ESCOLA POLITÉCNICA
DE PERNAMBUCO

UPE - POLI
Engenharia de Computação
GRADUAÇÃO

Trabalho de Conclusão de Curso

Engenharia de Computação

UMA NOVA ABORDAGEM PARA GERAÇÃO DE LEGENDAS ACESSÍVEIS A DEFICIENTES AUDITIVOS, CAPAZ DE CODIFICAR VISUALMENTE EMOÇÕES

Autor: Marlon Chalegre de Paula

Orientador: Prof. Dr. Fernando Buarque de Lima

Recife, Dezembro de 2010

UMA NOVA ABORDAGEM PARA GERAÇÃO DE LEGENDAS ACESSÍVEIS A DEFICIENTES AUDITIVOS, CAPAZ DE CODIFICAR VISUALMENTE EMOÇÕES

Dedico este trabalho ao meu irmão Michel Chalegre. Sem ajuda dele eu não estaria realizando mais este sonho.

Agradecimentos

Agradeço a minha família por todo o apoio que tive em minha vida e graças a esse apoio hoje posso concluir o curso de Engenharia de Computação na Universidade de Pernambuco.

Agradeço também a todos os professores com quem tive contato durante esses cinco anos de curso, tendo paciência e não apenas passando conhecimento técnico como realmente instruindo a nós alunos a como proceder na vida. Agradeço em especial o professor Fernando Buarque que me deu a oportunidade de fazer este trabalho.

E por fim, um agradecimento especial a minha noiva que ficou ao meu lado durante todos essas anos.

Resumo

As legendas de televisão e cinema de puro texto, como são geradas atualmente, por si só não são capazes de levar ao deficiente auditivo toda a informação que esta transcorrendo no diálogo exibido em cada cena. Isso reduz a sua capacidade de ter um entendimento melhor sobre o que esta se passando no programa ou filme que ele esta assistido.

Este trabalho propõem um novo modelo de legendas mais acessíveis que possa levar ao deficiente o conteúdo emocional de dialogo tornando a percepção do que esta acontecendo mais simples.

Utilizando *Hidden Markov Models,* uma técnica inteligente de classificação, e um processo automatizado para extração de características foi possível desenvolver o sistema capaz de identificar emoções em um dialogo e codifica-las na legenda do programa.

Como resultado deste trabalho temos uma nova forma de se assistir a programas legendados, capaz de ampliar a capacidade do deficiente auditivo de entender o que está sendo discutido no dialogo apresentado.

Abstract

The subtitles of pure text, in movies and television, such as they are currently made, by themselves are not capable of conveying to hearing impaired persons all information that is being produced out of the dialogues displayed. This reduces the capacity of the hearing impaired to have a better understanding about what is happening in the program or movie currently watches.

This monograph presents a new model that affords subtitles to be conveyed to the hearing impaired the emotional content of dialogues in order for them to make sense of what is happening in an easier manner.

The system is able to identify some basic emotions in a speech and encode them in the caption of the movie. For that we used Hidden Markov Models and automated feature extraction that has produced improvements in the understanding capacity of hearing impaired persons of on-line dialogues.

Sumário

Índice de Figuras

Tabela de Símbolos e Siglas

As siglas aparecem em ordem alfabética.

ABNT – Associação Brasileira de Normas Técnicas

API - Application Programming Interface

ASS - Advanced SubStation Alpha

DCT – Discrete Cosine Transform

GISTAL – Grup d'Investigació en Sordeses i Transtorns del Llenguatge

HMM – Hidden Markov Model

MFCC – Mel-Frequency Cepstrum Coefficients

UAB - Universitat Autónoma de Barcelona

WAV - Waveform Audio File Format

XML - Extensible Markup Language

Capítulo 1

Introdução

1.1 *Caracterização do Problema*

Em todo o mundo, 278 milhões de pessoas têm uma estimativa de perda auditiva moderada a profunda em ambas os ouvidos. Menos de um em cada 40 pessoas que precisam de próteses auditivas as possui. Oitenta por cento das pessoas com deficiência auditiva vivem em países de baixa e média renda, e um quarto de deficiência auditiva começa na infância [1].

A deficiência auditiva é a incapacidade parcial ou total de audição. Esta deficiência trás consigo sérios problemas sociais. Em crianças pode retardar o desenvolvimento da linguagem e habilidades cognitivas, que podem impedir o progresso na escola. Em adultos, a deficiência auditiva muitas vezes torna difícil de obter, executar e manter empregos.

> “São várias as causas que levam à deficiência auditiva. A deficiência auditiva condutiva, por exemplo, tem como um dos fatores o acúmulo de cera no canal auditivo externo, gerando perda na audição. Outra causa são as otites. Quando uma pessoa tem uma infecção no ouvido médio, essa parte do ouvido pode perder ou diminuir sua capacidade de "conduzir" o som até o ouvido interno. No caso da deficiência neurossensorial, há vários fatores que a causam, sendo um deles o genético. Algumas doenças, como rubéola, varíola ou toxoplasmose, e medicamentos tomados pela mãe durante a gravidez podem causar rebaixamento auditivo no bebê. Também a incompatibilidade de sangue entre mãe e bebê (fator RH) pode fazer com que a criança nasça com

problemas auditivos. Uma criança ou adulto com meningite, sarampo ou caxumba também pode ter como sequela a deficiência auditiva. Infecções nos ouvidos, especialmente as repetidas e prolongadas e a exposição frequente a barulho muito alto também podem causar deficiência auditiva". [1]

Portadores de deficiência auditiva encontram dificuldades para assistir a filmes ou programas televisivos. Pesquisadores do Centro de Investigação em deficiência auditiva e Aquisição da Linguagem da *Universitat Autónoma de Barcelona* (UAB) conhecidos como GISTAL[2], estudaram o nível de compreensão de programas de televisão legendados, por grupos de alunos que têm deficiência auditiva severa ou profunda. Os resultados demonstram que as crianças e adolescentes surdos têm dificuldades em seguir legendas e imagens em conjunto, devido à velocidade com que as legendas aparecem e a transcrição literal dos diálogos [2]. O resultado desta pesquisa mostra claramente a dificuldade que os deficientes auditivos têm de se integrar na sociedade.

1.2 *Motivação*

Além de outras mídias de massa, a televisão hoje é uma das principais fontes de informação e entretenimento da população brasileira. Na seção anterior foi descrito a dificuldade que os deficientes auditivos possuem para obter um entendimento global do que esta sendo televisionado apenas utilizando as legendas. Neste cenário é possível notar que esta fatia da população acaba não tendo o acesso à informação transmitida através deste meio de comunicação de maneira adequada a sua deficiência.

O conteúdo emotivo de uma cena pode ser identificado por um humano utilizando a visão (análise da face) e principalmente, utilizado a audição. No caso de um

[1] Retirado do sítio internet http://www.crfaster.com.br/auditiv.htm, em 16/11/2010.

[2] Grup d'Investigació en Sordeses i Transtorns del Llenguatge

deficiente auditivo esta parte importante da informação contida na cena é perdida ao assistir um programa televisivo legendado, já que uma analise facial dos falantes não pode ser realizada enquanto se está lendo as legendas.

É através do conteúdo emotivo que é possível distinguir uma ironia de uma afirmação ou facilmente identificar se existe ou não uma agressividade no que é dito. Esta informação é completamente perdida no momento em que ocorre a transcrição literal dos diálogos nas legendas.

A inteligência computacional, uma sub-área importante da ciência da computação, vem se popularizando rapidamente pois pode, por exemplo, utilizar várias técnicas capazes de extrair características e classificar informações de forma eficaz e eficiente. Estas técnicas podem ser aplicadas a uma vasta classe de problemas complexos dando à máquina o poder de reconhecer padrões simbólicos. O poder das técnicas inteligentes para classificação pode ser visto no trabalho de T. Nwe [3] onde é criado um classificador inteligente utilizando Hidden Markov Models (Modelos Escondidos de Markov) capaz de identificar emoções baseado na fala.

1.3 *Objetivos e Metas*

O objetivo deste projeto é extrair o conteúdo emocional dos diálogos utilizando algoritmos inteligentes e criando um novo modelo de legendas animadas pré-gravadas que sejam capazes de passar o máximo possível de informação para os deficientes auditivos, acarretando em uma melhora substancial do programa que esta sendo visto e trazendo um maior conforto nas horas de lazer.

Para isto este projeto tem como meta estudar o modelo atual de criação de legendas, os métodos de extração de características utilizados neste tipo de problema e desenvolver um software que seja capaz de criar uma nova legenda que codifique as emoções em cores através do áudio de um programa no formato WAV (Waveform Audio File Format) e sua legenda em formato SRT (SubRip File Format).

1.4 *Resultados Esperados*

Espera-se que ao utilizar implementações piloto deste trabalho o deficiente auditivo possa assistir a um programa com legenda pré-gravada em formato que facilite a compreensão do que está sendo falado melhorando o entendimento global sobre o programa ou filme, reduzindo as barreiras de comunicação e trazendo um melhor conforto ao deficiente auditivo.

1.5 *Organização do Documento*

O documento está dividido em cinco capítulos, resumidos a seguir:

Capitulo 1: Introdução

Contém o texto introdutório sobre o trabalho, caracterizando o problema, abordando a motivação para resolvê-lo e apresentando os objetivos, metas e resultados esperados.

Capítulo 2: As legendas e os Deficientes auditivos

Neste capítulo está descrito como se dá o processo de criação de legendas acessíveis a deficientes auditivos, discute a experiência do usuário, mostra trabalhos relacionados e as técnicas que serão utilizadas para criação do sistema proposto.

Capítulo 3: O Sistema Proposto

Este capítulo descreve o funcionamento do sistema, mostrando seus diagramas de blocos, diagramas de classes e descreve a arquitetura do sistema.

Capítulo 4: Resultados Obtidos

Aqui está descrito quais foram os dados utilizados para realização dos testes e mostra também os resultados obtidos nesses testes.

Capítulo 5: Conclusão e Trabalhos Futuros

Nesse último capítulo são comentados os resultados obtidos como também formuladas as conclusões globais acerca destes resultados. Logo em seguida são feitas as considerações finais e listados os possíveis trabalhos futuros.

Capítulo 2

As legendas e os Deficientes auditivos

Neste capítulo é descrito como se dá o processo de criação de legendas acessíveis a deficientes auditivos, discute a experiência do usuário e mostra trabalhos relacionados.

2.1 *Acessibilidade em comunicação na televisão*

A Associação Brasileira de Normas Técnicas (ABNT) criou a norma 15.290/2005 que estabelece diretrizes gerais a serem observadas para acessibilidade em comunicação na televisão, consideradas as diversas condições de percepção e cognição, com ou sem a ajuda de sistema assistivo ou outro que complemente necessidades individuais [4].

O texto define todas as características necessárias para que uma legenda, seja ela ao vivo ou pré-gravada, possa ser considerada acessível. Abaixo estão descritos algumas partes da norma referentes a legendas pré-gravadas:

- Efeitossonoros: Devem ser transcritos e indicados entre colchetes todos os sons não literais, importantes para a compreensão do texto. Por exemplo: [Latidos], [Criança chorando], [Trovoadas], [Porta rangendo] etc. Sons importantes devem ser descritos. (portas abrindo, telefone tocando, gritos, etc).

- Fala e ruídos: Quando houver informações simultâneas de fala e sons não literais, a fala deve estar posicionada próxima ao falante e o som não literal deve vir informado entre colchetes ([ ]).

- Identificação dos falantes: Quando a situação cênica não permite a identificação sobre quem está falando, ou o personagem está fora de cena (em off), o nome

do personagem ou algum tipo de informação que o identifique deve ser informado entre colchetes. Ex.: [João]; [Menino]; [Policial] etc.

- Itálico: Deve ser usado o itálico para indicar falas fora de cena (em off), narração, enfatizar entonação e para palavras em outra língua.

- Música: O símbolo da nota musical deve ser usado para diferenciar a música da palavra falada:
  - a informação sobre a música (se é fundo musical, rock, música romântica ou de suspense, se é cantada etc.) deve vir entre notas musicais;
  - no caso de transcrição da letra da música, duas notas musicais seguidas, ao final da transcrição, devem indicar seu termino;
  - sempre que possível, a letra da música deve ser transcrita.

- Onomatopeias: O uso da informação literal do som (latidos) deve ter preferencia em relação ao uso da onomatopeia (au-au). Programas e filmes infantis ou cômicos podem fazer uso de onomatopeias.

As legendas ao vivo não necessitam possuir todas as estas características, o motivo para que isso ocorra é a necessidade de que ela seja produzida em tempo hábil evitando o máximo possível, atrasos no que esta sendo dito e no que esta sendo descrito na legenda.

Adicionar todas essas informações textuais à legenda pré-gravada torna o texto longo, reduzindo o tempo útil gasto na leitura da legenda e retirando o foco do deficiente auditivo para o conteúdo informativo da cena.

2.2 *O processo de criação de legendas*

As legendas podem ser de dois tipos ao vivo ou pré-gravada. As duas formas mais comuns de produção de legendas ao vivo nas transmissões da TV brasileira são por (i) estenotipia informatizada e o de (ii) reconhecimento de voz.

Na estenotipia existe a figura humana treinada para digitar rapidamente equipada de um teclado especial que representa letras e grupos de fonemas. O estenotipista registra o que ouve no mesmo momento em que o telespectador assiste ao programa

em seu televisor. Neste método alguns erros de digitação ocorrem, além do atraso natural entre a fala, digitação e exibição do resultado na TV [5] [10].

Utilizando um sistema de reconhecimento de voz o computador torna-se capaz de interpretar vozes e produzir o texto das legendas. As vozes são transmitidas a um profissional em uma sala isolada sendo esta pessoa responsável por repetir pausadamente o que houve ao computador. O *software* não é totalmente preciso, podendo confundir alguns fonemas como "lhe" e "lie". Além do trabalho vocal, o profissional responsável por ditar acrescenta via teclado informações sobre outros sons do ambiente para depois liberar a transmissão da legenda. A principal dificuldade deste método é quando há momentos de sobreposição de falas, tumultos e entrevistas, quando então o primeiro método ainda se mostra bem mais eficiente [4][5].

No caso das legendas pré-gravadas o método utilizado é conhecido como *transcript*. O *transcript* é a passagem do áudio para texto feito por um profissional de digitação. Após a transcrição ocorrer, a legenda passa por uma revisão ortográfica e são adicionados todos os elementos referentes à acessibilidade [5]. O sistema de legendas pré-gravadas permite incluir a transcrição de sons não literais e recursos, como diferentes posicionamentos da legenda, informações sobre o falante e informações sobre personagens em *off* (fora da cena) [4].

Geni Aparecida Fávero, coordenadora do grupo de trabalho que criou a norma 15.290/2005, da Associação Brasileira de Normas Técnicas (ABNT), cujo texto oferece diretrizes de acessibilidade em comunicação na tevê diz, que "No sistema ao vivo o texto da legenda deve ter no mínimo 98% de acerto e no pré-gravado, 100%. Portanto, num programa com legenda pré-gravada não devem ocorrer erros de grafias ou omissão de palavras como acontece" [4][5].

2.3 *A experiência do deficiente auditivo*

A pesquisa realizada pelo GISTAL mostrou que o entendimento obtido pelos deficientes auditivos ao ler uma legenda esta diretamente ligado à idade e ao tipo de programa que está sendo assistido. De acordo com os pesquisadores, jovens com deficiência auditiva têm dificuldades para entenderem do que se trata o programa que

estão assistindo. Nos adultos esta dificuldade praticamente não existe, acostumados a ler textos rapidamente os adultos não sofrem tanto quanto os jovens.

Dois outros estudos foram realizados pelos pesquisadores do GISTAL com participantes mais jovens: um deles consistiu em piloto com sete crianças com idades entre 6 e 7, enquanto o outro era formado por 16 crianças de 7 a 10. Ambos os grupos viram um fragmento do desenho animado "Shin-Chan", mas ao segundo grupo foi mostrado o desenho com legendas criadas pelo próprios professores (com nova velocidade e critérios de seleção de texto). No primeiro grupo, apenas 2% dos participantes entenderam do que a animação se tratava. No segundo grupo, a compreensão global do fragmento atingiu 65,5%.

Para C. Cambra [2], não existe a necessidade de legendar tudo o que é dito pelos personagens, deixando algumas partes chaves da estória sem legenda. Por exemplo, personagens expressando um estado emocional, fazendo com que o deficiente auditivo possa contemplar a imagem e através dela obter o significado da cena.

No trabalho de D. Lee [7] é reforçada a necessidade para as legendas se tornarem muito mais do que são atualmente, puro texto. Em especial considerando o crescimento e popularização da TV Digital onde nos é permitido criar softwares específicos para melhorar a maneira e qualidade das legendas. Além disso, telespectadores estão começando a exigir mais e melhor qualidade na legenda que é produzida. Neste trabalho ainda é exposto uma abordagem para exibição de emoção nas legendas através de cores e ações, porém esse trabalho é feito manualmente.

Existem outros trabalhos sobre legendas na TV com o foco em analisar as características sobre um ponto de vista formal, como as preferências dos usuários em relação cor, tamanho, número de caracteres, local da legenda na tela [8][9], ou no estilo linguístico utilizado nas legendas, propondo procedimentos alternativos [6].

2.4 *Extração de características de sinais de áudio*

Extração de características é um processo de extrair parâmetros especiais que sejam capazes de identificar um objeto. Essas informações podem ser utilizadas posteriormente, por exemplo, para identificação de padrões. Neste trabalho é utilizado o *Mel-Frequency Cepstrum Coefficients* (MFCC) como as características que devem ser extraídas.

O MFCC é uma representação do espectro de energia de curto prazo de um som. Estes coeficientes são derivados de um tipo de representação cepstral [3] de um sinal de áudio. A escala mel[4] é utilizada para posicionar as bandas de frequência de maneira que fiquem igualmente separadas, o que aproxima o modelo do comportamento do sistema auditivo humano, uma vez que a percepção das frequências dos sons por seres humanos é dita não-linear [11]. A figura 1 mostra um gráfico que relaciona e escala mel com a frequência.

[3] O cepstro é o resultado da aplicação da transformada de Fourier em um sinal, também chamado de espectro de um espectro.

[4] Escala que representa a frequência fundamental percebida de um som.

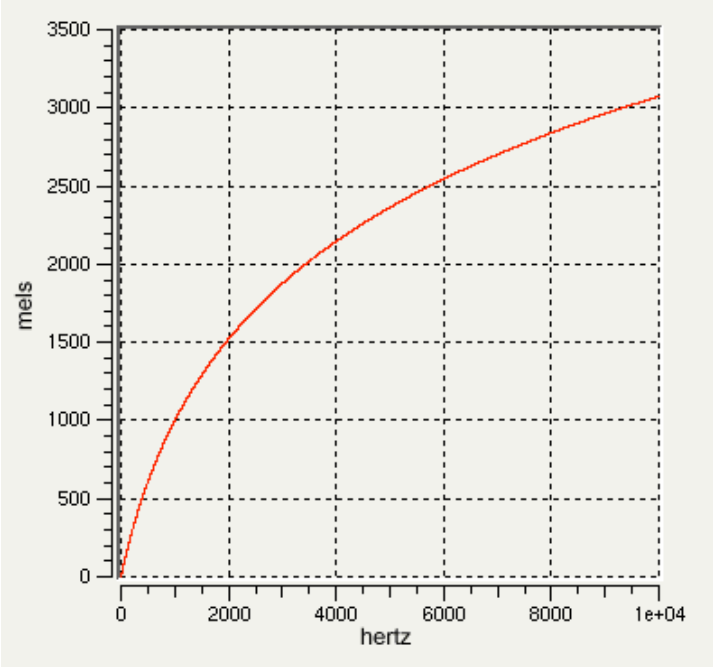

Figura 1 –Mel x Frequência

O MFCC contém tanto informações sobre o tempo como informações sobre a frequência do sinal, tornando-o ideal para sistemas de Reconhecimento Automático de Fala e Reconhecimento Automático de Emoção [12].

A Figura 2 mostra o processo básico utilizado para extrair o MFCC. Os primeiros passos tem como o objetivo compensar atenuação do sinal (Pré-ênfase) e dividi-lo em *frames* comumente aplicando uma função para criação de janelas (*Windowing*). A função para criação de janelas mais utilizada é a *Hamming Windowing*. Após este passo é aplicado a Transformada Rápida de Fourier (FFT) para cada *frame* e depois é mapeado a energia do espectro obtida através da escala mel utilizando *mel-filter banks* (banco de filtros na escala mel). *Mel-filter Banks* são filtros triangulares do tipo passa-baixa comumente utilizados sem reconhecimento de áudio, figura 3. Na maioria das aplicações são utilizados 24 filtros para simular o processamento do ouvido humano. Por fim, a Transformada Direta do Cosseno (DCT) é aplicada a cada *frame* e é obtido um vector com valores reais que representam a característica [14].

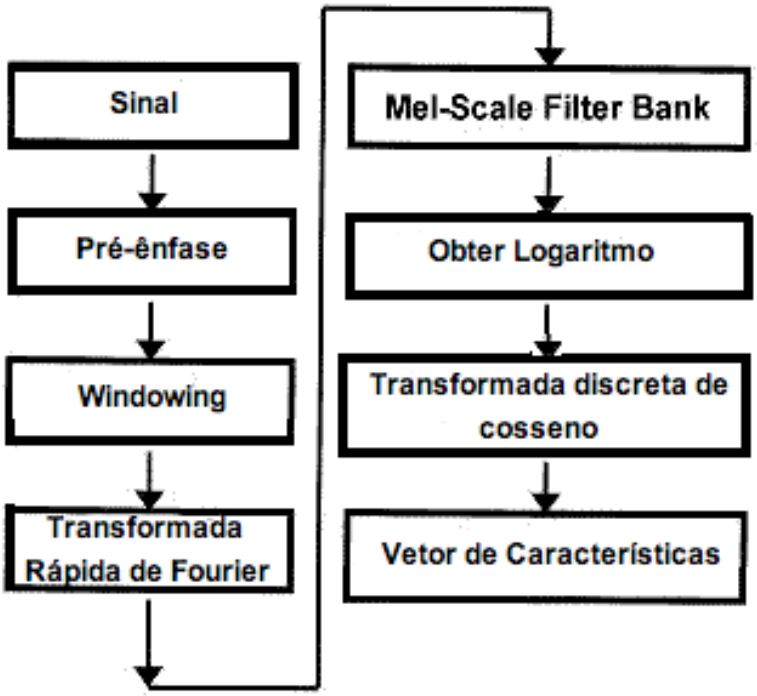

Figura 2 – Extração de MFCCs

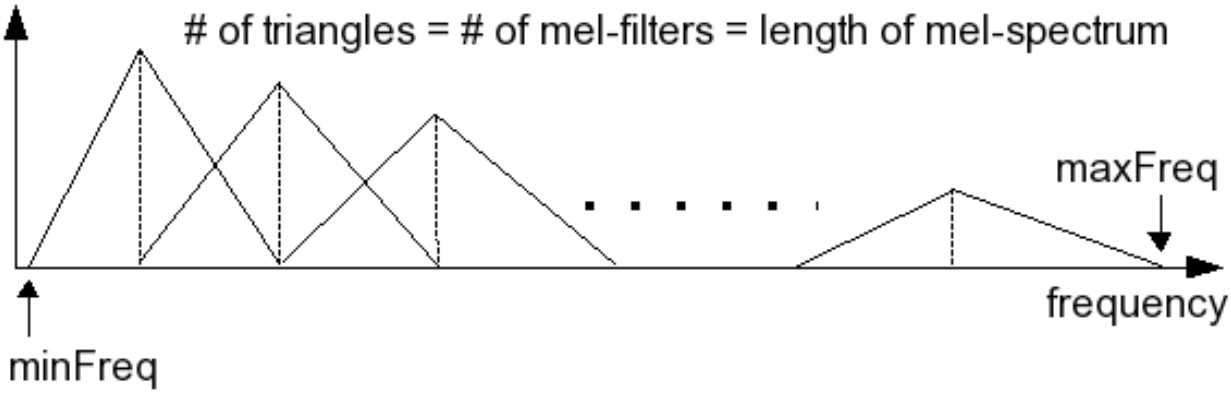

Figura 3 – Mel-filter bank

Mais informações sobre MFCC pode ser encontrado no trabalho de L. Rabiner [13].

Outra característica que ajuda na identificação de emoções a partir de áudio é a Delta e Delta-Delta. Esses parâmetros são obtidos através das derivadas de primeira (Delta) e segunda ordem (Delta-Delta) das características de voz. São utilizadas para

representar as mudanças dinâmicas no espectro de voz e assim detectar variações bruscas dentro do espectro como também analisar a variação dinâmica das características. A adição desta característica ao vetor final aumenta consideravelmente o desempenho dos classificadores, já que estes trabalham melhor com padrões de entrada estáticos [17].

2.5 *Classificação dos Padrões*

Existem vários algoritmos capazes de classificar padrões, dentre eles podemos destacar Redes Neurais Artificiais, Árvores de Decisão, Algoritmos Genéticos entre outros. No caso de MFCCs o método mais utilizado para classificar estas informações é o Hidden Markov Model.

Os HHMs são modelos matemáticos de processos estocásticos, ou seja, processos que geram sequências aleatórias de resultados de acordo com as probabilidades determinadas [15]. Uma cadeia de Markov representa vários estados possíveis para uma determinada situação e as transições, entre um estado e outro, que ocorrem segundo uma certa probabilidade. As primeira aplicações dessa modelagem estavam voltadas para o reconhecimento de fala, sendo os trabalhos de F. Jelinek (IBM) e J. K. Baker (Carnegie Mellon University - CMU), no começo dos anos 70, pioneiros no uso de HMM. Na segunda metade da década de 80, HMM foi aplicado em sequenciamento de DNA, alcançando posteriormente grande importância em todo o campo da bioinformática [19].

Uma HMM é caracterizada pelos seguintes itens [16]:

- N, o número de estados no modelo. Embora os estados estejam escondidos, para muitas aplicações práticas, muitas vezes há algum significado físico ligado aos estados ou aos conjuntos de estados do modelo. Assim, em moeda jogando experimentos, cada estado corresponde a uma moeda distinta tendenciosa. Normalmente os estados estão interligados uns aos outros.

- M, o número de símbolos distintos para cada observação por estado. Os símbolos de observação correspondem à saída física do sistema a ser modelado. No

exemplo de lançamento de moedas esses símbolos seriam *Head* (Cara) ou *Tail* (Coroa).

- A, a distribuição de probabilidade de transição de estado.

Um exemplo simples de um processo estocástico é uma sequência de jogadas de moedas. Imagine que Você está em uma sala dividida em duas partes separadas por uma barreira de tal forma que você não sabe o que acontece do outro lado. No outro lado da sala existe uma outra pessoa que está lançando para o alto uma ou mais moedas. A outra pessoa não irá dizer a você o resultado de cada lançamento de moeda realizado. Nestas condições uma sequência escondida de lançamentos de moedas é realizada tendo como resultado uma sequência de observação consistindo em uma série de Caras e Coroas. Dado este cenário, o problema consiste em como podemos construir uma HMM capaz de explicar a sequência observada de Caras e Coroas. O primeiro problema consiste na escolha da quantidade de estados do modelo. Uma possível escolha poderia ser assumir que apenas uma moeda viciada esta sendo lançada. Neste caso podemos criar um modelo com apenas dois estados onde cada estado corresponde a um lado da moeda, Figura 4. Para completar a especificação deste modelo faz-se necessário apenas decidir qual o melhor valor para o limiar (no caso da figura 4 seria o valor de P(H), probabilidade de dar cara).

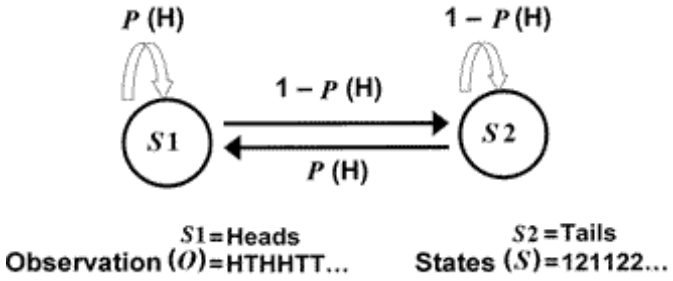

Figura 4 – Modelo de Markov para o lançamento de moedas

Este é apenas um dos modelos possíveis. Também é possível criar um modelo que leve em consideração duas ou mais moedas sendo as próprias moedas

representadas pelos estados. Vale ressaltar que quanto maior a quantidade de estados maior é o número de variáveis que devem ser conhecidas.

O artigo de L. Rabiner [15] contém maiores informações sobre a utilização de HMMs no domínio de aplicações no reconhecimento de padrões em fala, além de informações sobre o processo de treinamento e os tipos de HMMs.

Capitulo 3

O Sistema Proposto

Este capítulo descreve o modelo de legenda proposta e o funcionamento do sistema, mostrando seus diagramas de blocos, diagramas de classes e descreve a arquitetura do sistema.

3.1 *O modelo de legenda*

A legenda que vai ser produzida pelo sistema tenta minimizar a perda de informação sobre o diálogo apresentado utilizando cores para codificar o conteúdo emocional da cena. A emoção extraída da cena será codificada em cores distintas. Para esta primeira versão do sistema o padrão de cores e as emoções a serem codificadas escolhidas foram:

- Raiva: Legenda em vermelho e de tamanho um pouco maior;
- Felicidade: Legenda em azul;
- Tristeza: Legenda em lilás;
- Neutro: Legenda em branco.

3.2 *Descrição do Software Produzido*

O software é responsável por analisar um arquivo de áudio no formato WAV e gerar uma legenda no formato ASS (*Advanced SubStation Alpha*). Para isso é necessário que o utilizador do sistema o alimente com o áudio e uma legenda no formato SRT para ser realizada a extração das características e a sincronização da legenda gerada. O software foi desenvolvido utilizando a linguagem Java.

No todo o sistema possui três módulos mostrados na Figura 5. O primeiro modulo é o responsável por extrair as características do áudio, neste modulo é utilizado a API Sphinx-4. A API Sphinx-4 foi desenvolvida em Java e contém classes necessárias para o processamento e reconhecimento de áudio. O segundo modulo é o classificador, nele se encontram as HMMs responsáveis por determinar qual a emoção que o trecho de áudio possui. O ultimo modulo tem como objetivo gerar a legenda ASS sincronizada.

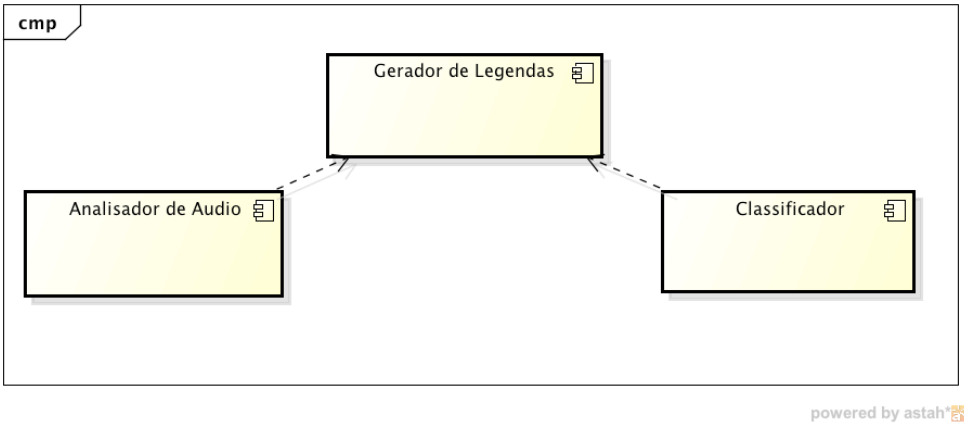

Figura 5 – Módulos do Sistema

A Figura 6 mostra a interface de entrada do software, ela é composta por duas caixas de texto para entrada dos endereços dos arquivos de áudio e da legenda no formato SRT e uma caixa de texto que recebe o endereço com nome do arquivo de saída.

Figura 6 – Tela Principal do Sistema

3.3 *Analise do áudio*

Para processamento do áudio foi utilizado a API Shpinx-4. Esta API implementa todas as classes necessárias para extração de características de arquivos de áudio, bastando o desenvolvedor determinar como será o pipeline de execução e os parâmetros de cada fase. A Figura 7 mostra a sequência de execução que foi utilizada neste projeto. Esta sequência é configurada através de um arquivo XML e pode ser vista no Anexo I.

O primeiro passo é a leitura do arquivo que contém o áudio, logo após é utilizado um filtro de pré-ênfase. O terceiro passo é a criação dos *frames* de acordo com a função para criação de janelas *Hamming Windowing* (1). O tamanho da janela foi ajustado para 25 ms com atraso de 10 ms para criação da próxima. No quarto passo cada *frame* é submetido à FFT .

$$W(n) = 0.54 - (0.46 * \cos\left(\frac{(2*\pi*n)}{N-1}\right)) \quad (1)$$

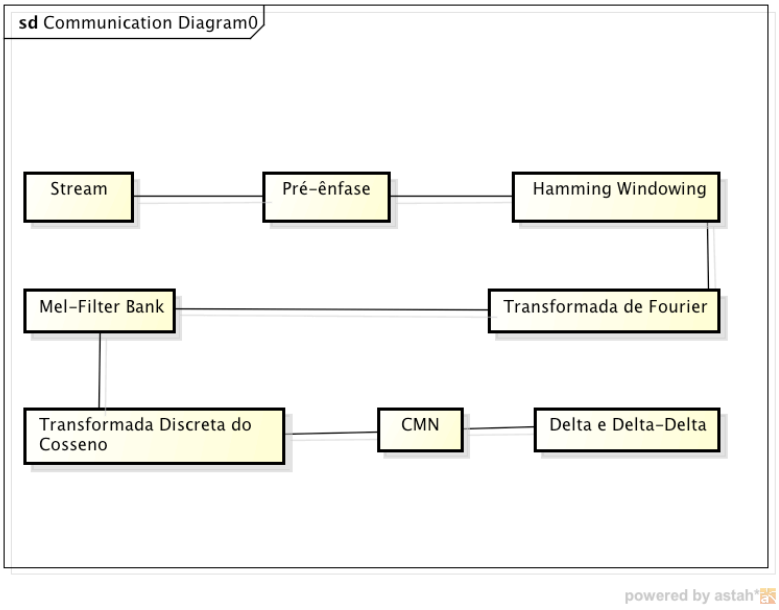

Figura 7 – Sequência de Execução do Pipeline do Sphinx-4

No quinto passo do pipeline é utilizado o *Mel-Filter Bank*, que aplica a função logarítmica (2) para a obtenção de valores na escala Mel. As configurações utilizadas para o *Mel-Filter Bank* foram: 24 filtros, a frequência mínima configurada foi de 300hz e a máxima 3200hz.

$$melFrequency = 2593 * \log\left(1 + \frac{linearFrequency}{700}\right) \quad (2)$$

No sexto passo do pipeline, cada frame é submetido ao DCT. Para melhorar a qualidade dos dados de saída o sétimo passo utiliza o LiveCMN. Esta classe é responsável por aplicar a normalização cepstral média (CMN), às vezes chamado canal médio de normalização, tendo como objetivo reduzir a distorção causada pelo canal de transmissão.

O ultimo passo é responsável por adicionar mais duas características ao vetor de saída chamadas de Delta e Delta-Delta. O vetor final contém 39 posições compostas pelas características: MFCC, Delta e Delta-Delta.

3.4 *Utilizando HMMs para classificação de emoções*

A implementação das HMMs utilizadas neste projeto foram baseadas no framework *OpenSouce Jahmm*, algumas poucas alterações foram realizadas no framework para a leitura dos dados base de arquivos externos e construção do modelo a partir de um arquivo externo

Foram criados quatro modelos de HMMs treinadas com áudio retirados de seriados, filmes e outros tipos de programas televisivos. Cada modelo é responsável por classificar um tipo de emoção e são elas: Raiva, Tristeza, Felicidade e Neutro.

Cada modelo possui quatro estados, e segue a estrutura ergódica apresentada na Figura 8, onde todos os estados estão interligados e existe uma ligação para ele mesmo. O treinamento foi feito em duas etapas, na primeira um modelo intermediário foi gerado utilizando o algoritmo *K-Means* e depois foi aplicado o algoritmo de *Baum–*

Welch, mais informações sobre este tipo de treinamento pode ser encontrado no artigo de L. Rabiner [16].

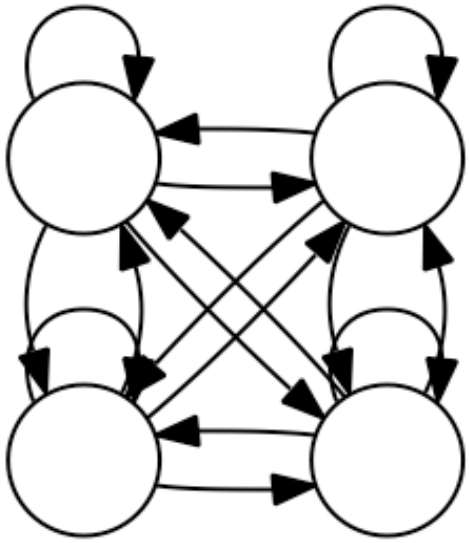

Figura 8 – Modelo de HMM Ergódica

3.5 *Criação da legenda final no formato ASS*

O formato ASS possui a capacidade de aplicação de estilos à legenda. Esses estilos podem conter cor, fonte, tamanho da fonte entre outras inúmeras características que podem ser adicionadas. Para maiores informações sobre a especificação deste tipo de legenda veja a referência [18].

Para criação da legenda final no formato ASS foi desenvolvido um algoritmo que lê a legenda de entrada em formato SRT, identifica os trechos em tempos de milissegundos onde ocorreram diálogos e procura no arquivo com os vetores de características aqueles nos quais os *frames* foram gerados no mesmo intervalo de tempo. Quando as características são obtidas o algoritmo passa todas as características para o classificador, este é responsável por identificar qual emoção é predominante naquele intervalo de tempo e retorna esta informação para a classe

responsável por criar as legendas. Esta comunicação pode ser vista na Figura 9, que mostra um diagrama de sequência demonstrando essas chamadas.

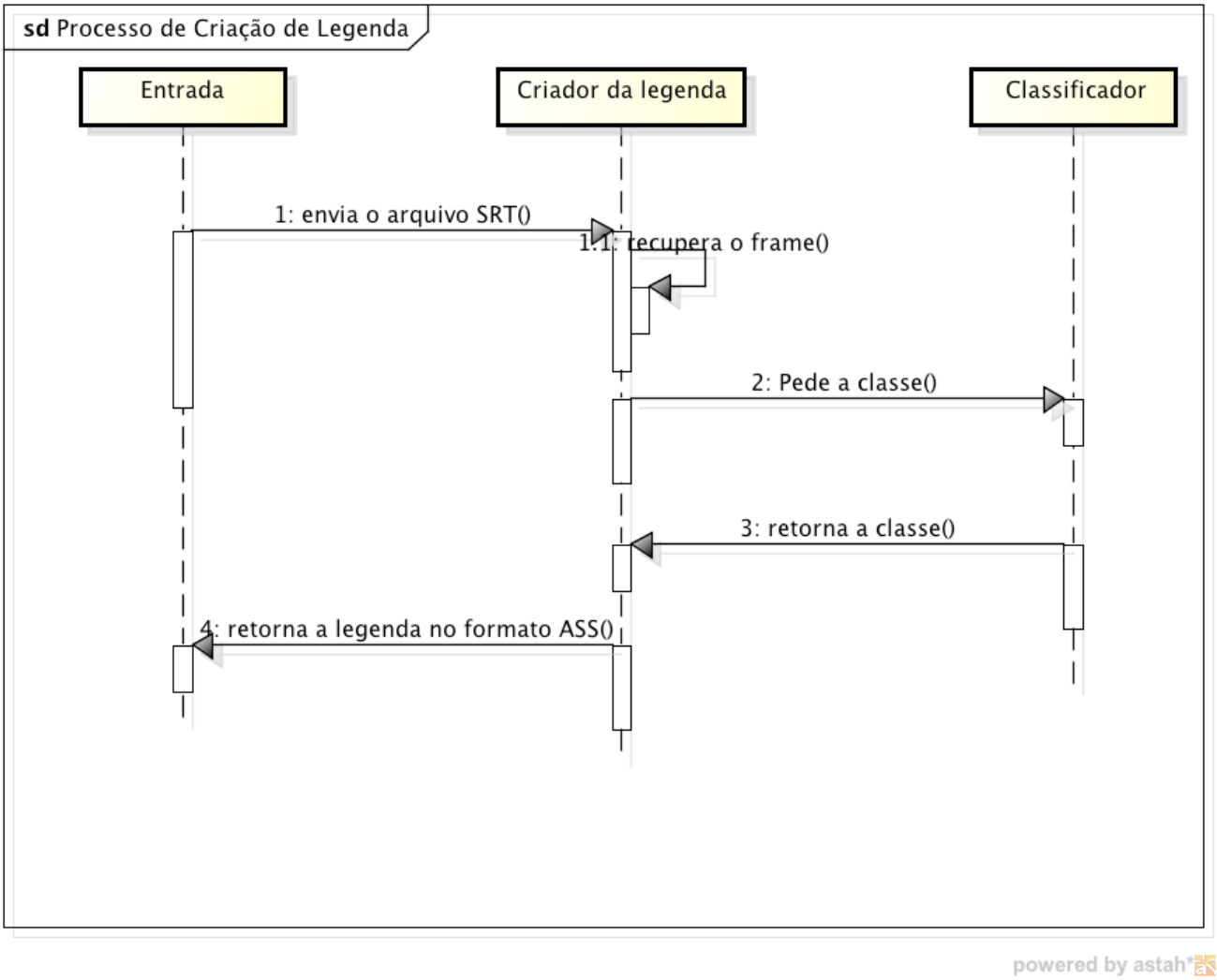

Figura 9 – Comunicação entre os módulos do Sistema

Após todo o arquivo com a legenda em formato SRT ser lido, uma legenda ASS é gerada e configurada com os tempos da legenda SRT e os estilos de cada dialogo determinado pela emoção resultante da operação.

3.6 *Padrões e Arquitetura utilizada*

No desenvolvimento do *software* foram utilizados dois padrões de projetos para facilitar futuras mudanças. Foi aplicado o padrão *Strategy* para criar uma família de algoritmos de classificação, podendo assim, facilmente ser alterado, mudando por exemplo de HMM para Rede Neural Artificial. O segundo padrão de projeto utilizado foi o *Facade* com o intuito de centralizar as chamadas ao core do sistema criando um ponto único de entrada e saída de informações.

O diagrama de classes do sistema criado pode ser visto na Figura 10, que mostra todas as classes, a divisão de pacotes, o relacionamento entre as classes além de mostrar algumas das principais funcionalidades de algumas classes.

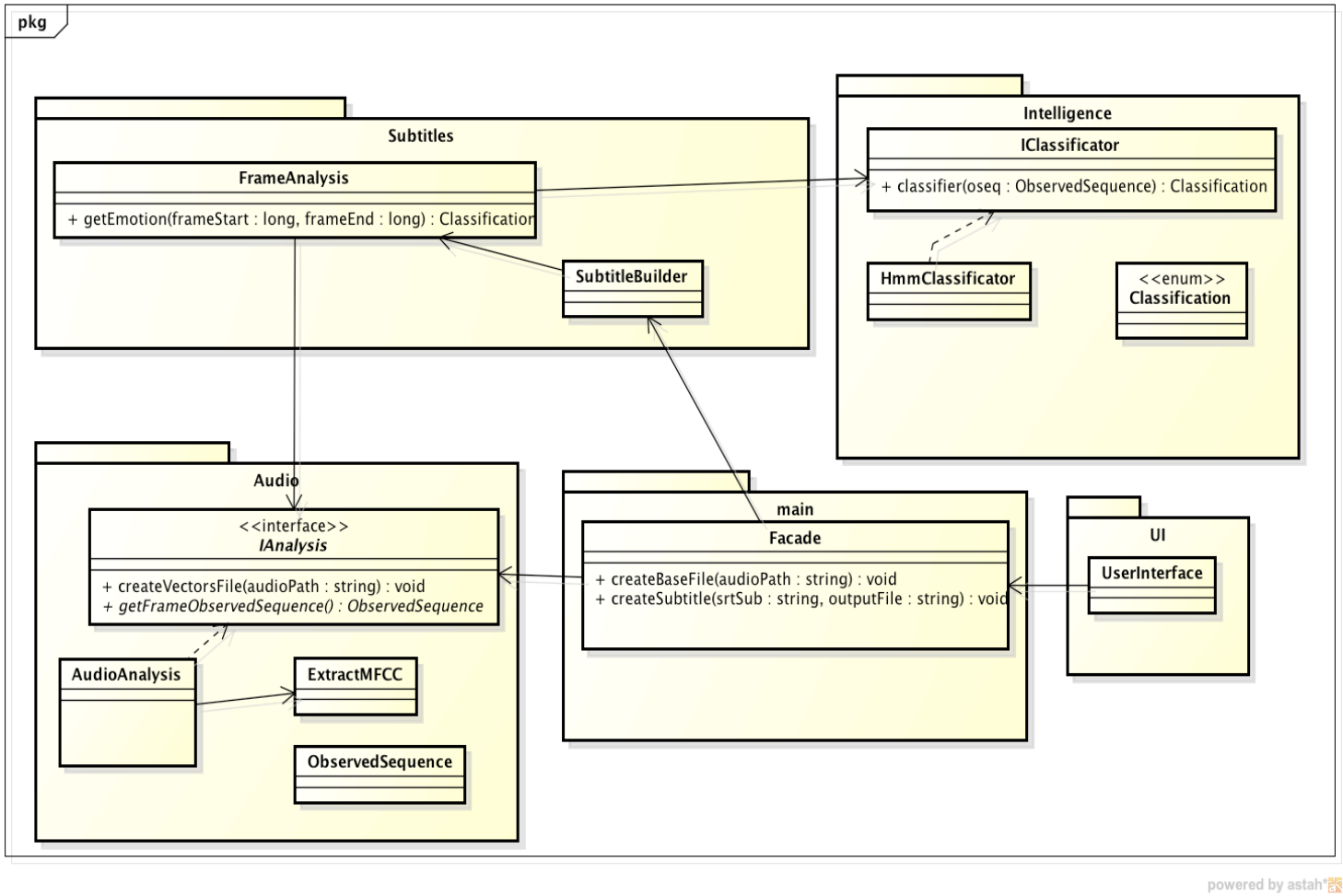

Figura 10 – Diagrama de Classes do Sistema

Responsabilidades das principais classes desenvolvidas para este sistema são:

AudioAnalysis: Percorre todo o áudio e utiliza a classe ExtractMFCC para criar um arquivo texto, que servirá como base de informação para a classe FrameAnalysis, contendo todas as características extraídas do áudio.

FrameAnalysis: Busca na base gerada pela analise do áudio anteriormente realizada, resgatando as informações extraídas de um dialogo e enviado-as para o classificador.

HMMClassificator: Contém os modelos de markov criados para aplicação. Possui o método responsável por realizar o treinamento dos modelos e um método que, dado um conjunto de observações, retorna o melhor resultado.

SubtitleBuilder: Sua responsabilidade é informar ao FrameAnalysis em quais faixas de tempo ocorrem diálogos e escreve o arquivo final contendo a legenda.

Capitulo 4

Resultados Obtidos

4.1 *Material Utilizado*

Pela natureza do problema o banco de dados precisa ser composto de diálogos humanos nas diversas situações selecionadas e com a maior quantidade de pessoas diferentes possíveis, abrangendo assim uma maior quantidade de variações de reações e emoções. Dada a exigüidade de tempo não foi possível fazer um banco de dados extenso. Deste modo foi utilizado um banco de dados criado manualmente através de cortes de áudios de tamanho máximo de dois segundos, extraídos de séries e filmes americanos. O formato padrão de áudio adotado para o sistema tem a seguinte configuração: Formato WAVE, Codificação Microsoft PCM de 16 bit, Mono e taxa de amostragem 22050 Hz.

Pela dificuldade de extração e posterior classificação das informações para montagem do banco de dados, não foi possível montá-los com um grande numero de informações, sendo o banco de dados composto por: 47 arquivos de áudio representando a emoção Raiva, 40 representando a emoção Felicidade, 48 representando a emoção Tristeza e 47 representando Neutralidade

4.2 *Extração das Características e Classificador Utilizado*

Para extração de características, os parâmetros utilizados para o MFCC foram de 24 filtros logaritmos (*Mel-Filters*) para uma faixa de frequência de 300-3200 Hz. O fator utilizado no pré-ênfase foi de 0.97, o tamanho da janela foi definido como 25 ms tendo um atraso de 10 ms para o inicio da próxima. Ao final do processo foram adicionadas as características Delta e Delta-Delta, fazendo com que o vetor final de características contivesse 39 posições.

As HMMs foram configuradas para conter 4 estados. Cada modelo foi treinado utilizando 80% do banco de dados existentes em dois passos. O primeiro passo do treinamento envolve em encontrar um modelo aproximado utilizando o algoritmo de *K-Means* e no segundo passo é utilizado o algoritmo de *Baum-Welch* para aprimoramento do modelo.

4.3 *Experimentos*

O primeiro experimento envolveu o teste dos modelos em um ambiente controlado com os 20% restantes do banco de dados de áudio. Este experimento teve como objetivo avaliar o resultado do treinamento das HMMs utilizadas.

Para o segundo experimento foi extraído o áudio de um episódio da serie How I Met Your Mother. O áudio foi convertido utilizando a ferramenta *Audacity* para o formato padrão estabelecido. Foi definido que a legenda resultante teria as seguintes características: Vermelho e um aumento no tamanho da fonte quando for identificado raiva, Lilás quando for identificado Tristeza, Azul quando identificado Felicidade e Branco quando for neutro.

4.4 *Resultados*

Os resultados do primeiro experimento demonstram que a ferramenta é capaz de classificar emoções com uma boa taxa de acertos, validando o seu uso no sistema construído. Na Tabela 1 pode ser vistos os valores aproximados dos resultados obtidos no primeiro experimento.

Emoção	% de acerto
Raiva	40
Felicidade	60
Tristeza	80
Neutro	60

Tabela 1 – Resultados do Primero Experimento

Para demonstrar o resultado do segundo experimento algumas capturas de telas foram feitas. A Figura 11 mostra uma cena, cujo conteúdo emocional é de Tristeza, utilizando a legenda comum no formato SRT e a Figura 12 mostra a mesma cena com a legenda gerada pelo sistema.

Figura 11 – Cena triste com a legenda em SRT

Figura 12 – Cena triste com a legenda gerada pelo sistema

Abaixo, algumas outras imagens que mostram as legendas para cada tipo de sentimento que o sistema pode codificar. A Figura 13 mostra uma cena de felicidade, A Figura 14 uma cena classificada como Neutro e a Figura 15 uma cena classificada como Raiva.

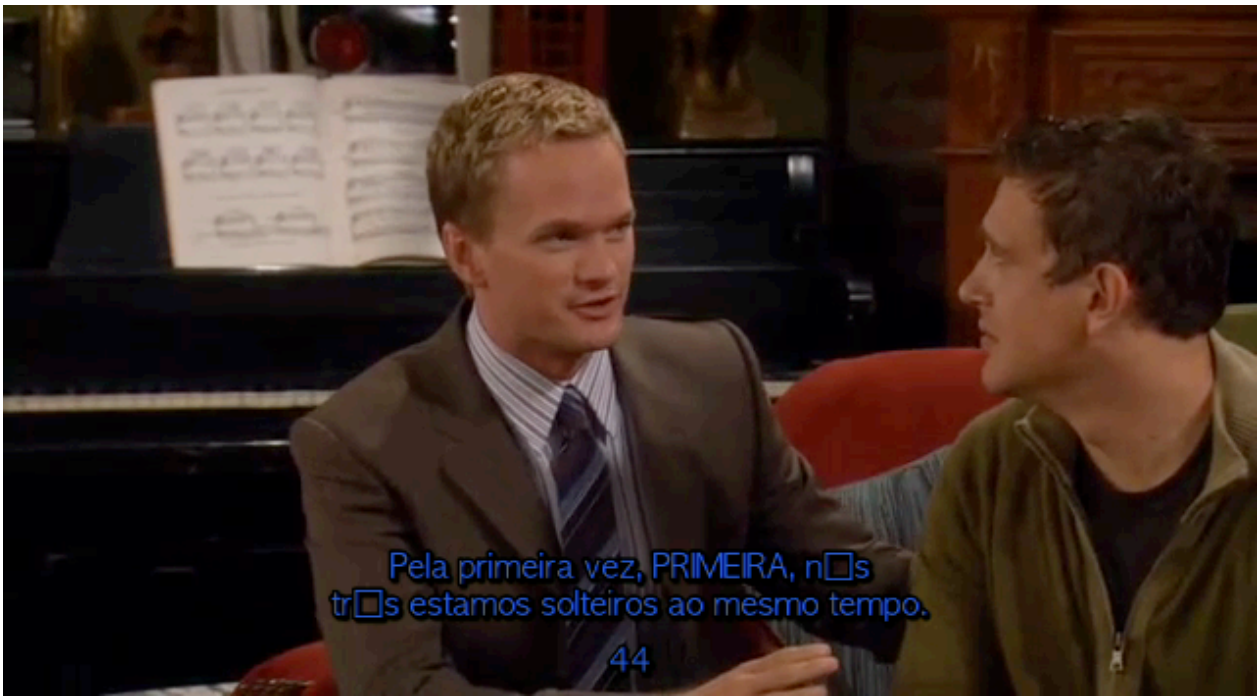

Figura 13 – Uma cena de alegre

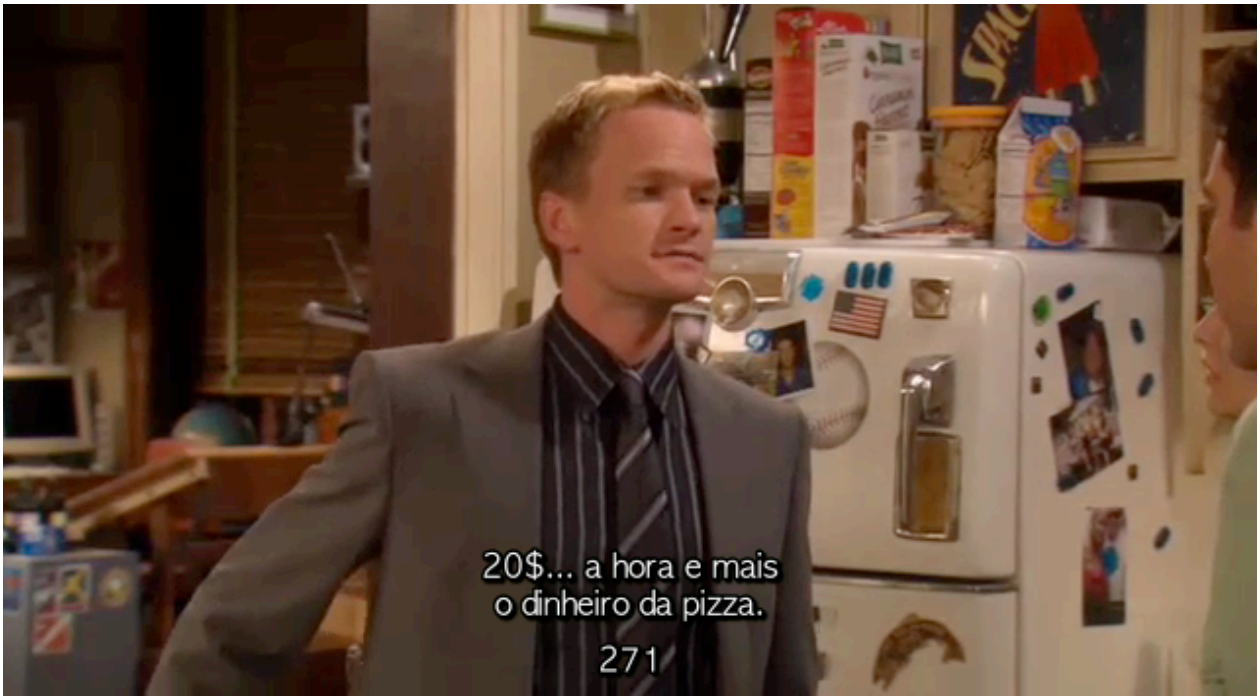

Figura 14 – Cena classificada como Neutra

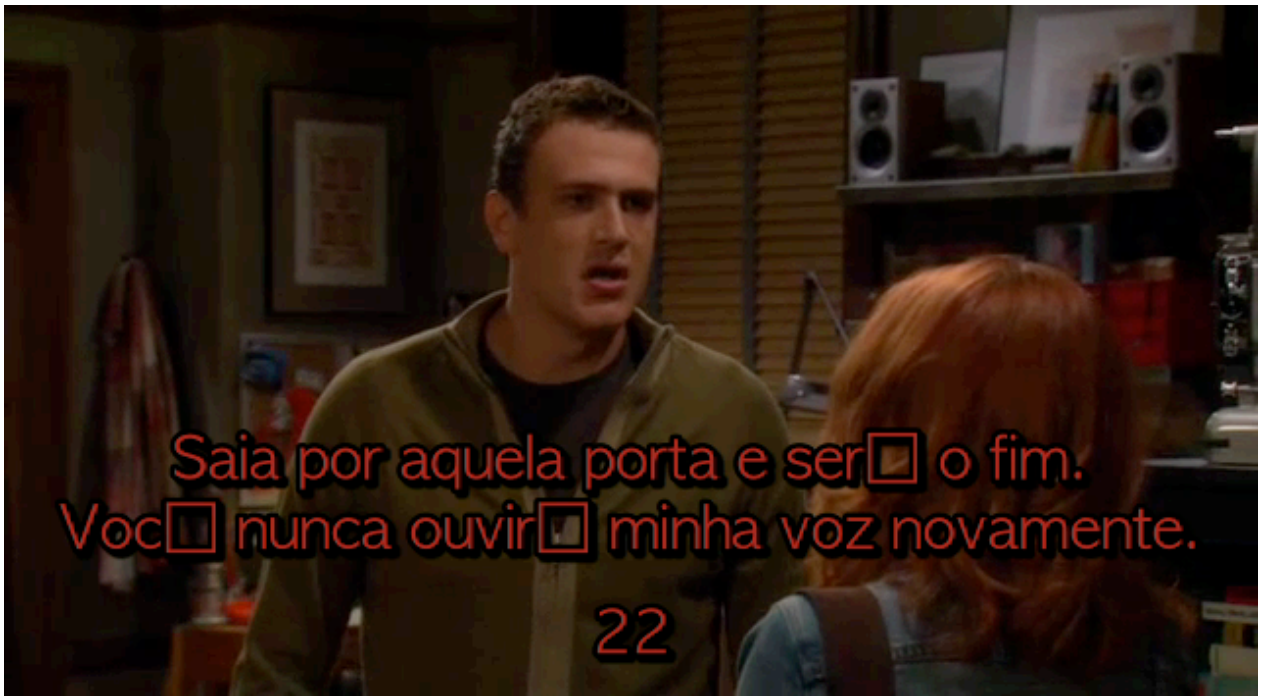

Figura 15 – Cena classificada como Raiva

Durante a visualização do vídeo foi identificado que o erro mais comum de classificação estava em emoções fortes tais como classe Raiva e Tristeza. Estas estavam sendo classificadas de maneira semelhante. Este tipo de erro acontece por que algumas emoções básicas possuem um vetor de características semelhante dificultando a classificação em um ambiente onde o treinamento não teve dados suficientes para serem realizados. Isto pode ser afirmado através da analise da Tabela 1, que contém os resultados do primeiro experimento que demonstram valores aceitáveis na classificação do áudio para as emoções, Felicidade, Tristeza e Neutro. Uma melhora no banco de dados, pode nos dar a chance de treinar melhor os modelos utilizados e aumentado a capacidade de generalização e acertos.

Outro erro identificado na legenda final foi a codificações dos caracteres (charset), o problema não foi solucionado pois ficou fora do escopo do trabalho.

Um exemplo contendo uma parte da legenda gerada pelo sistema proposto pode ser visto no Anexo II.

Capitulo 5

Conclusão e Trabalhos Futuros

5.1 *Conclusão*

O modelo atual do sistema se mostrou capaz, mesmo com poucas informações, de trazer uma nova maneira, mais eloqüente, de se assistir a programas televisivos. Esta prova de conceito evidenciou que a idéia é viável e pode ser utilizada para ajudar pessoas que possuem deficiência auditiva que existem nesse mundo reduzindo o abismo existente nos dias atuais entre as pessoas com deficiências e as ditas normais.

O sistema no estado atual, como toda prova de conceito, necessita de mais testes e refinamentos. Entretanto ele demonstra que se mais funcionalidades forem adicionadas, existe uma grande capacidade de se tornar um produto viável.

Por fim, este trabalho também demonstrou que é possível através dos conhecimentos obtidos na graduação de engenharia de computação a criação de ferramentas capazes de ajudar a melhora da vida de outras pessoas.

5.2 *Trabalhos Futuros*

Muito trabalho deve ser realizado, dentre eles destacam-se:

- **Melhoria no banco de dados para treinamento**

Como já discutido na seção anterior, o banco de dados atual limita os resultados que a ferramenta pode obter. A solução para este problema pode ser de duas naturezas: (i) a compra de um banco de dados, como também (ii) a extração de áudio de filmes, séries ou qualquer outro tipo de mídia que contenha diálogos.

- **Aumentar a quantidade de informação das legendas**

A legenda pode carregar consigo mais informações do que apenas emoção e transcrição de texto. Um exemplo de informação que poderia ser agregado as legendas são: Volume e *Pitch*[5]**.** Essas informações poderiam ser colocadas na legenda em forma de tipo da fonte e tamanho da fonte, respectivamente.

- **Ajustar a legenda final para conter as características da norma da ABNT referente a acessibilidade.**

As restrições de tempo não permitiram adicionar as regras que a ABNT define para as legendas pré-gravadas. Isso seria uma importante adição de qualidade ao produto final.

[5] Representa a frequência fundamental percebida de um som.

Bibliografia

[1] ORGANIZAÇÃO Mundial de Saúde. Magnitude and causes of visual impairment. Disponível: site Organização Mundial de Saúde. Disponível em: http://www.who.int/mediacentre/factsheets/fs282/en, Acessado em: outubro de 2010.

[2] C. Cambra and A. Leal, Comprehension of Television Messages by Deaf Students at Various Stages of Education, American Annals of the Deaf, vol. 153, 2009, pp. 425-434.

[3] T. Nwe, F. Wei, and L.D. Silva, Speech based emotion classification, Proceedings of IEEE Region, 2001.

[4] Associação Brasileira de Normas Técnicas, NBR 15290, Acessibilidade em comunicação na televisão, Rio de Janeiro, 2005

[5] SILVA, A. A. Legenda oculta permite a surdos e pessoas com deficiência auditiva assistir programas, mas precisa melhorar sincronização ao vivo: entrevista. [9 de junho, 2008]. São Paulo: *Revista Sentidos*. Entrevista concedida a Claudete Oliveira.

[6] Cross, A., Segarra, M., & Torrent, A. M. (2000). Llengua oral i llengua escrita a la televisió [Oral and written language on television]. Barcelona, Spain: Publicaciónes de l'Abadia de Montserrat.

[7] D. Fels, D. Lee, and C. Branje, Emotive Captioning and access to Television, Americas Conference on, 2005, pp. 2330-2337.

[8] C. Jensema Viewer reaction to different television captioning speeds. American Annals of the Deaf, 143(4), 318–324.

[9] C. Jensema, R. McCann, and S. Ramsey, Closed-captioned television presentation speed and vocabulary. American Annals of the Deaf, vol. 141, 1996, pp. 284-292.

[10] O Processo de Legendagem no Basil,Vera Lúcia Santiago Araújo, Universidade Estadual do Ceara.

[11] F.S. a., V.K. V.R., R.S. a., A. Jayakumar, and B.A. P., Speaker Independent Automatic Emotion Recognition from Speech: A Comparison of MFCCs and Discrete Wavelet Transforms, 2009 International Conference on Advances in Recent Technologies in Communication and Computing, Oct. 2009, pp. 528-531.

[12] K. Ismail, S.H. Salleh, A.K. Arif, and A. Chowdhury, Heart Sound Analysis Using MFCC and Time Frequency Distribution, Biomedical Engineering, vol. 14, 2006, pp. 946-949.

[13] L. R. Rabiner, and B. H. Juang, “Fundamentals of Speech Recognition,” Prentice Hall, Englewood Cliffs, N.J, 1993

[14] R. Thangarajan and A.M. Natarajan, Robust Front-End Processor combining Mel Frequency Cepstral Coefficient and Sub-band Spectral Centroid Histogram methods for Automatic Speech Recognition, Pattern Recognition, vol. 2, 2009, pp. 67-74.

[15] P.D. Polur and G.E. Miller, Effect of high-frequency spectral components in computer recognition of dysarthric speech based on a Mel-cepstral stochastic model. Journal Of Rehabilitation Research And Development, vol. 42, 2004, pp. 363-371.

[16] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, vol. 77, 1989, pp. 257-286.

[17] ICSI Speech, Título: What are delta features? How do you calculate them?, Disponível em: http://www.icsi.berkeley.edu/speech/faq/deltas.html, Data de acesso: Novembro de 2010

[18] Matroska, Título: SSA/ASS Subtitles, Disponível em: http://www.matroska.org/technical/specs/subtitles/ssa.html , Data de acesso: Novembro de 2010.

[19] L. da Silveira Espindola, Um Estudo sobre Modelos Ocultos de Markov HMM-Hidden Markov Model, Disponível em: http://www.inf.pucrs.br/peg/pub/tr/TI1_Luciana.pdf, Acesso em: Dezembro de 2010.

.

ANEXO I

Arquivo XML utilizado para configuração do SPHINX-4

```
<?xml version="1.0" encoding="UTF-8"?>
<config>
  <component name="cepstraFrontEnd"
type="edu.cmu.sphinx.frontend.FrontEnd">
    <propertylist name="pipeline">
      <item>streamDataSource</item>
      <item>preemphasizer</item>
      <item>windower</item>
      <item>fft</item>
      <item>melFilterBank</item>
      <item>dct</item>
      <item>liveCMN</item>
      <item>featureExtraction</item>
    </propertylist>
  </component>

  <component name="preemphasizer"
          type="edu.cmu.sphinx.frontend.filter.Preemphasizer" />

  <component name="windower"
          type="edu.cmu.sphinx.frontend.window.RaisedCosineWindower">
    <property name="windowSizeInMs" value="25.0" />
```

```
            <property name="windowShiftInMs" value="10.0" />
        </component>

        <component name="fft"
                   type="edu.cmu.sphinx.frontend.transform.DiscreteFourierTransform"
/>

        <component name="melFilterBank"

      type="edu.cmu.sphinx.frontend.frequencywarp.MelFrequencyFilterBank">
            <property name="minFreq" value="300" />
            <property name="maxFreq" value="3200" />
            <property name="numberFilters" value="24" />
        </component>

        <component name="dct"
             type="edu.cmu.sphinx.frontend.transform.DiscreteCosineTransform">
        </component>

        <component                                          name="liveCMN"
type="edu.cmu.sphinx.frontend.feature.LiveCMN"/>
        <component                                  name="featureExtraction"
type="edu.cmu.sphinx.frontend.feature.DeltasFeatureExtractor"/>

        <component name="streamDataSource"
                   type="edu.cmu.sphinx.frontend.util.StreamDataSource">
            <property name="sampleRate" value="22050" />
        </component>
    </config>
```

ANEXO II

Trecho da legenda gerada

[Script Info]
; Script generated by NoName
; mcp@ecomp.poli.br
Title: Default
Original Script: marlonchalegre
Update Details:
ScriptType: v4.00+
Collisions: Normal
PlayDepth: 0
Timer: 100,0000

[V4+ Styles]
Format: Name, Fontname, Fontsize, PrimaryColour, SecondaryColour, OutlineColour, BackColour, Bold, Italic, Underline, StrikeOut, ScaleX, ScaleY, Spacing, Angle, BorderStyle, Outline, Shadow, Alignment, MarginL, MarginR, MarginV, Encoding
Style: Neutral,Arial,26,&H00FFFFFF,&H000000FF,&H00000000,&H00000000,0,0,0,0,100,100,0,0,1,2,2,2,10,10,10,0
Style: Angry,Arial,29,&H001D1DA2,&H000000FF,&H00000000,&H00000000,0,0,0,0,100,100,0,0,1,2,2,2,10,10,10,0

Style: Happy,Arial,26,&H00CD4826,&H000000FF,&H00000000,&H00000000,0,0,0,0,100,100,0,0,1,2,2,2,10,10,10,0

Style: Sad,Arial,26,&H00921C8C,&H000000FF,&H00000000,&H00000000,0,0,0,0,100,100,0,0,1,2,2,2,10,10,10,0

[Events]

Format: Layer, Start, End, Style, Name, MarginL, MarginR, MarginV, Effect, Text

Dialogue: 0,0:00:01.47,0:00:02.66,*Neutral,,0000,0000,0000,,Ok. Onde est·vamos?\N\N2

Dialogue: 0,0:00:03.30,0:00:04.66,Neutral,,0000,0000,0000,,Era junho de 2006.\N\N3

Dialogue: 0,0:00:04.98,0:00:07.27,Happy,,0000,0000,0000,,E a vida deu uma virada inesperada.\N\N4

Dialogue: 0,0:00:07.51,0:00:10.28,Happy,,0000,0000,0000,,Pai, n„o pode apenas pular pra\Nparte onde vocÍ conhece a mam„e?\N\N5

Dialogue: 0,0:00:10.50,0:00:12.23,Neutral,,0000,0000,0000,,Para que vocÍ ta falando a um ano.\N\N6

Dialogue: 0,0:00:13.03,0:00:14.94,Happy,,0000,0000,0000,,Querida, tudo isso que estou\Nte falando È importante.\N\N7

Dialogue: 0,0:00:15.19,0:00:16.46,Neutral,,0000,0000,0000,,… tudo parte da histÛria.\N\N8

Dialogue: 0,0:00:16.70,0:00:18.46,Happy,,0000,0000,0000,,- Posso ir ao banheiro?\N- N„o.\N\N9

Dialogue: 0,0:00:20.00,0:00:23.81,Happy,,0000,0000,0000,,{\i1}O ver„o de 2006 foi\Nmaravilhoso e terrÌvel.{\i0}\N\N10

Dialogue: 0,0:00:24.57,0:00:27.63,Happy,,0000,0000,0000,,{\i1}Pra mim comeÁou muito bem. Na\Nverdade, o primeiro dia foi incrÌvel.{\i0}\N\N11

Dialogue: 0,0:00:28.06,0:00:30.17,Sad,,0000,0000,0000,,{\i1}Eu finalmente fiquei junto da Robin.{\i0}\N\N12

Dialogue: 0,0:00:31.16,0:00:33.53,Sad,,0000,0000,0000,,{\i1}Mas enquanto eu estive fora tendo\Numa das melhores noites da minha vida,{\i0}\N\N13

Dialogue: 0,0:00:34.93,0:00:37.74,Sad,,0000,0000,0000,,{\i1}seu tio Marshall estava tendo uma\Ndas piores noites da vida dele.{\i0}\N\N14

Dialogue: 0,0:00:42.45,0:00:44.09,Sad,,0000,0000,0000,,… isso? Estamos terminando?\N\N15

Dialogue: 0,0:00:44.56,0:00:46.06,Sad,,0000,0000,0000,,Marshall, sinto muito.\N\N16

Dialogue: 0,0:00:46.20,0:00:49.96,Sad,,0000,0000,0000,,Eu tenho que ir pra S„o Francisco\Ne fazer essa escola de arte\N\N17

Dialogue: 0,0:00:50.16,0:00:52.47,Sad,,0000,0000,0000,,e descobrir quem eu\Nsou alÈm de nÛs dois.\N\N18

Dialogue: 0,0:00:53.43,0:00:56.04,Sad,,0000,0000,0000,,E a ˙nica maneira de\Neu fazer isso È se...\N\N19

Dialogue: 0,0:00:57.40,0:00:58.85,Sad,,0000,0000,0000,,a gente n„o se falar por uns tempos.\N\N20

Dialogue: 0,0:01:00.09,0:01:03.86,Angry,,0000,0000,0000,,Uau.. tente nunca! OK?\N\N21

Dialogue: 0,0:01:04.12,0:01:08.78,Angry,,0000,0000,0000,,Saia por aquela porta e ser· o fim.\NVocÍ nunca ouvir· minha voz novamente.\N\N22

Dialogue: 0,0:01:09.62,0:01:10.66,Sad,,0000,0000,0000,,Eu deveria ligar pra ela.\N\N23

Dialogue: 0,0:01:10.85,0:01:13.82,Sad,,0000,0000,0000,,N„o! N„o. Se vocÍ ligar\Nquando ela te pediu pra n„o,\N\N24

Dialogue: 0,0:01:13.92,0:01:15.78,Sad,,0000,0000,0000,,vai te fazer parecer fraco\Ne vocÍ vai se arrepender.\N\N25

Dialogue: 0,0:01:15.93,0:01:17.74,Neutral,,0000,0000,0000,,Olha, qualquer hora\Nque vocÍ sentir que\N\N26

Dialogue: 0,0:01:17.85,0:01:19.55,Neutral,,0000,0000,0000,,deve ligar pra ela, me procure primeiro.\N\N27

Dialogue: 0,0:01:20.53,0:01:22.02,Happy,,0000,0000,0000,,E eu vou te dar um soco na cara.\N\N28

Dialogue: 0,0:01:22.93,0:01:24.43,Happy,,0000,0000,0000,,VocÍ È um bom amigo Ted.\N\N29

Dialogue: 0,0:01:26.53,0:01:29.70,Sad,,0000,0000,0000,,Ei! Ent„o, ouviu as boas notÌcias?\N\N30

Dialogue: 0,0:01:29.96,0:01:32.87,Sad,,0000,0000,0000,,VocÍ esta falando de como a Lily e o\NMarshal terminaram e que a Lily se foi\N\N31

Dialogue: 0,0:01:33.05,0:01:35.72,Sad,,0000,0000,0000,,e nada È mais, nem remotamente\Nimportante do que isso? Sim.\N\N32

Dialogue: 0,0:01:35.93,0:01:36.73,Sad,,0000,0000,0000,,Acho que ele sabe.\N\N33

Dialogue: 0,0:01:37.20,0:01:38.14,Happy,,0000,0000,0000,,Oh meu Deus!\N\N34

Dialogue: 0,0:01:38.58,0:01:41.27,Happy,,0000,0000,0000,,Sinto muito... o que aconteceu?\N\N35

Dialogue: 0,0:01:41.53,0:01:43.66,Happy,,0000,0000,0000,,Bem, ela foi embora.\N\N36

Dialogue: 0,0:01:44.10,0:01:46.00,Happy,,0000,0000,0000,,E eu nem sei se ela vai voltar.\N\N37

Dialogue: 0,0:01:46.32,0:01:50.03,Sad,,0000,0000,0000,,Eu n„o recebi sua mensagem atÈ\Neu acordar. Cara, sinto muito.\N\N38

Dialogue: 0,0:01:50.20,0:01:50.92,Sad,,0000,0000,0000,,Obrigado.\N\N39

Dialogue: 0,0:01:51.11,0:01:53.59,Sad,,0000,0000,0000,,Eu sei que deve ser\NdifÌcil, mas vocÍ esta pronto\N\N40

Dialogue: 0,0:01:53.72,0:01:55.52,Happy,,0000,0000,0000,,pra ouvir algo que n„o sÛ\Nvai te fazer sentir melhor\N\N41

Dialogue: 0,0:01:56.19,0:01:57.71,Happy,,0000,0000,0000,,mas vai efetivamente te excitar?\N\N42

Dialogue: 0,0:01:58.39,0:01:59.35,Happy,,0000,0000,0000,,Claro!\N\N43

Dialogue: 0,0:01:59.76,0:02:04.70,Happy,,0000,0000,0000,,Pela primeira vez, PRIMEIRA, nÛs\NtrÍs estamos solteiros ao mesmo tempo.\N\N44

Dialogue: 0,0:02:05.92,0:02:09.82,Happy,,0000,0000,0000,,Eu sonhei com este dia,\Ngalera, e ser· legend·rio.\N\N45

Dialogue: 0,0:02:10.91,0:02:12.95,Happy,,0000,0000,0000,,Juntos iremos dominar a cidade.\N\N46

Dialogue: 0,0:02:13.44,0:02:18.60,Happy,,0000,0000,0000,,A qualquer hora que alguma garota\Nquiser voltar com um ex, estaremos l·.\N\N47

Dialogue: 0,0:02:19.54,0:02:22.54,Sad,,0000,0000,0000,,A hora que uma garota\Nquiser resolver questıes\N\N48

Dialogue: 0,0:02:22.75,0:02:26.95,Sad,,0000,0000,0000,,com o pai atravÈs de sexo\Ne ·lcool, estaremos l·.\N\N49

Dialogue: 0,0:02:28.35,0:02:30.73,Sad,,0000,0000,0000,,A hora que uma festa de noiva\Nestiver cruzando as ruas de Nova York\N\N50

Dialogue: 0,0:02:30.91,0:02:33.04,Sad,,0000,0000,0000,,numa limusine, colocando\Na cabeÁa para fora do teto\N\N51

Dialogue: 0,0:02:33.21,0:02:38.18,Sad,,0000,0000,0000,,solar gritando "e aÌ Nova\NYork!", nÛs seremos o Nova York.\N\N52

Dialogue: 0,0:02:39.49,0:02:41.58,Sad,,0000,0000,0000,,Senhores, estamos prestes a embarcar...\N\N53

Dialogue: 0,0:02:46.65,0:02:49.42,Happy,,0000,0000,0000,,Poxa gente, vocÍs transaram, n„o foi?\N\N54

Dialogue: 0,0:03:05.36,0:03:06.57,Happy,,0000,0000,0000,,{\i1}Uma coisa que eu\Naprendi no ˙ltimo ver„o{\i0}\N\N55

Dialogue: 0,0:03:06.66,0:03:07.78,Happy,,0000,0000,0000,,{\i1}È que quando um\Namor esta nascendo...{\i0}\N\N56

Dialogue: 0,0:03:10.54,0:03:12.36,Sad,,0000,0000,0000,,{\i1}e quando est· acabando...{\i0}\N\N57

Dialogue: 0,0:03:12.60,0:03:14.81,Sad,,0000,0000,0000,,{\i1}Os 30 primeiros dias\Ns„o muito parecidos.{\i0}\N\N58

Dialogue: 0,0:03:16.31,0:03:18.55,Sad,,0000,0000,0000,,{\i1}A primeira coisa È que vocÍ passa\Na maior parte do tempo na cama.{\i0}\N\N59

Dialogue: 0,0:03:24.04,0:03:25.92,Sad,,0000,0000,0000,,{\i1}Seus amigos n„o conseguem\Nficar te ouvindo.{\i0}\N\N60

Dialogue: 0,0:03:29.88,0:03:33.81,Sad,,0000,0000,0000,,Era uma canÁ„o muito bonita, ent„o...\N\N61

Dialogue: 0,0:03:35.05,0:03:36.35,Sad,,0000,0000,0000,,BeyoncÈ est˙pida!\N\N62

Dialogue: 0,0:03:36.86,0:03:38.26,Sad,,0000,0000,0000,,E vocÍ parece que nunca veste calÁas...\N\N63

Dialogue: 0,0:03:43.84,0:03:46.17,Sad,,0000,0000,0000,,- Ei Marshall!\N- Ei Tedd.\N\N64

Dialogue: 0,0:03:46.74,0:03:47.57,Happy,,0000,0000,0000,,Est· com fome?\N\N65

Dialogue: 0,0:03:48.09,0:03:51.03,Sad,,0000,0000,0000,,Qual È o ponto de comer\Nalgo se ela vai me deixar?\N\N66

Dialogue: 0,0:03:52.04,0:03:54.09,Sad,,0000,0000,0000,,Bem, pelo menos nesse caso,\NÈ vocÍ que "joga fora!"\N\N67

Dialogue: 0,0:03:55.95,0:03:58.01,Sad,,0000,0000,0000,,Vamos, È domingo!\NDia de panquecas!\N\N68

Dialogue: 0,0:03:59.00,0:04:03.44,Happy,,0000,0000,0000,,A Lily fez panquecas? Cara, eu\Namo as panquecas que ela faz.\N\N69

Dialogue: 0,0:04:03.95,0:04:08.07,Happy,,0000,0000,0000,,T„o macias, t„o quentinhas,\Nt„o bem moldadas...\N\N70

Dialogue: 0,0:04:09.45,0:04:11.00,Happy,,0000,0000,0000,,Ainda estamos falando\Ndas panquecas dela?\N\N71

Dialogue: 0,0:04:12.50,0:04:14.47,Happy,,0000,0000,0000,,Vamos, vocÍ tem que comer\Nalguma coisa. O que vocÍ quer?\N\N72

Dialogue: 0,0:04:14.77,0:04:15.48,Sad,,0000,0000,0000,,Cerveja...\N\N73

Dialogue: 0,0:04:15.61,0:04:17.01,Happy,,0000,0000,0000,,N„o, isso È o que\NvocÍ teve no jantar.\N\N74

Dialogue: 0,0:04:17.90,0:04:20.73,Happy,,0000,0000,0000,,Tudo bem. Eu fico com as sobras...\N\N75