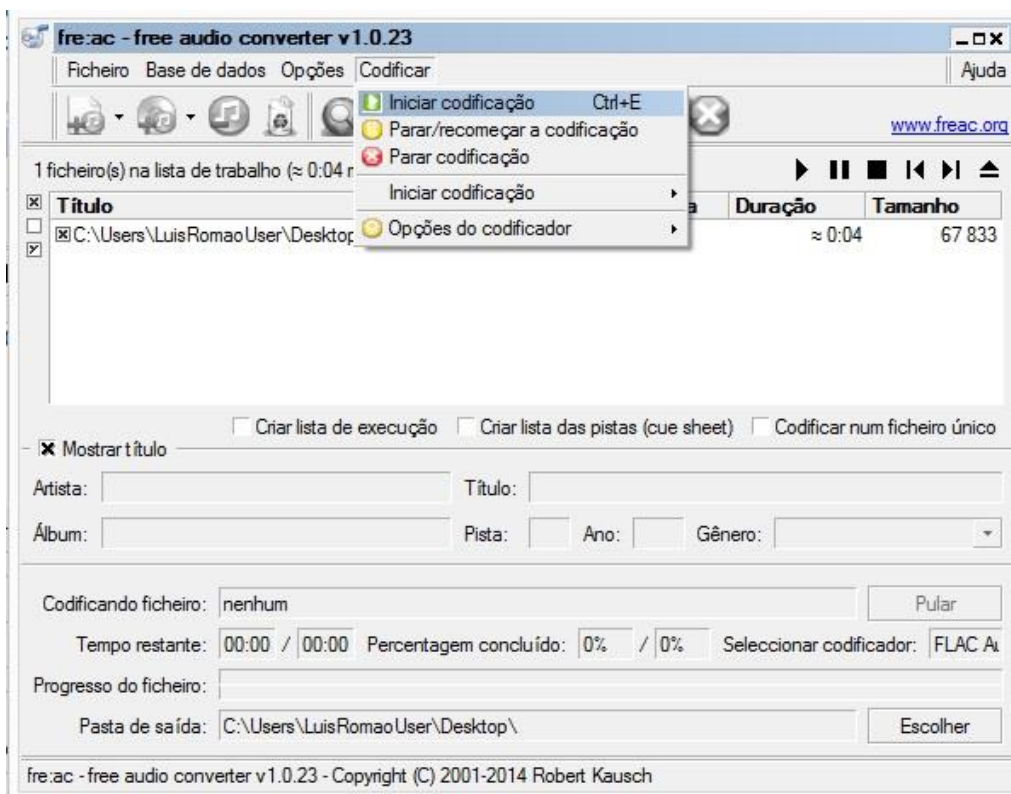


Imagem 21 – Iniciar a conversão de wma para flac.

Os ficheiros de áudio convertidos estarão na pasta de saída (ver Imagem 21).

3. Comunicar com o Google Speech API

3.1. Instalar local webservice, tais como, XAMPP ou EasyPHP. Usei EasyPHP DevServer 14.1 VC11. Disponível: <http://www.easyphp.org/easyphp-devserver.php>

3.1.1. Ir a “C:\o-seu-caminho\EasyPHP-DevServer-14.1VC11\data\localweb” e criar nova pasta, exemplo, FLAC_GoogleSpeech

3.2. Instalar Python. Usei “Python 2.7.10rc1 - 2015-05-11” versão “Windows x86-64 MSI installer”. Disponível: <https://www.python.org/downloads/windows/>

3.2.1. Instalar pasta Requests.

3.2.1.1. Download “requests-master” (carregar em download ZIP).

Disponível: <https://github.com/kennethreitz/requests>

3.2.1.2. Extrair e abrir a pasta “requests-master” e copiar pasta “requests”

3.2.1.3. Colar pasta “requests” na pasta “Lib”. A pasta “Lib” encontra-se em “C:\o-seu-caminho\Python27\Lib”

3.2.2. Copiar “python.exe” para FLAC_GoogleSpeech

3.3. Download “Python Google Speech to Text API implementation” de Travis Payton. Disponível: <http://codeabito.com/2014/09/python-google-speech-api/#comment-223>

3.3.1. Extrair “gsst.rar”. Abrir “gsst.py” e introduzir a “server key” (**ver Imagem 10**) na variável “self.api_key” (encontra-se aproximadamente na linha 86).

3.3.2. Testar Script de Travis Payton.

3.3.2.1. Copiar “gsst.py” e “test.flac” para FLAC_GoogleSpeech

3.3.2.2. Criar script PHP (“runPython.php”) em FLAC_GoogleSpeech para correr o script de Travis Payton. Ver Algoritmo 1 de runPython.php

```

<!DOCTYPE html>
<html lang="pt-PT">
<head>
    <meta charset="utf8">
    <title>Run Script Travis Payton</title>
</head>
<body>
    <?php
        echo "#####<br>";
        echo "Results from Python Code<br><br>";
        $cmd = "python.exe gstt.py";
        system($cmd);
        echo "#####<br>";
    ?>
</body>
</html>

```

Algoritmo 1 – Algoritmo de runPython.php

3.3.2.3. Iniciar o EasyPHP e ligar a Internet.

3.3.2.3.1. Instalar “Visual C++ Redistributable for Visual Studio 2012 Update 4”.

Por ventura poderá ocorrer um erro aquando da ligação do EasyPHP. Se isso ocorrer este passo 3.3.2.3.1, terá que ser realizado.

3.3.2.3.1.1. Download do “Visual C++ Redistributable for Visual Studio 2012 Update 4” versão x86. Usei “VSU_4\vc redistrib_x86.exe”. Disponível: <https://www.microsoft.com/en-us/download/details.aspx?id=30679>

NOTA: mesmo que o seu computador seja x64, o EasyPHP só funcionará aquando da instalação da versão x86 do “Visual C++ Redistributable for Visual Studio 2012 Update 4”.

3.3.2.4. Abrir um novo separador no Google Chrome e escrever: 127.0.0.1

3.3.2.5. "Enter", seleccionar " FLAC_GoogleSpeech " e depois seleccionar "runPython.php". Ver resultado na Imagem 22.

```
#####  
Results from Python Code  
  
Filename: test.flac  
Duration filename: 5.016 seconds  
Bitrate filename: 44100  
Result Google: [{"result":[{"alternative":[{"transcript":"testing Google's transcription service 1234"}],"final":true}],"result_index":0}]  
Result Google duration transcript: 6.756 seconds  
#####
```

Imagem 22 – Resultado final na Web – Idioma Inglês

3.4. Interligação dos ficheiros flac e script de Travis Payton modificado em FLAC_GoogleSpeech (pasta criada no EasyPHP).

3.4.1. Colar os ficheiros flac em FLAC_GoogleSpeech

3.4.2. Script de Travis Payton modificado. O script de Travis Payton encontra-se bem implementado para ficheiros flac no idioma em inglês. Para o idioma Português de Portugal é necessário fazer umas alterações ao algoritmo desenvolvido por Travis Payton (“gsst.py”). Ver Algoritmo 2.

```
1 # -*- coding: utf-8 -*-
2 __author__ = 'Travis Payton'
3 # gsst.py
4 # Travis Payton
5 # June 10th, 2014
6 # Updated July 13, 2014
7 # Version 1.1
8 # Tested with Python 2.7.8

(...)

84 self.upstream_url = https://www.google.com/speech-api/full-duplex/v1/up?key=%\(key\)s&pair=%\(pair\)s&lang=pt-PT&client=chromium&continuous&interim&pFilter=0

(...)

219 input = "test.flac"

221 if input[-5:] != '.flac':
222     input += '*.flac'
223 files = glob.glob(input)
224 for flac in files:
225     print "Filename: %s<br>" %flac
226     valid_result = False
227     tries = 0
228     with open(flac, 'rb') as f:
229         objGoogleSpeech = GoogleSpeechAPI(f)
230         print "Duration filename: %.03f seconds<br>" %objGoogleSpeech.length
231         f.seek(0)
232         with Timer() as t:
233             objGoogleSpeech.start()
234         print "Bitrate filename: %s" %objGoogleSpeech.sampleRate + "<br>"

236     # clean up results into JSON String
237     JSONResult = '['
238     for x in objGoogleSpeech.result:
239         JSONResult += x + ', '
240     JSONResult = JSONResult[:-3] + ']'
241     #print JSONResult

243     print "Result Google: %s" %JSONResult + "<br>"
244     print "Result Google duration transcript: %.03f seconds" %t.interval + "<br>"
245 if __name__ == "__main__":
246     main(sys.argv[1:])
```

Algoritmo 2 – Script Travis Payton modificado (“gsst.py”).

Analisando o Algoritmo 2, o script de Travis Payton modificado (“gsst.py”) arredondará as 246 linhas de código. **Alterações:**

- Adicionar a linha 1
- Da linha 2 a 83 não existe nenhuma alteração a fazer
- Linha 84 modificar “lang=en-US” para “lang=pt-PT”
- Da linha 85 a 219 não existe nenhuma alteração a fazer
- A partir da linha 220 eliminar todo o conteúdo original e adicionar as linhas 221 a 246

Por fim guardar as alterações efetuadas.

3.4.3. Analisar os ficheiros flac.

3.4.3.1. Abrir o ficheiro “gsst.py” e escrever na linha 219 (ver Algoritmo 2) o nome do ficheiro flac presente em FLAC_GoogleSpeech que quer ser analisado, exemplo, frase7.flac. Guardar “gsst.py”

3.4.3.2. Iniciar o EasyPHP e ligar a Internet.

3.4.3.3. Abrir um novo separador no Google Chrome e escrever: 127.0.0.1

3.4.3.4. "Enter", seleccionar " FLAC_GoogleSpeech " e depois seleccionar “runPython.php”. Ver resultado na Imagem 23.

```
#####  
Results from Python Code  
  
Filename: frase7.flac  
Duration filename: 3.297 seconds  
Bitrate filename: 44100  
Result Google: [{"result":[{"alternative":[{"transcript":"o meu carro é amarelo","confidence":0.71846169}],"final":true}],"result_index":0}]  
Result Google duration transcript: 4.355 seconds  
#####
```

Imagem 23 Resultado final na Web – Idioma Português de Portugal

FIM!!

Autor deste tutorial: Luís Romão
Email:gaspar.luisromao@gmail.com
Ano:2015

A.2 Tutorial Tradutor LP para LGP

Este tutorial descreve como instalar o sistema desenvolvido desta dissertação num computador visto que o sistema não se encontra na Internet.

Tutorial: Instalação e Utilização do “Tradutor LP para LGP” num computador

Como o “Tradutor LP para LGP” não se encontra alojado num servidor Web não se pode utilizar diretamente. Estando alojado num computador é **necessário** efetuar os passos em baixo.

1. Instalar local webserver, tais como, XAMPP ou EasyPHP. Usei EasyPHP DevServer 14.1 VC11. Disponível:
<http://www.easyphp.org/easyphp-devserver.php>

- 1.1. Ir a “C:\o-seu-caminho\EasyPHP-DevServer-14.1VC11\data\localweb” e colar “Tradutor LP para LGP”

- 1.2. Iniciar o EasyPHP e ligar a Internet.

- 1.2.1. Instalar “Visual C++ Redistributable for Visual Studio 2012 Update 4”.

Por ventura poderá ocorrer um erro aquando da ligação do EasyPHP. Se isso ocorrer este passo 1.2.1, terá que ser realizado.

- 1.2.1.1. Download do “Visual C++ Redistributable for Visual Studio 2012 Update 4” versão x86. Usei “VSU_4\vc redistrib_x86.exe”. Disponível:
<https://www.microsoft.com/en-us/download/details.aspx?id=30679>

NOTA: mesmo que o seu computador seja x64, o EasyPHP só funcionará aquando da instalação da versão x86 do “Visual C++ Redistributable for Visual Studio 2012 Update 4”.

- 1.3. Abrir um novo separador no Google Chrome e escrever: 127.0.0.1

2. Importar bases de dados para “Tradutor LP para LGP”. Ver Imagem 1 a Imagem 5.



Imagem 1 – Carregar no botão “Casa” para aceder às configurações de administrador.

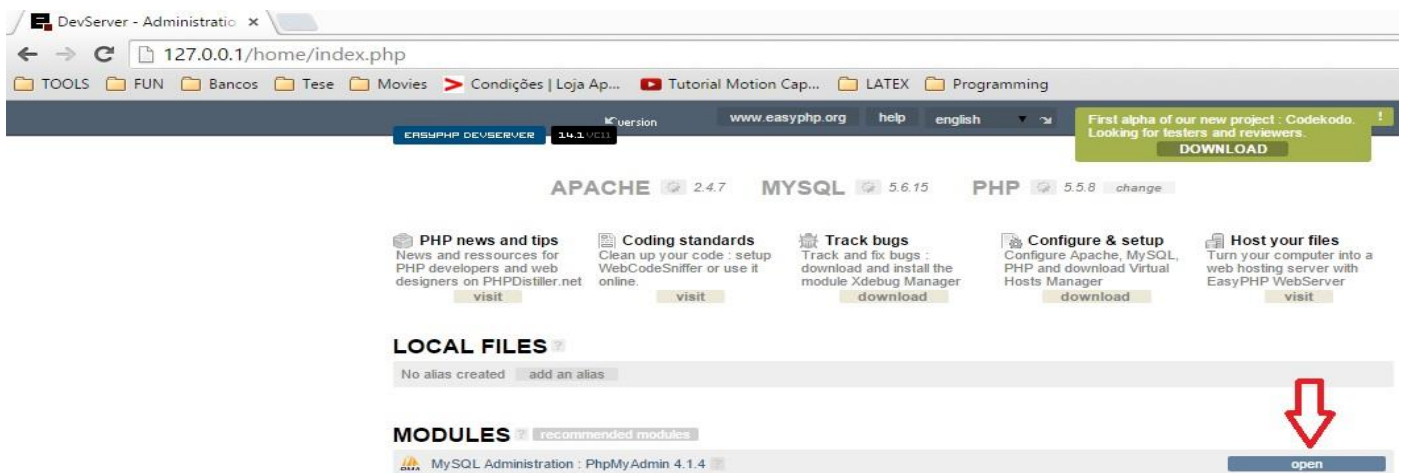


Imagem 2 – Carregar em “Open” para aceder às configurações da Base de Dados (BD)

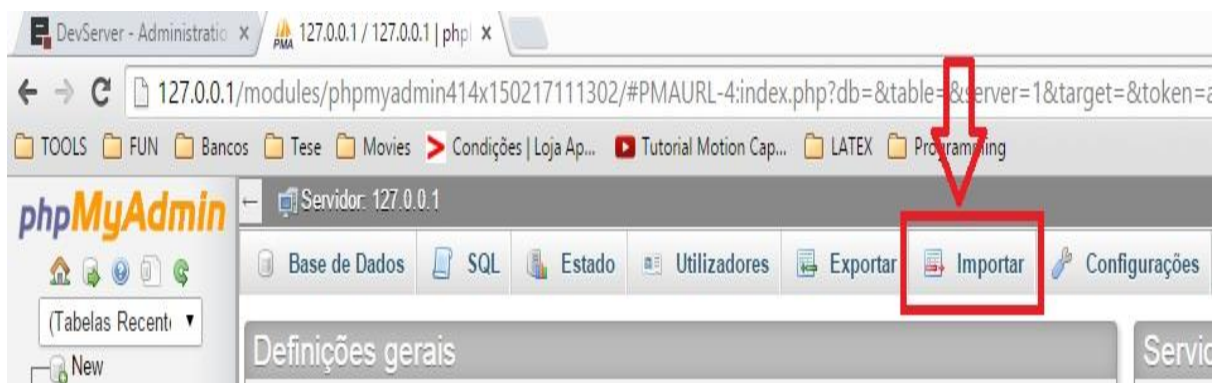


Imagem 3 – Carregar em “Importar” para importar as BD para “Tradutor LP para LGP”



Imagem 4 – Carregar em “Escolher ficheiro” para importar a BD que se encontram no seu computador. BD a importar: “modulo2_database.sql” e “modulo3_database.sql”, sequencialmente

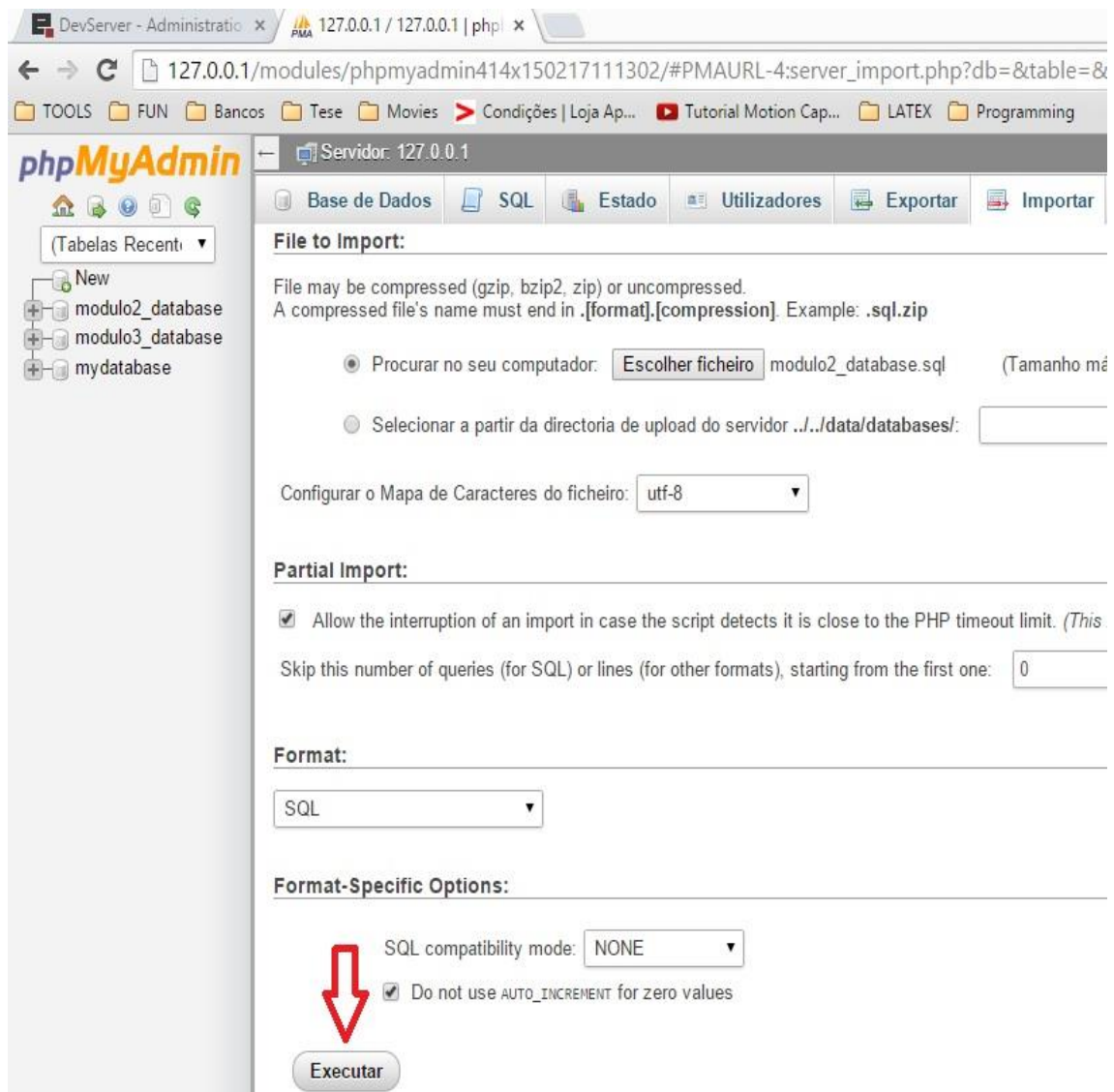


Imagem 5 – Carregar “Executar” para importar a BD escolhida.

Assim, importar BD “modulo3_database.sql” apenas será necessário repetir as Imagem 4 e Imagem 5.

3. Criar novo utilizador na Base de Dados. Ver imagens Imagem 6 a Imagem 9.



Imagem 6 - Carregar em “Utilizadores” criar um utilizador para “Tradutor LP para LGP”

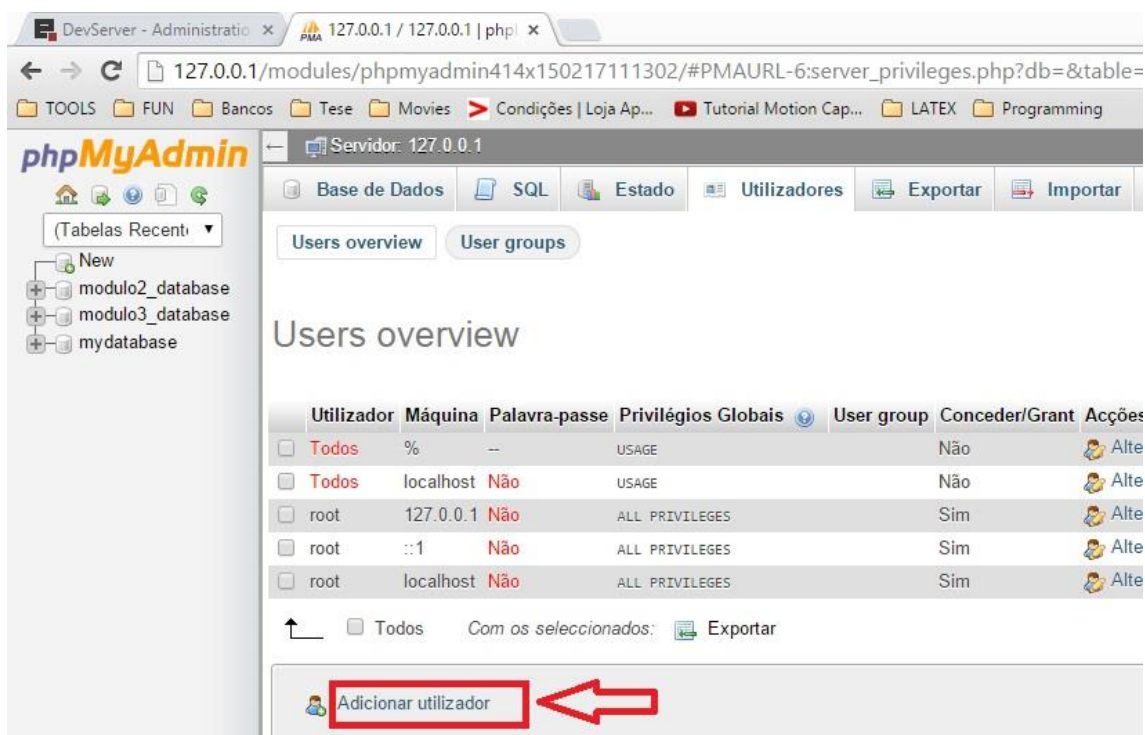


Imagem 7 – Carregar “Adicionar utilizador” para criar utilizador

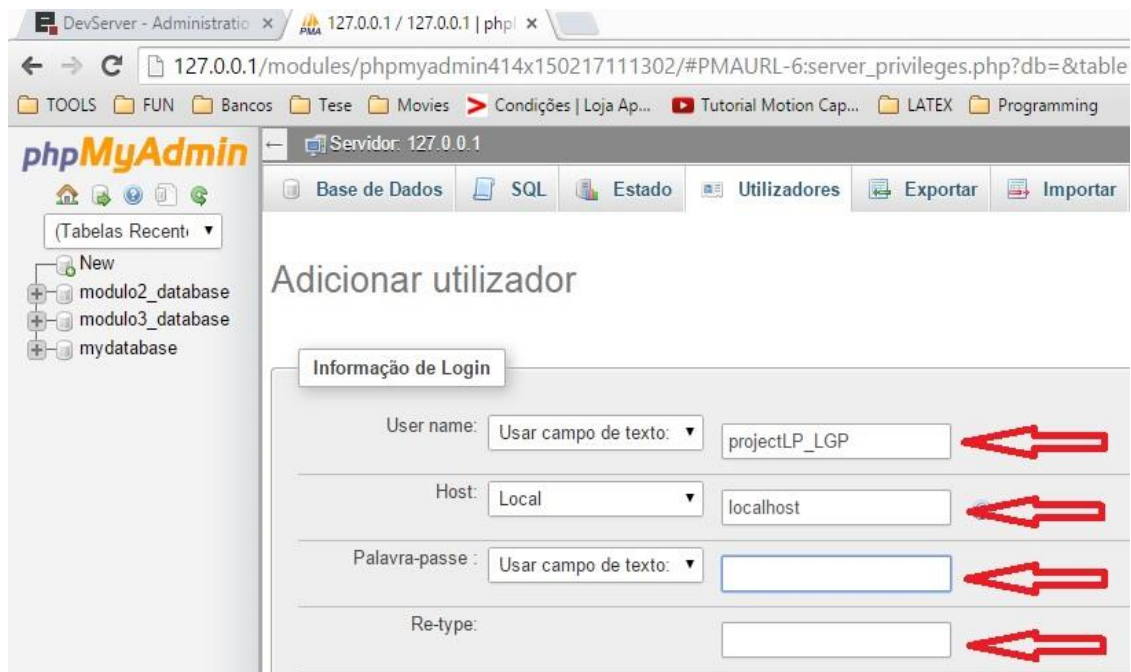


Imagem 8 – User name: projectLP_LGP; Host: local; Palavra-passe: password; Re-type: password

As BD “modulo2_database.sql” e “modulo3_database.sql” foram definidas como utilização:

- User name: projectLP_LGP
- Host: localhost
- Palavra-passe: password

A **não utilização** das configurações anteriores (ver Imagem 8) as BD **não funcionam** no “Tradutor LP para LGP”.

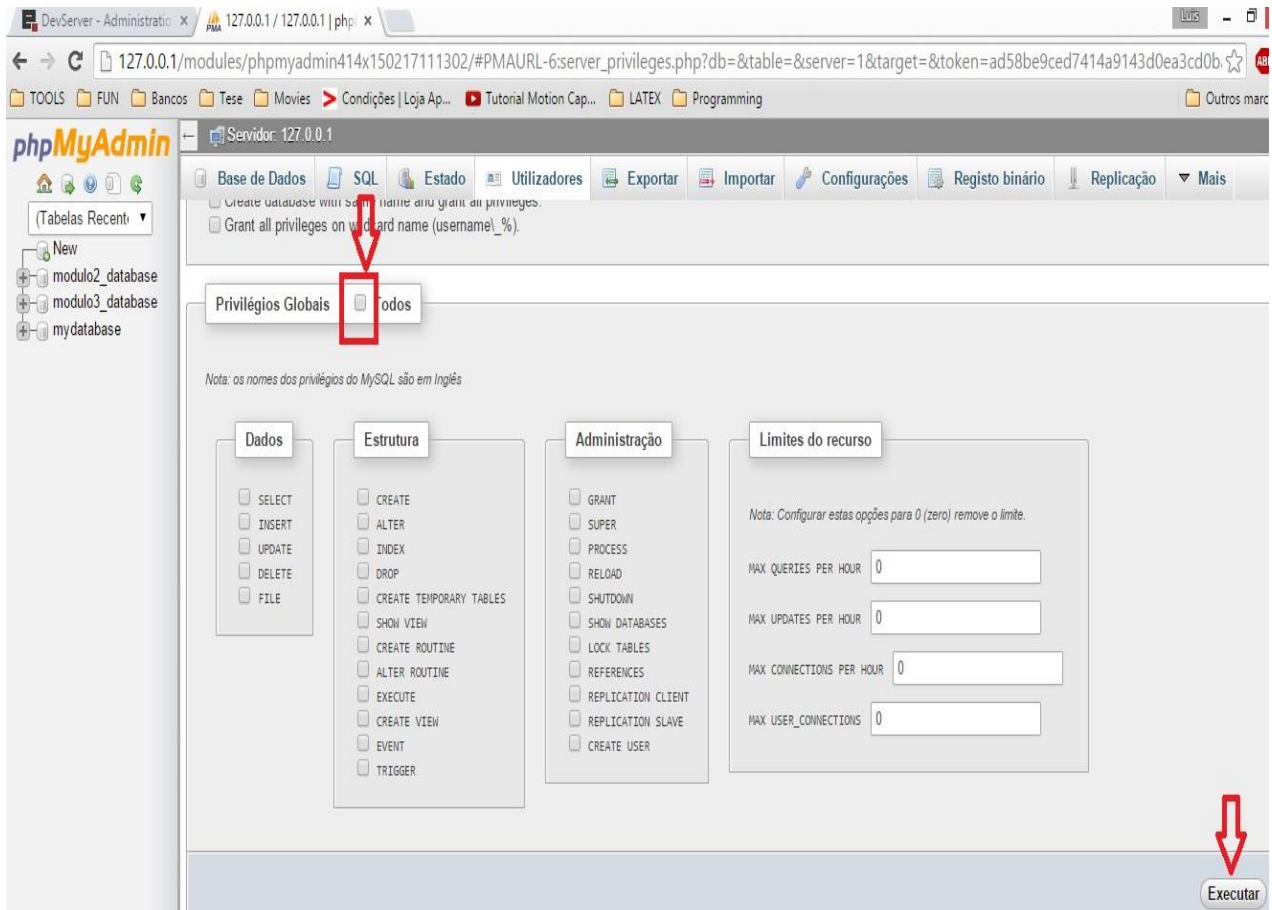


Imagem 9 – Concluir configurações do utilizador. Carregar “Todos” e depois “Executar”, criando assim um novo utilizador para “Tradutor LP para LGP”

4. Correr “Tradutor LP para LGP”

4.1. Iniciar o EasyPHP e ligar a Internet.

4.2. Abrir um novo separador no Google Chrome e escrever: 127.0.0.1

4.3. "Enter" e seleccionar “Tradutor LP para LGP”

Nas próximas utilizações do “Tradutor LP para LGP” apenas será necessário realizar o passo 4.

FIM!!

Autor deste tutorial: Luís Romão
Email: gaspar.luisromao@gmail.com
Ano: 2015

Anexo B

Questionário de Avaliação ao sistema

Este é o questionário de avaliação que as pessoas preencheram acerca do funcionamento do sistema desta dissertação.

Questionário de Avaliação do "Tradutor de LP para LGP"

Este questionário pretende saber a sua opinião sobre o "Tradutor de LP para LGP".

O "Tradutor de LP para LGP" tem como objectivo fazer a conversão da fala de Língua Portuguesa (LP) para Língua Gestual Portuguesa (LGP) através de vídeos com um gestuante humano.

Todas as suas respostas são confidenciais.
Obrigado pela sua disponibilidade.

* Required

Idade. *

Sexo. *

- Feminino
 Masculino

Profissão. *

Conhece a LGP de igual forma como a Língua Portuguesa? *

- Não
 Sim

Qual a sua área de conhecimento de acordo com a sua profissão? *

- Áudio
 Vídeo
 Informática
 Língua Portuguesa
 Língua Gestual Portuguesa
 Outra

A aplicação é útil? *

1 2 3 4 5

Pouco Útil Muito Útil

Como classifica a experiência/facilidade de usar a aplicação? *

1 2 3 4 5

Muito Má Muito Boa

A aplicação tem visual atrativo? *

1 2 3 4 5

Nada Atrativo Muito Atrativo

Avaliação do Reconhedor de Fala. *

(Todas as palavras proferidas foram reconhecidas?)

1 2 3 4 5 6 7 8 9 10

Nenhuma Todas

Avaliação da conversão para GLOSA. *

(GLOSA: sequência de gestos a serem visualizados no vídeo. Todas a palavras reconhecidas foram convertidas para GLOSA corretamente?)

1 2 3 4 5 6 7 8 9 10

Nenhuma Todas

Avaliação dos vídeos em LGP: tradução para LGP. *

(Todas as palavras em GLOSA foram visualizadas?)

1 2 3 4 5 6 7 8 9 10

Nenhuma Todas

Avaliação dos vídeos em LGP: tamanho da janela de vídeo. *

(O tamanho da janela de vídeo é adequado?)

- Pequeno
- Grande
- Bom

Avaliação dos vídeos em LGP: qualidade visual. *

(Como classifica a qualidade visual do vídeo?)

1 2 3 4 5 6 7 8 9 10

Muito Má Muito Boa

Comentários Finais... *

(O que gosta? O que não gostou? O poderia ser melhorado?)