UNIVERSIDADE FEDERAL DO PARANÁ

BRUNO CORTES LOPES SIQUEIRA
GUILHERME MACIEL DA SILVA
MOISÉS PEREIRA OLIVEIRA
WESLEY CALESSO ARANTES
WILLIAN BAIERLE DE OLIVEIRA

J-SOM

CURITIBA
2008

BRUNO CORTES LOPES SIQUEIRA
GUILHERME MACIEL DA SILVA
MOISÉS PEREIRA OLIVEIRA
WESLEY CALESSO ARANTES
WILLIAN BAIERLE DE OLIVEIRA

J-SOM

Trabalho de graduação apresentado à disciplina Trabalho Conclusão de Curso do curso de Tecnologia em Sistemas de Informação da Universidade Federal do Paraná.

Orientadores: Roberto Tadeu Raittz
Dieval Guizelini

CURITIBA
2008

DEDICATÓRIA

“Dedico este trabalho à minha família, meus amigos e colegas de projeto e a todos que ajudaram de algum modo a idealizar essa idéia e a concretizarmos nosso objetivo.”
Bruno Cortes Lopes Siqueira

“Dedico esse trabalho a todas as pessoas que direta ou indiretamente, contribuíram para o meu crescimento pessoal e profissional e a alcançar mais essa façanha de minha vida.”
Guilherme Maciel da Silva

“Dedico este trabalho a todos que contribuíram para a minha formação, moral, pessoal e acadêmica.”
Moisés Pereira Oliveira

“Dedico este trabalho a minha família, a todas as pessoas que acreditaram, me incentivaram e todos aqueles que ajudaram de alguma forma para tornar esse trabalho uma realidade.”
Wesley Calesso Arantes

“Dedico este trabalho a essa equipe da qual eu faço parte há três anos e a qual trabalhou duro para que tudo isso fosse realidade.”
Willian Baierle de Oliveira

AGRADECIMENTOS

Agradecemos a Deus por ter nos dado força, paciência e persistência no decorrer do projeto. Às nossas famílias pela compreensão, pela paciência e pelo incentivo. Às fonoaudiólogas Cleusa Pierin e Luciane Kowalewski que nos deram auxílio importantíssimo, tanto no conhecimento técnico quanto na companhia nas atividades nas escolas. Às pessoas que nos ajudaram com conselhos em relação à interface do programa, professora Sandramara Scandelari Kusano de Paula Soares e Paula Silvestre Guimarães. Aos orientadores, Roberto Tadeu Raittz e Dieval Guizelini. A todos os professores que contribuíram na formação do nosso conhecimento. Às escolas Catavento e Sol Criança, por terem nos atendido muito bem quando estávamos precisando de vozes de crianças para fazer os testes. À Marisa Rodrigues pela gravação dos arquivos de vídeo e áudio que fazem parte do sistema. Ao Giuliano Branco Dal Piva por ajudar em dúvidas técnicas, e em especial Guilherme Silva ao Felipe Pereira por substituí-lo no trabalho sempre que preciso. Agradecemos principalmente uns aos outros da equipe do projeto, pela determinação, pelo compromisso, pela responsabilidade no desenvolvimento do projeto e também por compartilhar suas experiências e seus conhecimentos.

Equipe J-Som.

RESUMO

Este sistema auxiliará profissionais da área da educação e saúde infantil, como educadores e fonoaudiólogos, no processo de aprendizado e também na detecção de problemas na fala de crianças na faixa etária entre 3 e 6 anos. Foi utilizado um conjunto de 26 fonemas, os quais englobam todo o quadro fonético da língua portuguesa, tais fonemas foram devidamente selecionados por fonoaudiólogas. A ferramenta também possui características de um jogo, no qual a criança aprende brincando. Foram utilizados artifícios da engenharia semiótica para a construção de uma interface atrativa e amigável.

O sistema analisa as palavras pronunciadas pela criança e retorna um *feedback* parabenizando-a quando estiver certa e incentivando a melhorar quando o resultado não for satisfatório e ao final de cada sessão um relatório que descreve o desempenho da criança é gerado. Este relatório pode ser analisado por um fonoaudiólogo para detectar eventuais problemas na fala das crianças.

Para fazer o reconhecimento dos fonemas foram utilizados conceitos de inteligência artificial, especificamente Rede Neural Artificial utilizando a técnica de treinamento supervisionado. A rede neural utilizada foi a J-Fan, manuseada pela interface Easy Fan. O "treinamento" da rede foi realizado com um conjunto de aproximadamente 700 amostras, na média de 25 exemplares para cada classe de fonema. Foram capturadas as vozes de aproximadamente 100 crianças na faixa etária mencionada anteriormente.

Palavras-chave: Reconhecimento automático de fala. Análise de pronúncia. Rede neural.

ABSTRACT

This system helps professionals from Healthy and Education areas, as teachers and audiologists, in the process of learning and also in the illness related to the voice from kids around 3 and 6 years old. It has been selected 26 phonemes, which encompass the entire phonetic table from the Portuguese language. All phonemes have been selected by professional audiologists. This tool works as a game, where the child learns while he is playing. Semiotic engineering has been used to build an attractive and friendly interface.

The system analyses the child's speech and returns a feedback, congratulating him if the system considers the speech correct or incentivizing him to improve his skills if the system considers the speech wrong. At the end of each session, a report that shows the child's performance is created. This report can be used by audiologists to help detecting voice illness in their patients.

To develop the speech recognition, it has been used Artificial Intelligence concepts, specifically Artificial Neural Network, using a supervised training system. The neural network used is the J-FAN, handle by the Easy-Fan interface. The neural network training has been done using a set of 700 samples, in an average of 25 samples for each phonetic. The samples were captured from the speech of approximately 100 children, from 3 to 6 years old.

Key-words: speech recognition, pronunciation analysis, neural network

LISTA DE FIGURAS

LISTA DE DIAGRAMAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

API - *Application Programming Interface* (Interface de Programação de Aplicativos).
CVS - *Concurrent Version System* (Sistema de Versões Concorrentes).
DVXXX – *Data View* XXX (Visão de Dados XXX).
GHz – Giga Hertz.
IBM - *International Business Machines*.
IDE - *Integrated Development Environment* (Ambiente de Desenvolvimento Integrado).
LAN - *Local Area Network* (Rede de Área Local).
MATLAB - *Matrix Laboratory*.
OCR - *Optical Character Recognition* (Reconhecimento de Caracteres Ópticos).
PDF - *Portable Document Format* (Formato de Documento Portável).
RAF – Reconhecimento Automático da Fala.
RNA - Rede Neural Artificial
UCXXX – *Use Case* XXX (Caso de Uso XXX)
UML – *Unified Modeling Language* (Linguagem de Modelagem Unificada).
WAV – *WAVEform audio format*.

SUMÁRIO

1. INTRODUÇÃO

A maior parte do aprendizado das letras e palavras é feita com base em escutar e repetir, seja em casa, na escola ou em outros lugares e tudo o que a criança ouve tende a ser repetido na fase de aprendizado. Isso faz com que a melhor maneira de ensiná-la seja falar-lhe as palavras e pedir para que repita. Esse mecanismo é feito involuntariamente quando os pais pedem para o bebê falar “papai” e “mamãe”, por exemplo. Na escola, o professor executa essa tarefa, mostrando em desenhos e figuras todas as letras do alfabeto e as sílabas que elas formam, para que a criança repita e aprenda os fonemas.

1.1. Tema

O J-Som é um avaliador de pronúncia de fonemas. Ele entra diretamente no mecanismo citado na seção anterior, grosso modo, sendo semelhante a um “professor eletrônico”, embora não tenha esse foco.

Ele consiste em uma voz reproduzindo sílabas ao mesmo tempo em que exibe a escrita das mesmas; após a reprodução, o sistema solicita a interação da criança pedindo a ela que repita a sílaba reproduzida anteriormente e a cada sílaba repetida apresenta uma mensagem de retorno em forma de animação, a qual informará se a reprodução foi correta ou incorreta. Com isso o professor pode ensinar o funcionamento do software e trabalhar de maneira mais distribuída com cada aluno.

Em casa, o sistema trabalhará com a criança da mesma forma que o professor trabalha em sala de aula, sendo uma espécie de lição de casa, onde no dia seguinte o aluno estaria mais treinado e ambientado com as sílabas e fonemas, possibilitando assim mais rapidez e qualidade ao aprendizado.

Nos consultórios de Fonoaudiologia ele será utilizado para uma espécie de pré-consulta ou triagem, antecedendo a consulta com o fonoaudiólogo. Com esse sistema, a criança pode interagir jogando e ser avaliada ao mesmo tempo, auxiliando o fonoaudiólogo fornecendo-lhe dados sobre a sua fala.

1.2. Problema

A educação infantil é a mais importante na vida da pessoa, onde ela tem as primeiras lições e aprende a gostar de aprender, mas na situação atual do Brasil ainda existem grandes problemas nessa área. Um deles é o fato de alguns professores que trabalham com crianças pequenas não estarem totalmente qualificados para isso, além de se depararem com turmas numerosas onde não é possível trabalhar com a atenção devida a todos os alunos, o que prejudica muito o desenvolvimento da criança.

Outro grande problema, segundo as profissionais de Fonoaudiologia que fazem parte da equipe do projeto, é o fato de não haver acompanhamento devido no desenvolvimento da fala da criança. A maioria das escolas menores, principalmente as públicas que ficam em cidades pequenas, não tem condições de ter em seu quadro de funcionários um profissional de Fonoaudiologia. Então resta disponibilizar consultas na rede pública de saúde para cada uma das crianças, o que nos dias atuais leva muito tempo, onde há casos de um aluno esperar até um ano para sua primeira consulta devido à grande fila que existe e ao pouco número de profissionais nas prefeituras. A perda de tempo prejudica muito a criança, pois quanto mais cedo se detecta o problema na fala, melhor será o resultado do tratamento e menos tempo será preciso para que ele dê resultado. Não é raro encontrar em séries avançadas do ensino fundamental alunos com problemas de fala que até então não tenham sido percebidos. Além de desmotivar a pessoa a aprender, tais problemas

podem fazer com que ela seja motivo de brincadeiras de mau gosto, discriminação e rejeição.

1.2.1. Resumo das Principais Necessidades dos Envolvidos

TABELA 1 – NECESSIDADES DOS ENVOLVIDOS

Necessidade	Prioridade	Preocupações	Solução Atual	Soluções Propostas
Ensino de pronúncia a crianças	Alta	Caso o aprendizado da criança seja ruim, o desenvolvimento dela pode ser comprometido. Salas de aula numerosas no ensino público.	Ensino baseado no fato de o professor ensinar e revisar todas as pronúncias da criança.	Com o J-Som, o professor não ficará sobrecarregado com tantos alunos ao mesmo tempo. Adotando o uso desse software, ele poderá ensiná-los a interagir com ele, permitindo ter uma atenção mais direcionada e forte para com os que têm dificuldades. Além disso, utilizando-o em casa, a criança chega à aula mais preparada tendo praticado anteriormente.
Identificação de problemas na fala	Alta	Problemas na fala são tratados e corrigidos com mais eficácia se constatados precocemente. Quanto mais tarde se identifica o problema, mais difícil será de corrigi-lo.	Em escolas públicas cabe aos pedagogos identificarem esses problemas, pois raramente as famílias têm condições de realizar consultas periódicas com um fonoaudiólogo.	Com o J-Som, em certo tempo de interação com a criança, é possível identificar se ela tem problemas de fala ou não, caso não consiga repetir a sílaba depois de numerosas tentativas.

1.3. Hipóteses

Um dos primeiros problemas que devem ser resolvidos é o acesso fácil do aluno ao fonoaudiólogo. É muito difícil uma escola pública de pequeno porte ter um profissional dessa área em seu quadro de funcionários, assim como é igualmente difícil levar todas as crianças aos consultórios. A contratação de mais profissionais pode a longo prazo ajudar a diminuir a espera na fila por uma consulta. Mas para

ajudar a amenizar esse problema em pouco tempo é preciso saber quem realmente necessita de acompanhamento profissional. Como os professores não têm todo o conhecimento de desenvolvimento da fala, problemas auditivos etc., ter uma ferramenta que faça essa observação inicial analisando a fala pode dar um grande auxílio na filtragem dos indivíduos que terão de ser avaliados pelos fonoaudiólogos, diminuindo assim a fila pela consulta. E se essa ferramenta indicar o modo certo de pronunciar cada fonema a ajuda será ainda maior, pois crianças cujos problemas de dicção são mínimos poderão corrigir suas falas somente interagindo com ela. O J-Som seria essa ferramenta.

1.3.1. Extração de características usando o gráfico do espectro sonoro do fonema

Em experimentos feitos com os sons gravados notou-se que os fonemas tinham características semelhantes quando seus respectivos gráficos eram desenhados. Desta forma, criou-se a hipótese de que a utilização desta técnica poderia proporcionar maiores taxas de acerto se comparados com outros métodos de extração de características no campo de Reconhecimento Automático da Fala.

Devido a essa semelhança optou-se em utilizar os gráficos dos fonemas no domínio do tempo para aplicar a extração das características. Pelo fato de ser uma técnica ainda não utilizada (não existem indícios da utilização dessa técnica em outros projetos), por conseqüência esta é a primeira a tentativa de implementação da técnica referenciada.

1.4. Objetivos

O J-Som tem como principal objetivo auxiliar crianças de 3 a 6 anos na fixação dos fonemas através de sílabas.

Com o J-Som a criança aprimorará sua pronúncia através de uma interação humano-computador de alto nível. Ele poderá ser utilizado tanto nas escolas, com o auxílio dos professores, quanto nas casas dos alunos, com o auxílio dos pais, e cada escola que aderir ao seu uso terá cópias para seus alunos da educação infantil para que possam instalar o programa nos computadores de suas casas.

O objetivo é fazer do J-Som um software utilizado em muitas escolas no auxílio ao aprendizado das crianças, possibilitando futuramente novas versões com funcionalidades mais completas, tais como reconhecimento de palavras e frases. Para atingir a meta do projeto, é necessária uma interface altamente voltada para crianças pequenas, pois quanto mais cores vivas e interatividade, mais fácil será prender a atenção delas. Na faixa de idade de 3 a 6 anos as crianças costumam ser hiperativas e inquietas, andando para todos os lados e realizando inúmeras atividades; deste modo, o primeiro passo é prender a sua atenção, para depois começar a interagir com elas. Visando essa dificuldade, é extremamente importante o auxílio de profissionais que trabalham diretamente com crianças, pois eles conhecem as melhores maneiras de se atingir essa meta.

1.5. Justificativa

Com a tecnologia cada vez mais presente em nossas vidas é inevitável sua participação no aprendizado das crianças, para aperfeiçoar o que já foi construído anteriormente na educação delas. Sendo assim, o J-Som entra na educação infantil como ferramenta tecnológica capaz de prover um grande suporte aos pais e professores nos primeiros anos de desenvolvimento da criança, além de detectar se ela tem algum problema em sua dicção, podendo priorizar um possível tratamento. Com uma melhoria na sua educação básica, mais fácil será investir no seu desenvolvimento.

Existem vários softwares que analisam a fala de crianças no ensino das letras, sílabas ou palavras, porém poucos que forneçam um retorno direto. Dentre os principais concorrentes do J-Som destaca-se o FonoFlex ® [10], porém este não interage diretamente com o usuário dando retornos, pois tem o foco em proporcionar relatórios detalhados ao profissional de Fonoaudiologia.

O fato de haver interação direta com a criança em forma de retorno, fazendo com que ela perceba de forma intuitiva como está sua fala e podendo melhorá-la na própria interação, torna o J-Som mais adequado para trabalhar-se em escolas com auxílio dos professores e em casa com auxílio dos pais.

Sua utilidade nos consultórios de Fonoaudiologia deve-se ao fato de que é difícil conquistar a confiança das crianças na primeira consulta. Segundo as fonoaudiólogas do projeto, geralmente elas ficam acanhadas e demoram a se comunicar plenamente com o profissional, o que torna mais lento o processo de avaliação da sua fala. Desenvolvido em forma de jogo, o J-Som interage com a criança, prende sua atenção e ao mesmo tempo avalia sua dicção na pronúncia das sílabas. A grande vantagem é que logo na primeira consulta o fonoaudiólogo consegue obter informações necessárias para uma pré-avaliação do indivíduo, o que seria mais difícil sem outra ferramenta ou jogo.

1.5.1. Motivo da utilização dos fonemas

O projeto inicial baseava-se na emissão sonora das letras do alfabeto como nominativo. Desta forma haveria algumas distorções na emissão, devido alguns fonemas ao ser pronunciados teriam o acréscimo de outros fonemas como é o caso do “F” (éfe); “J” (jota); “L” (éle); “M” (eme); “N” (ene); “R” (érre); “S” (ésse).

Desta forma, seguindo orientações de profissionais da área de Fonoaudiologia, foi abandonada a idéia de utilização de letras, para então adotar o uso de fonemas direcionados para a alfabetização, sendo esta, baseada na

aprendizagem de um fonema com o apoio de uma vogal (ex. “PA”), utilizado no método silábico, muito utilizado pelos profissionais de Fonoaudiologia. Se fosse isolado seria avaliado somente o som, (som isolado sem vogal, ex “P”), também utilizados por estes profissionais.

Baseado nisso, foi utilizada a sílaba para o levantamento sonoro, a fim de que num segundo momento seja possível isolar com o programa, somente o fonema (o som).

A fim de simplificar e evitar as análises fonéticas e lingüísticas e executando os métodos de alfabetização foi desenvolvida uma forma onde o programa pode decodificar os sons de uma maneira lógica sendo possível sua análise e avaliação de acordo com os parâmetros utilizados pelos profissionais de Fonoaudiologia.

Os resultados foram obtidos com o apoio das profissionais e dos programas sugeridos e utilizados como apoio.

O programa ora desenvolvido atende as necessidades atuais dos profissionais de Fonoaudiologia e dos de Educação como um material de suporte.

2. REVISÃO DA LITERATURA

O campo da informática que utiliza Inteligência artificial tem chamado a atenção de pesquisadores principalmente pelo fato de possibilitar a execução de atividades que antes só eram possíveis com a participação do ser humano, e em muitos casos ainda que o mesmo fosse especialista no assunto. Como, por exemplo, seria possível enviar soldados a países estrangeiros e os mesmos pudessem se comunicar com outras raças sem a ajuda de um tradutor? Graças a anos de pesquisas desenvolvidas neste campo, fatos como esse começam a se tornar realidade, com a ajuda de dispositivos equipados com sistemas inteligentes muitas tarefas substituem a deficiência e inexperiência humana e agilizam os meios de comunicação. Em muitos destes casos, os sistemas inteligentes utilizam o reconhecimento de voz como parâmetro para seu treinamento. É através deste método que os sistemas conseguem ganhar experiências em determinadas atividades. A seguir se encontram alguns desses exemplos.

2.1. Sistemas anteriores

2.1.1. Programa transforma voz em linguagens de sinais

A IBM, com o apoio de algumas universidades inglesas desenvolveu um sistema que transforma sinais de áudio em linguagens de sinais [14]. O programa, chamado de “SiSi - Say it sing it” possui um personagem digital que reproduz as vozes faladas em gestos que podem ser identificados por deficientes auditivos. A idéia é que o sistema possa ser utilizado em equipamentos de comunicação, como televisões, videogame e DVD *players*. O dispositivo seria parte do pacote de opções do equipamento e poderia ser acionado em qualquer situação.

2.1.2. Sistema tradutor de línguas

A DARPA - Agência para projetos de pesquisas avançadas de defesa dos EUA - desenvolveu um sistema chamado TRANSTAC [16] cujo objetivo é traduzir em tempo real vozes geradas entre as línguas árabes e inglesas. A idéia é possibilitar a comunicação entre os soldados americanos e civis de países árabes facilitando o trabalho de inspeções dos soldados americanos e dispensando a ajuda de um tradutor. O sistema é de extrema complexidade, pois além de trabalhar com as dificuldades naturais em fazer o reconhecimento de voz ainda trabalha com ambientes onde o sinal ruído é inerente. A tecnologia já foi utilizada no Afeganistão e mais recentemente no Iraque. A DARPA ainda trabalha em um projeto chamado GALE (Exploração da linguagem autônoma global), um programa que recebe fluxo de telejornais estrangeiros e traduz para o inglês. A expectativa é que o programa possa traduzir várias línguas, em tempo real, com uma precisão próxima de 90%.

2.1.3. Atendimento Telemar com reconhecimento de voz

A Telemar utiliza um sistema no qual uma atendente virtual pode realizar o pré-atendimento ao cliente [15]. Neste caso, ao invés de o cliente ter que digitar números para escolher o departamento adequado para resolver seu problema, ele simplesmente conversará com a atendente virtual, que se fará o encaminhamento, de acordo com a solicitação do usuário. Esta solução foi criada pela Nuance, empresa americana com experiência neste tipo de projeto. A idéia é agilizar o processo de pré-atendimento e encaminhar o cliente à área desejada, utilizando para isso apenas comandos de voz. Em alguns casos o cliente nem chega a falar com um atendente real, em tarefas como, “solicitar segunda via da nota fiscal” e “solicitar o valor gasto até o momento”.

2.1.4. Reconhecimento de voz dependente de locutor utilizando RNAs

Este trabalho foi efetuado por um aluno da Engenharia da Computação da Universidade de Pernambuco [5]. Seu objetivo é identificar com o máximo de precisão fonemas e frases faladas por um locutor.

A divisão básica deste trabalho é a seguinte:

- Pré-processamento.
- Extração das características do sinal da fala.
- Treinamento e Teste

2.1.4.1.1. Pré-Processamento:

A fase de pré-processamento envolveu os processos de captura do áudio até a extração das características para fazer o treinamento da rede.

Após a captura do áudio feita com um microfone a voz capturada é passada a um filtro na fase chamada de “pré-ênfase”. A finalidade do filtro é compensar a atenuação nas altas freqüências do sinal da fala, gerado pelo processo de produção da fala na glote, tornando seu espectro de freqüência mais plano. Após a aquisição da fala e a etapa de pré-ênfase os sinais são divididos em quadros e janelamentos. O janelamento é utilizado para aumentar as informações espectrais de uma amostra de sinal, este aumento de informações é decorrente da minimização das margens de transição em forma de ondas truncadas e de uma melhor separação de sinais com amplitudes diferentes e freqüências parecidas. Após esta fase chegou-se ao processo chamado de *Endpoints*. O objetivo principal deste processo é determinar o início e o fim de uma locução e eliminar as partes silenciosas. Isto foi feito com o auxílio de um algoritmo que faz uma estimativa na amplitude média do sinal, neste caso, os 100ms iniciais e os 30ms finais são considerados como ruídos de fundo.

2.1.4.1.2. Extração das características do sinal da fala

Para fazer a extração das características foram utilizados os métodos de Banco de filtros, transformada rápida de Fourier e a de codificação por predição linear, entretanto estas apresentaram algumas restrições com o trato vocal, cujo problema foi resolvido com o método cepstrum. Os coeficientes mel cepstrais são obtidos em freqüência na escala mel. Os coeficientes mel cepstrais são obtidos após alguns processamentos:

- Aplicação do banco de filtro triangular em escala mel e cálculo do logaritmo da energia de saída de cada filtro;
- Cálculo da transformada discreta inversa do co-seno do vetor do logaritmo da energia de saída do banco de filtros.

2.1.4.1.3. Treinamento e teste

O treinamento foi feito utilizando uma rede neural do tipo SOM (Self Organizing Map). Estas redes foram idealizadas por Kohonen. São baseadas no aprendizado competitivo, onde os neurônios competem entre si e o neurônio vencedor atualiza seus pesos. Estas redes mapeiam o conjunto de entrada em um mapa topográfico, unidimensional ou bidimensional.

Os neurônios são ativados seletivamente de acordo com os padrões de entrada durante o processo de aprendizado. Esta ativação é feita baseada na medida da distância euclidiana, de forma que o estado de ativação de um neurônio é determinado pela distância entre seu peso e o vetor de entrada.

Os testes foram implementados com o auxílio da ferramenta Matlab; foi utilizado um grupo contendo as vogais e a construção de 10 frases. Os resultados foram executados levando em consideração normalização e dimensão de um mapa topográfico. Os melhores resultados obtidos ficaram aproximadamente em 86%.

3. METODOLOGIA

3.1. Fluxo macro da solução

Para o sistema chegar à conclusão de que uma criança pronunciou “A” por exemplo, há um longo caminho a ser percorrido, pois existem processos pré-determinados para que se chegue ao resultado esperado.

O projeto J-Som é baseado no conceito de Reconhecimento Automático da Fala (RAF) [8], o qual consiste em três fases: aquisição do sinal da fala, extração de parâmetros e reconhecimento do padrão.

A aquisição do sinal da fala é feita utilizando um transdutor, repassando o sinal a uma interface analógica/digital [2], que faz a decodificação de um sinal analógico, na entrada da informação, em formato digital, na saída do processamento.

A extração de parâmetros é o pré-processamento, o qual extrai do sinal capturado as características que descrevem adequadamente o sinal de voz.

O reconhecimento do padrão consiste em comparar os dados extraídos na fase de pré-processamento com os padrões armazenados anteriormente, medindo a similaridade entre eles escolhendo o padrão que melhor representa o sinal da fala.

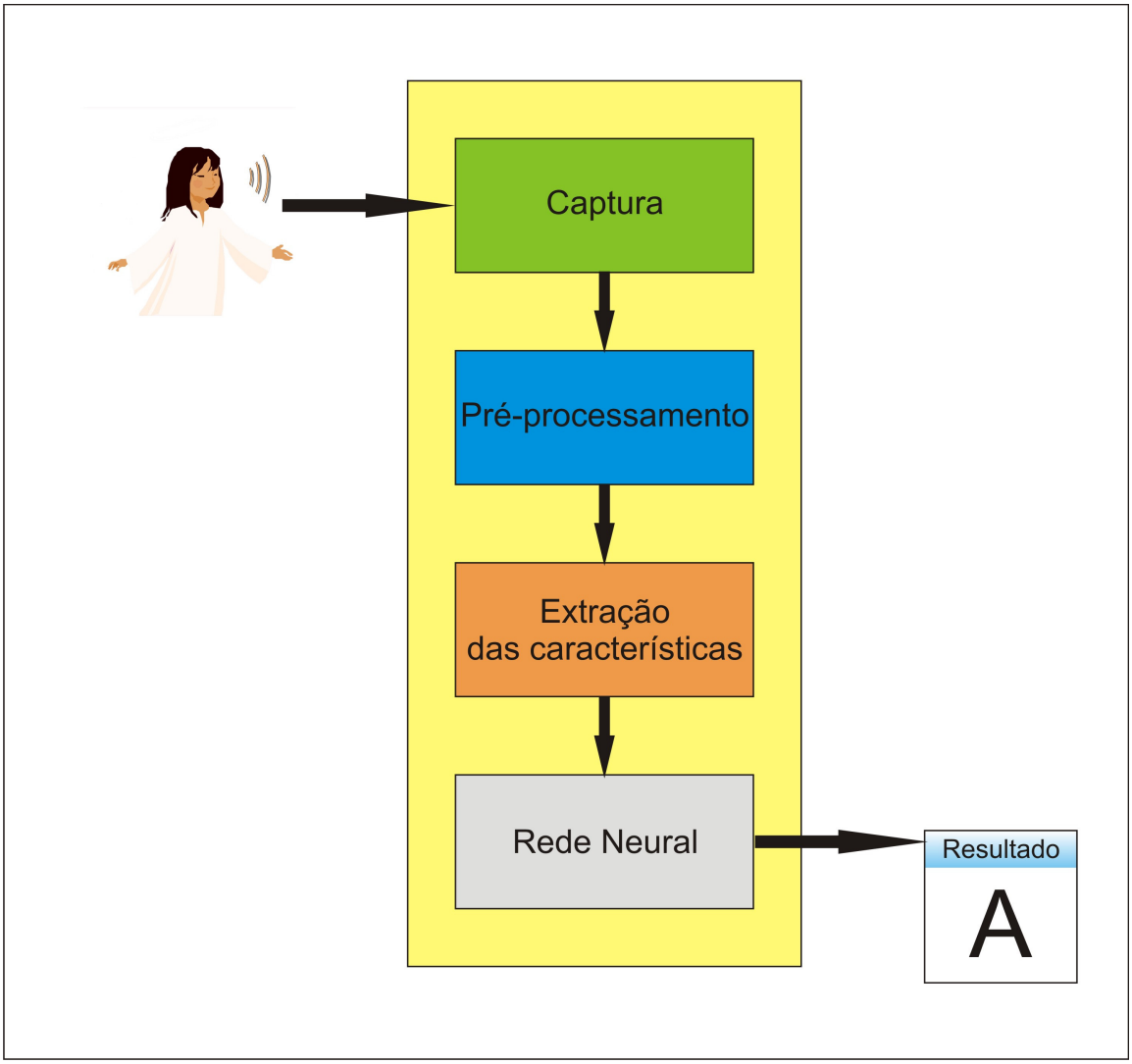


FIGURA 1 – PROCESSOS UTILIZADOS NO J-SOM

A figura 1 descreve resumidamente quais etapas foram utilizadas no sistema J-Som para chegar ao fonema pronunciado pela criança. A etapa da aquisição é representada pelo processo de captura. A etapa de extração de parâmetros é representada pelo processo de pré-processamento e extração de características. Por último, a etapa de reconhecimento do padrão é representada pelo processo Rede Neural. Na figura 2 estão detalhados todos os processos envolvidos na classificação do fonema.

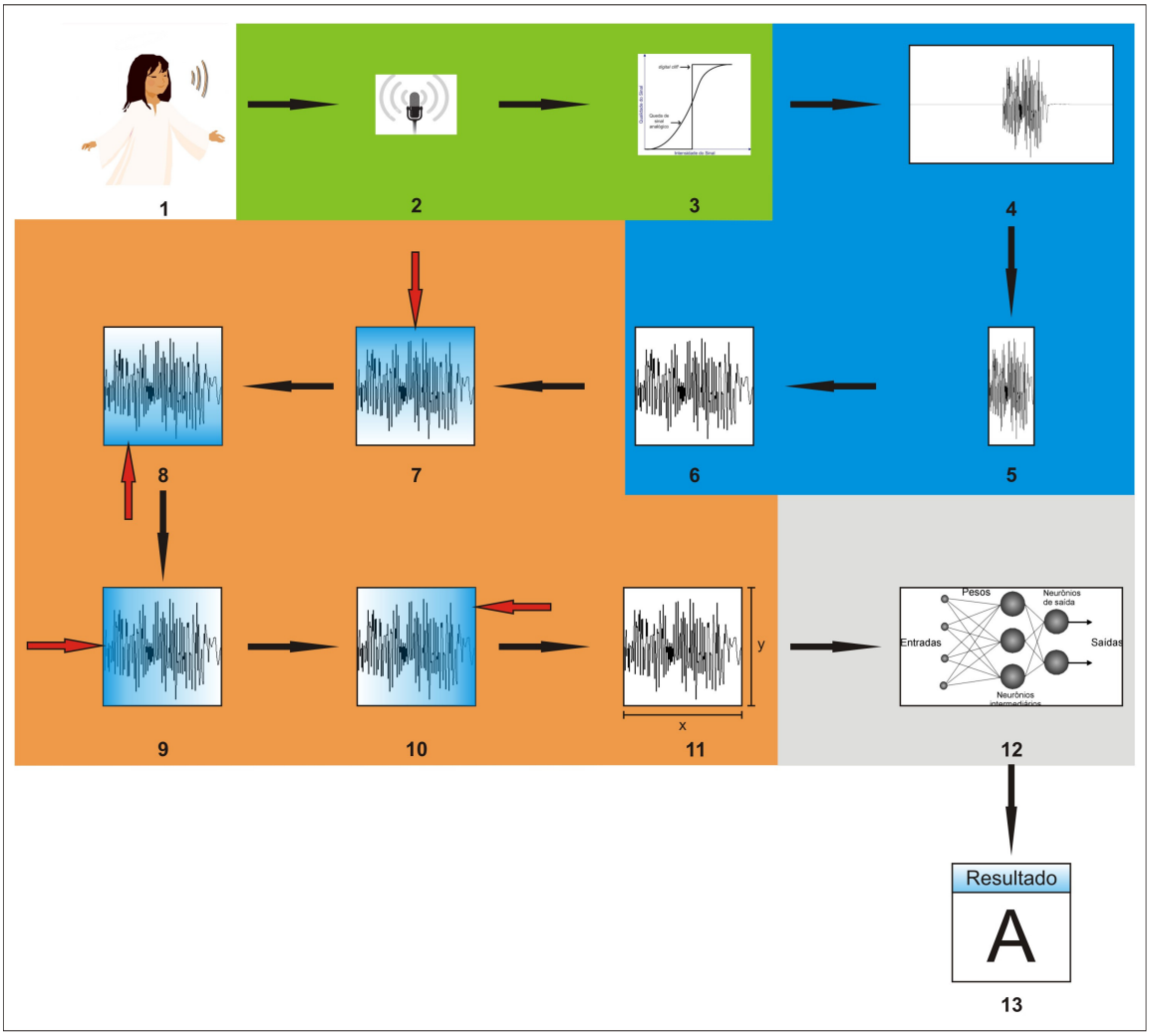


FIGURA 2 - PROCESSOS DETALHADOS

1. A criança pronuncia um determinado fonema, deste modo, inicia-se a cadeia de processos para o reconhecimento do fonema;
2. A voz da criança é capturada pelo microfone, o qual representa o transdutor, e é transformada em sinal analógico;
3. O sinal analógico passa pela interface analógica/digital, a qual é representada pela placa de som, e é transformado de sinal analógico para sinal digital, possibilitando assim a manipulação do mesmo;
4. Através do sinal digital recebido, é desenhado um gráfico no domínio do tempo para representar as amplitudes do som pronunciado. Utilizando um

exemplo real, na figura 3 temos o gráfico que representa o fonema “Mê” dito por uma criança de 6 anos;

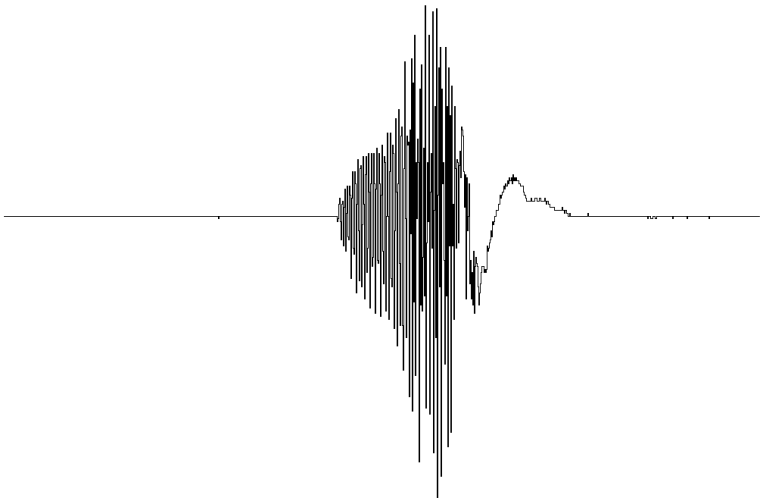


FIGURA 3 – GRÁFICO DO FONEMA MÊ

5. Com o gráfico desenhado, é aplicada a técnica de *trimming*, a qual consiste em cortar as retas sem variação de amplitude. Essas retas significam que não houve barulho algum durante aquele período, isto pode ser interpretado como silêncio do locutor;
6. Depois de retirada a parte que não interessa para a extração das características, a imagem representada pelo gráfico é redimensionada para que todos os gráficos fiquem na mesma escala. A figura 4 representa o gráfico desenhado anteriormente com a técnica de *trimming* e o redimensionamento já aplicados;

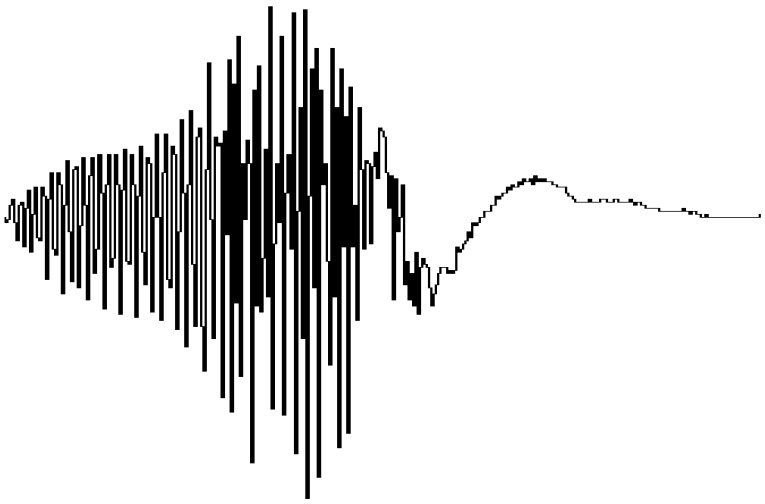


FIGURA 4 – GRÁFICO DO FONEMA MÊ COM A TÉCNICA TRIMMING E REDIMENSIONAMENTO APLICADOS

7. Com o gráfico na escala certa, outra técnica é aplicada, a qual é denominada de *Optical Character Recognition* (OCR), neste passo é aplicado o primeiro estágio da técnica (varredura de cima para baixo). A figura 5 representa o primeiro estágio da técnica OCR, onde a parte em azul indica a área em que foi aplicada a varredura. A OCR está descrita na seção Glossário;

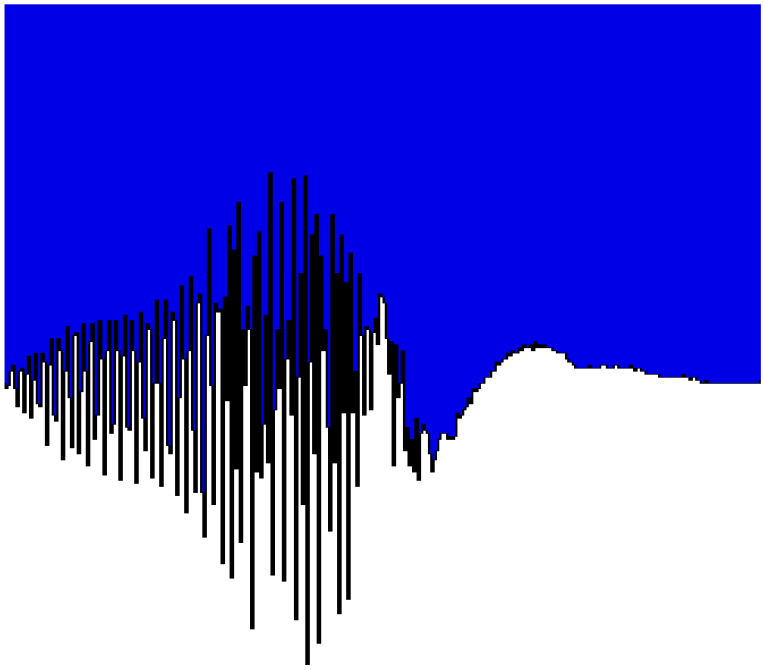


FIGURA 5 – GRÁFICO DO FONEMA MÊ COM O PRIMEIRO ESTÁGIO DA OCR

8. Neste passo é aplicado o segundo estágio da técnica OCR (varredura de baixo para cima). A figura 6 representa o segundo estágio;

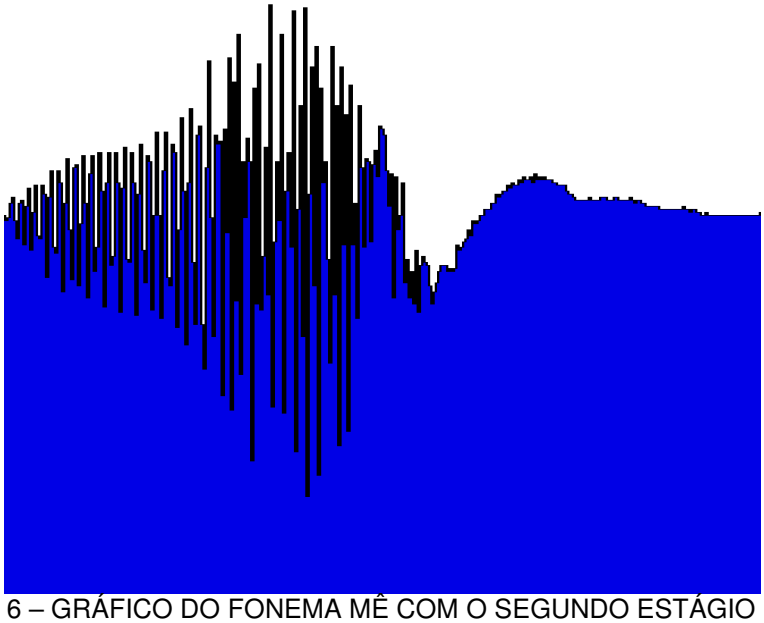


FIGURA 6 – GRÁFICO DO FONEMA MÊ COM O SEGUNDO ESTÁGIO DA OCR

9. Neste passo é aplicado o terceiro estágio da técnica OCR (varredura da esquerda para direita). A figura 7 representa o terceiro estágio;

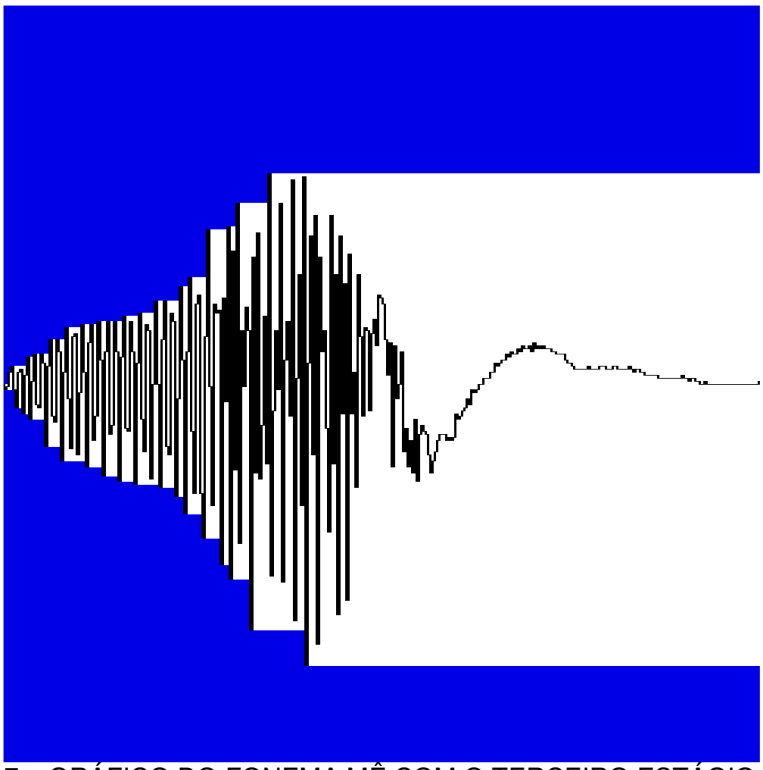


FIGURA 7 – GRÁFICO DO FONEMA MÊ COM O TERCEIRO ESTÁGIO DA OCR

10. Neste passo é aplicado o quarto estágio da técnica OCR(varredura da direita para esquerda). A figura 8 representa o terceiro estágio;

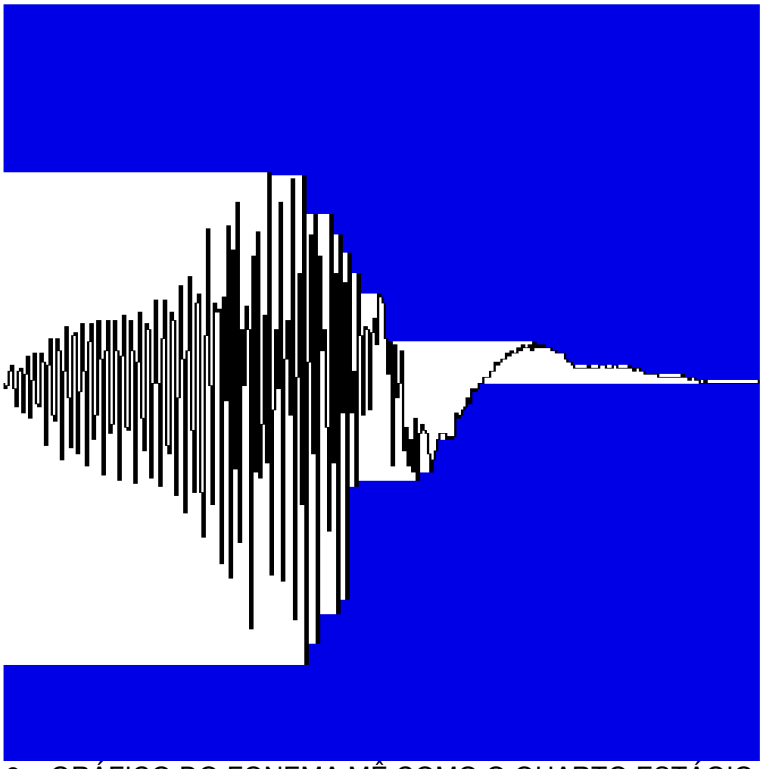


FIGURA 8 – GRÁFICO DO FONEMA MÊ COMO O QUARTO ESTÁGIO DA OCR

11. Após a geração das 4 primeiras características, é gerada a quinta – e última – característica, a qual consiste na divisão da altura pela largura do gráfico;
12. Com as 5 características retiradas do gráfico, é hora de enviar à rede neural para a mesma fazer o reconhecimento do padrão.
13. Finalmente o resultado final, a rede neural retorna a qual classe o gráfico analisado pertence.

3.2. Ferramentas e Tecnologias

Para construir o projeto J-Som foi necessário o uso de algumas ferramentas para o gerenciamento dos processos e do próprio projeto, também o uso de algumas tecnologias para a implementação do mesmo. Segue a lista de ferramentas e tecnologias utilizadas nesse projeto:

3.2.1. Ferramentas

3.2.1.1. Auxiliares na produção da documentação

Microsoft Office Word: utilizado para formatar os documentos do projeto. A versão utilizada é a 2003, ou superior. A utilização deste *software* é feita sobre uma licença adquirida particularmente, mas há alternativas gratuitas para o uso do mesmo, uma delas é o BR Office, o qual pode ser acessado através do link http://www.broffice.org/.

Microsoft Office Excel: utilizado como auxiliar na formatação dos documentos do projeto. A versão utilizada é a 2003, ou superior. A utilização deste *software* é feita sobre uma licença adquirida particularmente, mas há alternativas gratuitas para o uso do mesmo, uma delas é o BR Office, o qual pode ser acessado através do link http://www.broffice.org/.

Microsoft PowerPoint: utilizado como auxiliar na criação da apresentação do projeto. A versão utilizada é a 2003, ou superior. A utilização deste *software* é feita sobre uma licença adquirida particularmente, mas há alternativas gratuitas para o uso do mesmo, uma delas é o BR Office, o qual pode ser acessado através do link http://www.broffice.org/.

3.2.1.2. Desenvolvimento de diagramas

Jude/Community: utilizado para fazer os diagramas de UML* do projeto, bem como atualizações nos mesmo. A versão utilizada é a 5.2.1, a mais estável disponível no site até o momento. A utilização do mesmo é feita sobre licença livre, e pode ser adquirido através do link http://jude.change-vision.com/jude-web/index.html, é necessário um cadastro no site para a aquisição.

FreeMind: utilizado para converter em diagrama os conceitos extraídos através de *brainstorms*[*] realizados em reuniões do projeto. A versão utilizada é a 0.8.1. A utilização do mesmo é feita sobre licença livre, e pode ser adquirido através do link http://freemind.sourceforge.net/wiki/index.php/Download.

Gantt Project: utilizado para fazer e monitorar as atividades e as pessoas envolvidas no projeto, em relação à alocação das pessoas e controle de execução do mesmo. A versão utilizada é a 2.0.6 e a utilização do mesmo é feita sobre licença livre, e pode ser adquirido através do link http://ganttproject.biz/download.php.

3.2.1.3. Utilitários

Sound Forge: utilizado para manipular dados em formato "WAV" vindos da captação externa para amostra de dados, ele será responsável em gerar os gráficos para análise das características do arquivo de som. A versão utilizada é a 9.0 e a utilização deste *software* é feita sobre uma licença adquirida particularmente, mas a versão de demonstração pode ser adquirida através do link http://baixaki.ig.com.br/download/Sony-Sound-Forge.htm.

Free Sound Recorder: utilizado na gravação dos fonemas para a extração de características, treino e testes da rede neuronal. Estas gravações foram feitas nas escolas que se disponibilizaram a nos ajudar, fornecendo assim um tempo de suas aulas para que pudéssemos realizar a gravação das vozes. O link para o software é http://baixaki.ig.com.br/download/Free-Sound-Recorder.htm

Cool Record Edit Pro: utilizado na edição dos sons gravados. Esse software vem junto com o Free Sound Recorder e sua licença permite somente testes.

Adobe® Audition™: utilizado na edição dos sons gravados. A versão utilizada é a 3.0 e a utilização deste *software* é feita sobre uma licença adquirida

particularmente, mas a versão de demonstração pode ser adquirida através do link http://baixaki.ig.com.br/downloads/adobe-audition.htm

3.2.1.4. Integrated Development Environment (IDE)

Eclipse: IDE escolhida para ser utilizada na implementação do código, pois em comum acordo entre os integrantes da equipe é a melhor IDE a ser utilizada com a linguagem JAVA atualmente. A versão escolhida será a PDT all-in-one, pois é a que suporta maiores recursos até o momento. O link para download é http://www.eclipse.org/downloads/download.php?file=/tools/pdt/downloads/drops/R20080103/all-in-one/pdt-all-in-one-R20080103-win32.zip. Este é um projeto da Eclipse.org que visa integrar várias linguagens de desenvolvimento em uma IDE somente.

3.2.1.5. Centralização de informações

Yahoo Grupos: Ferramenta do site Yahoo para centralizar as informações e os documentos. Esta ferramenta é uma interface direta entre os integrantes do projeto, onde foram tiradas dúvidas, marcadas reuniões, armazenados os documentos desenvolvidos e compartilhadas informações importantes para o conhecimento de todos. O nome do grupo é tccjason, e a forma de comunicação com o mesmo é através do e-mail tccjason@yahoogrupos.com.br. Através do link http://br.groups.yahoo.com/group/tccjason/?v=1&t=search&ch=web&pub=groups&sec=group&slk=1 pode-se visitar a página principal do projeto, mas a associação ao grupo é feita somente com a solicitação para os moderadores do mesmo.

***Moodle*:** Ferramenta utilizada para fazer uma interface entre os integrantes do grupo e os orientadores do projeto. Nela houve sessões para os orientadores postarem informações importantes, como requisição de documentos, por exemplo,

experimentos realizados no projeto e uma ferramenta wiki* para a descrição do desenvolvimento do projeto. O link para acesso é http://www.tsi.ufpr.br/course/view.php?id=41, é necessário ser registrado para ter acesso a tal ferramenta.

3.2.1.6. Centralização de código

Current Version System (CVS): para centralizar o código fonte do projeto, foi utilizado o *software* CVS, na sua versão 1.11.21. Esse *software* foi escolhido por ser de licença livre e não consumir muita memória* do servidor*. Ele pode ser encontrado para *download* através do *link*: http://download.savannah.gnu.org/releases/cvs/binary/stable/x86-woe/cvs-1-11-22.zip.

Hamachi: é um programa de computador que simula uma Rede local, ou LAN (Local Area Network), permitindo que pessoas à distância tenham acesso ao computador ou computadores alheios, principalmente para jogos, como se estivessem realmente ligados em LAN. Link http://baixaki.ig.com.br/download/hamachi.htm

3.2.2. Tecnologias

Java: Linguagem escolhida para o desenvolvimento do código do projeto por ter sido utilizada para os trabalhos acadêmicos feitos anteriormente. Todo o processamento de negócio e comunicação com a rede neural é feito com JAVA. Esta linguagem além de ser livre, possui uma série de vantagens, as quais podem ser verificadas no *site** da SUN* através do *link** http://java.sun.com/javase/downloads/index_jdk5.jsp (em inglês). A distribuição escolhida foi a J2SE, na sua versão 5.0, pois em reunião com o grupo, foi decidido

que esta distribuição atende às necessidades do projeto e esta versão é a mais estável para usar neste momento. O link para *download** é http://java.sun.com/javase/downloads/index_jdk5.jsp.

Adobe® AIR™: O runtime do Adobe AIR [9] permite que desenvolvedores usem tecnologias comprovadas da Web para criar aplicações ricas para internet para implantação no *desktop* e execução em sistemas operacionais. É uma extensão do FLEX ™ para *desktop*, o qual é uma linguagem script desenvolvida pela Adobe para combinar Action Script com solução prática e fácil para desenvolver aplicações. A camada de visão do J-Som é desenvolvida em *Flex* rodando com a *runtime* do AIR. A licença para esta tecnologia é para meios acadêmicos, a qual foi conseguida através de pedido ao instituto Adobe mediante a retratação da carteirinha do curso. O link para *download** é http://get.adobe.com/air/?promoid=BUIGQ.

JFAN: API escrita em Java desenvolvida por alunos do curso de graduação em Tecnologia em Informática da UFPR. A JFan é uma API que possibilita a consulta a uma rede de Free Associative Neurons (FAN), baseada na tese de Roberto Tadeu Raittz [11], que tenha sido previamente treinada, com isso a integração de redes neurais em projetos de *software* ocorre de maneira mais rápida e fácil. Os arquivos .enn gerados pelo Easy Fan são utilizados pela API para a classificação das características enviadas pelo *software* Java. No J-Som o JFAN é utilizado para o reconhecimento dos fonemas. A versão utilizada é a 2.0 e essa ferramenta é de licença livre para uso dentro da Universidade Federal do Paraná.

3.3. Introdução a redes neuronais

3.3.1. Conceito

As redes neurais artificiais [13] foram desenvolvidas na década de 40 pelo neurofisiologista Warren McCulloch, do MIT e pelo matemático Walter Pitts, da Universidade de Illinois [1]. Esses pesquisadores criaram um sistema baseado no funcionamento das células nervosas. A partir da década de 80, com avanço da tecnologia várias redes neurais artificiais foram criadas aperfeiçoando cada vez mais este mecanismo.

As redes neurais são formadas por um conjunto de vários neurônios. Alguns dispositivos fazem a “entrada” dos dados em uma área que simula a captação de estímulos de um sistema nervoso, estas entradas fazem a comunicação com os neurônios intermediários, os quais fazem a comunicação com os neurônios de saídas. Estas comunicações representam, em comparação com um neurônio biológico, o contato dos dendritos com outros neurônios.

A estrutura de uma rede neural pode variar, o projetista pode manipular a sua arquitetura e achar uma estrutura ideal para a aplicação da sua rede, levando em consideração que a construção da rede está sendo preparada para a resolução de determinados problemas. Basicamente a estrutura que compõem uma RNA e que pode sofrer modificações, como já citado, são estas:

- Conexões entre camadas;
- Camadas intermediárias;
- Quantidade de neurônios;
- Função de transferência;
- Algoritmo de aprendizado.

As RNAs “aprendem” a resolver problemas através de exemplos que são apresentados para a rede. A rede utiliza padrões, regularidades e correlações para agrupar os conjuntos de dados em classes. Este processo é chamado de treinamento de rede. Na fase de treinamento os neurônios competem entre si para estabelecer um padrão e armazenar determinadas informações, isto é feito através de um sistema de ajuste dos pesos.

3.3.1.1. Neurônio Artificial

No neurônio artificial os dendritos são substituídos por entradas, as entradas recebem pesos, que substituem as sinapses. Uma “função de soma” processa os estímulos captados pelas entradas.

4. RECURSOS

4.1. Humanos

4.1.1. Equipe

TABELA 2 – INFORMAÇÕES DA EQUIPE

Nome	Descrição	Responsabilidades
Roberto Tadeu Raittz - Orientador	Professor orientador do projeto; especialista em Inteligência Artificial; já trabalhou com reconhecimento de voz, de imagens e atualmente realiza pesquisas na Bioquímica.	Tem como principal objetivo orientar a equipe no desenvolvimento do projeto com suas experiências na área de sua abrangência, especificamente a respeito das redes neurais.
Dieval Guizelini – Orientador	Professor orientador do projeto; especialista em Banco de Dados e mineração de dados, entre outras diversas.	Com seu grande conhecimento técnico em sistemas diversos, tem como objetivo orientar a equipe do projeto com técnicas e estruturas de desenvolvimento e documentação, auxiliando o orientador.
Bruno Côrtes Lopes Siqueira – Equipe do projeto	Autor da idéia do projeto; experiência em sistemas *web* com linguagem Java e sistemas bancários com linguagem COBOL.	Responsável por expor toda a idéia, além de manter contato com especialistas externos diretamente ligados às áreas de direta influência (Psicologia, Pedagogia e Fonoaudióloga); desenvolvimento técnico da interface do projeto.
Guilherme Maciel da Silva – Equipe do projeto	Experiência em sistemas com linguagem Java.	Responsável pelo estudo dos gráficos de áudio e das redes neurais para contribuir diretamente no desenvolvimento técnico do projeto.
Moises Pereira Oliveira – Equipe do projeto	Experiência acadêmica com linguagem Java.	Responsável pelo estudo de projetos de conclusão de curso que abrangeram a mesma área do projeto em questão e das estruturas de documentação adotadas pelos mesmos; desenvolvimento técnico do projeto.
Wesley Calesso Arantes – Equipe do projeto	Experiência em sistemas *web* com linguagem Java.	Responsável por implantar a metodologia de gerenciamento de projetos e estruturas de documentação; desenvolvimento técnico do projeto.
Willian Baierle de Oliveira – Equipe do projeto	Experiência em sistemas *web* com linguagem Java e sistemas bancários com linguagem COBOL.	Responsável pela implantação da ferramenta para versionamento de codificação e pelo estudo das redes neurais; desenvolvimento técnico do projeto.
Cleusa Pierin –	CRFa 5518/PR	Responsável por guiar a equipe de projeto através das reais necessidades da

Fonoaudióloga	16 anos de experiência na área	Fonoaudiologia que serão atendidas com o desenvolvimento do projeto.
Luciane de Souza Kowalewski – Fonoaudióloga	CRFa 5090/PR Especialista em Audiologia Clínica nº 293/97 19 anos de experiência na área	Responsável por guiar a equipe de projeto através das reais necessidades da Fonoaudiologia que serão atendidas com o desenvolvimento do projeto; fornecer as sílabas que se encontram na primeira versão funcional do programa e proporcionar base teórica para justificar o uso das mesmas.

4.1.2. Usuários

TABELA 3 – INFORMAÇÕES DOS USUÁRIOS DO SISTEMA

Nome	Descrição	Responsabilidades
Professor de Educação Infantil	É a pessoa que auxiliará a criança – principal usuário – na sala de aula a utilizar o sistema.	Descreve quais formas de abordagem devem ser empregadas para prender da melhor maneira a atenção da criança. Guia a criança no uso do sistema.
Pais	São as pessoas que auxiliarão a criança a utilizar o sistema em casa.	Guiam a criança no uso do sistema.
Fonoaudiólogo	Utiliza o sistema para identificar crianças com problemas na fala.	Descreve que pontos deverão ser abordados para identificar possíveis problemas de fala e que espécie de dados deve ser utilizada no sistema.
Aluno de Educação Infantil	Usuário-alvo do sistema.	Utiliza o sistema para melhorar sua dicção na reprodução dos fonemas através de sílabas.

4.2. Materiais

TABELA 4 – RECURSOS MATERIAIS

Tipo	Notebook	Notebook	Notebook	Desktop	Desktop	Desktop
RAM	2 Gb.	2 Gb	4 Gb	1.5 Gb	1 Gb	1 Gb
Tipo Processador	Turion 64 X2	Core 2 Duo	Core 2 Duo	Sempron 3000+	Pentium D	Athlon 64 3200
Velocidade Processador	1.6 GHz	1.6 GHz	2.0 GHz	1.8 GHz	2.6 GHz	2.2 GHz
Caixas Som	Integrada	Integrada	Integrada	Satelite 2.0	Satelite 2.0	Satelite 5.1

Sensibilidade Microfone	2 a 58dB	2 a 58dB	2 a 58dB	2 a 58dB	2 a 58dB	2 a 58dB
Sistema Operacional	Windows XP	Windows XP	Windows XP	Windows XP	Windows XP	Windows XP

5. O SISTEMA J-SOM

5.1.1. Riscos

5.1.1.1. Diagrama de Ishikawa

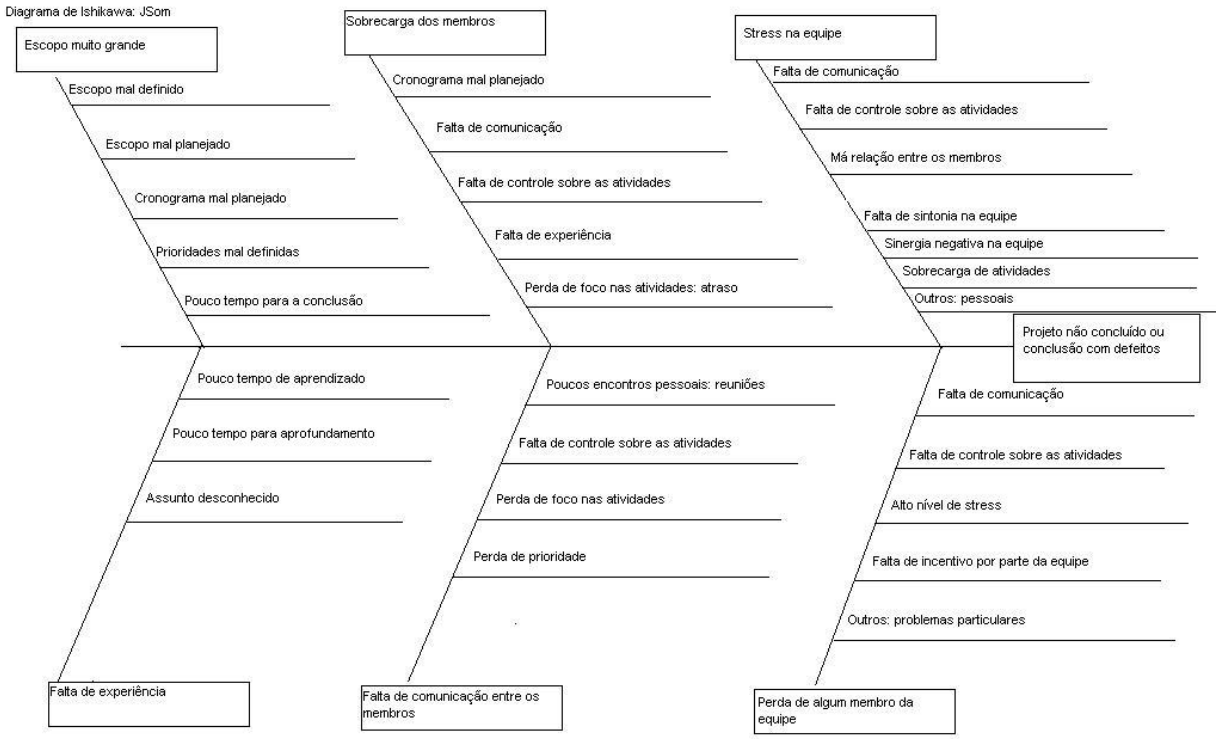


DIAGRAMA 1 – DIAGRAMA DE ISHIKAWA

5.1.1.2. Escopo complexo demais para o projeto.

Gravidade do risco

Prejudicial

Descrição

O sistema pode chegar a sua fase final sem executar funcionalidades antes previstas no escopo inicial.

Impactos

- Falta de tempo para implementar todo o escopo.
- Funcionalidades defeituosas.
- Funcionalidades previstas não implementadas.

Estratégia de diminuição

- Divisão do escopo de acordo com as prioridades, em “escopo primário” com funcionalidades que deverão ser implementadas sem exceção e “escopo secundário” com funcionalidades que poderão ser implementadas de acordo com o andamento do projeto e o tempo disponível para a conclusão do mesmo.

Plano de Contingência

- Nenhum

5.1.1.3. Falta de experiência da equipe

Gravidade do risco.

O mais prejudicial.

Descrição

A equipe não possui experiência com este tipo de projeto

Impactos

- Dificuldade para implementar as funcionalidades descritas no escopo.
- Falta de tempo para a equipe se especializar na tecnologia utilizada.
- Dificuldade em integrar tecnologias que serão usadas no sistema.

Estratégia de diminuição.

- Buscar ajuda de pessoas que conhecem a tecnologia utilizada no projeto, tais pessoas podem ser: orientadores, professores, profissionais da área, alunos e ex-alunos etc.

- Pesquisar em materiais que possam trazer explicações sobre o assunto utilizando-se dos mais variados meios como Documentos de TCC concluídos, revistas, livros, artigos e Internet.

Plano de contingência

- Quanto maior a complexidade da tecnologia mais tempo e atenção devem ser disponibilizadas para tais. Revisão do cronograma com diminuição de tempo de atividades menos complexa.
- Agendar reuniões com pessoas que dominem a tecnologia utilizada.
- Pesquisar materiais externos.

5.1.1.4. Perda de algum membro da equipe.

Gravidade do risco

Incalculável. A gravidade pode variar de acordo com os conhecimentos e participação dos membros da equipe assim como a data de saída de um integrante de acordo com as fases do projeto.

Descrição

Um integrante deixa de fazer parte da equipe por algum motivo em particular ou por decisão majoritária da equipe.

Impactos

- Sobrecarga de atividades.
- Impacto sobre o cronograma.

Obs.: Os impactos podem variar de acordo com os conhecimentos e participação dos membros da equipe assim como a data de saída de um integrante de acordo com as fases do projeto.

Estratégia de diminuição

- Manter uma boa relação entre os membros da equipe.
- Manter uma boa comunicação entre os membros da equipe.

- Efetuar reuniões capazes de mensurar o nível de satisfação e de stress de cada integrante da equipe.
- Manter o controle sobre as atividades, evitando que algum membro da equipe fique com sobrecarga de atividades.
- Manter um nível de estímulo e energia positiva para impedir que membros da equipe possam se desestimular pelo projeto em si.

Plano de contingência

- Revisar o cronograma para reajustar as atividades.
- Redistribuir as atividades antes designadas ao “membro desligado”.

5.1.1.5. Falta de comunicação entre os membros.

Gravidade do risco.

O mais prejudicial.

Descrição

Pode ocorrer falta de entendimento ou divergências entre os membros da equipe.

Impactos

- Indecisão para distribuir as atividades.
- Indecisão para priorizar as atividades.
- Indefinição sobre o que deve ser feito, qual caminho deve ser tomado.
- Atraso no cronograma.
- Sobrecarga de atividades para algum membro da equipe.
- Aumentar o nível de stress dos membros da equipe.
- Diminuir a sintonia e a energia positiva da equipe.
- Desinteresse pelo projeto.
- Perda de algum membro da equipe.

Estratégia de diminuição

- Escolha de um líder capaz de estimular a equipe e manter a equipe focada no projeto.
- Manter um controle sobre as atividades.
- Manter uma sintonia e uma energia positiva entre os membros da equipe.
- Manter uma boa relação entre os membros da equipe.
- Manter uma boa comunicação entre os membros da equipe.
- Efetuar reuniões capazes de mensurar o nível de satisfação e de stress de cada integrante da equipe.

Plano de contingência

- Efetuar reunião para manter o controle das atividades.
- Procurar ajuda de profissionais que possam auxiliar no estado psicológico da equipe.

5.1.1.6. Stress na equipe.

Gravidade do risco

O mais prejudicial.

Descrição

Um alto nível de stress pode atingir um ou mais membros da equipe.

Impactos

- Incapacidade psicológica para efetuar alguma atividade.
- Falta de concentração.
- Falta de comunicação entre os membros da equipe.
- Desinteresse pelo projeto.
- Diminuir a sintonia e a energia positiva da equipe.
- Atraso das atividades.
- Atraso no cronograma.

- Perda de algum membro da equipe.

Estratégia de diminuição

- Escolha de um líder capaz de estimular a equipe e manter a equipe focada no projeto.
- Manter um controle sobre as atividades.
- Manter uma sintonia e uma energia positiva entre os membros da equipe.
- Manter uma boa relação entre os membros da equipe.
- Manter uma boa comunicação entre os membros da equipe.
- Efetuar reuniões capazes de mensurar o nível de satisfação e de stress de cada integrante da equipe.

Plano de contingência.

- Efetuar reuniões capazes de mensurar o nível de satisfação e de stress de cada integrante da equipe.
- Procurar ajuda de profissionais que possam auxiliar a equipe psicologicamente e manter a equipe focada no projeto.

5.1.1.7. Sobrecarga dos membros.

Gravidade do risco.

Prejudicial

Descrição.

Um ou mais membro da equipe pode ficar com excesso de atividades.

Impactos

- Atraso ou não cumprimento das atividades.
- Insatisfação ou desinteresse por parte de algum integrante.
- Aumento do nível de stress de algum integrante.
- Divergência entre os membros da equipe.

Estratégia de diminuição

- Escolha de um líder capaz de distribuir as atividades com bom senso e inteligência.
- Manter o controle sobre as atividades.

Plano de contingência.

- Reestruturação do cronograma.
- Reestruturação das atividades.
- Mensurar o tempo que os membros da equipe têm disponível para suas atividades.
- Efetuar reuniões que proporcionem o status e as dificuldades para se implementar certas atividades.
- Proporcionar auxílio para o complemento das atividades.

5.2. Arquitetura do sistema

5.2.1. Descrição das camadas

O J-Som tem, além de um processamento complexo, uma interface que atende às necessidades do usuário. Para tornar mais fácil o desenvolvimento da interface e propiciar mais interatividade com o usuário resolveu-se adotar a tecnologia Adobe® Flex™ versão 3.0 [17], utilizando recursos do Adobe® Air™ [9] para executar em ambiente *desktop*. O Flex contém muitos recursos específicos para interface, possibilitando um alto nível de interatividade.

Devido ao uso de duas tecnologias diferentes, é preciso ter uma comunicação entre a camada de visão (Flex) e a camada de negócio (JAVA). Tal comunicação é realizada através de mensagens entre as camadas, no formato AMF3. Essa mensagem é assíncrona, fazendo com que cada envio de mensagem

não espere necessariamente uma resposta. Para isso o desenvolvimento deve suportar corretamente cada fluxo existente na aplicação.

5.2.2. Casos de Uso

5.2.2.1. Diagrama de casos de uso

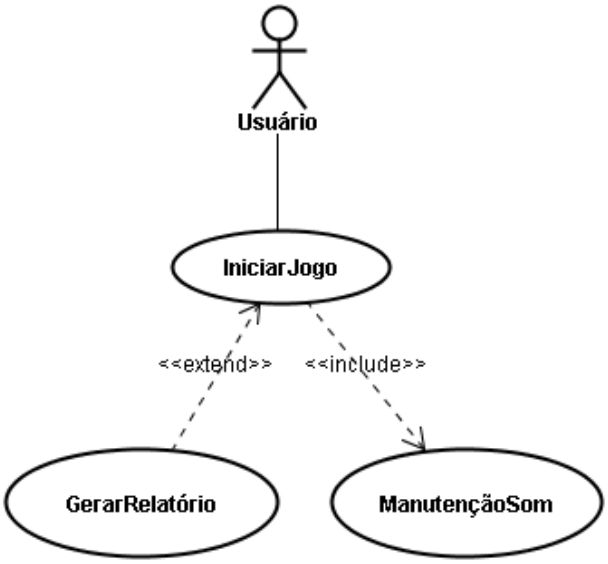


DIAGRAMA 2 – DIAGRAMA DE CASOS DE USO

5.2.2.2. UC001 – Iniciar Jogo

5.2.2.2.1. Descrição

Este Caso de Uso serve para iniciar o jogo J-SOM

5.2.2.2.2. Pré-Condições

Este Case de Uso inicia somente se:

1. Tela DV001 estiver disponível

5.2.2.2.3. Pós-Condições

1. Sistema deve ter os dados necessários para a geração do relatório.

5.2.2.2.4. Ator primário

Usuário

5.2.2.2.5. Atores secundário

Sistema

5.2.2.2.6. Fluxo Principal de eventos

1. O usuário clica no botão “Iniciar o Jogo” (A4)
2. O sistema exibe a tela DV002
3. O usuário clica no fonema desejado (A1)
4. O sistema exibe a tela DV003
5. O sistema reproduz o vídeo com o som e pronuncia do fonema escolhido
6. O Sistema exibe a tela DV004
7. O sistema solicita ao usuário para repetir o fonema reproduzido
8. O usuário repete o fonema
9. O sistema inicializa o caso de uso UC002 – Manutenção Som
10. O sistema identifica que o fonema foi repetido corretamente (A2) (R1)
11. O sistema exibe a tela DV005 com uma animação parabenizando o usuário
12. O sistema retorna à tela do sumário, DV002
13. O usuário clica em outro fonema (A1)
14. O sistema retorna ao passo 4 do fluxo principal de eventos

15. Caso de uso é finalizado

5.2.2.2.7. Fluxos Alternativos

A1. Usuário clica no botão “Sair do Jogo”

1. O sistema inicializa o caso de Uso UC003 - Gerar Relatório
2. O sistema retorna à tela DV001 e exibe a mensagem "Jogo encerrado. Para jogar novamente clique em 'Iniciar o Jogo' ".
3. Caso de Uso é finalizado.

A2. Fonema foi repetido incorretamente

1. O sistema exibe a tela DV006 com uma animação que representa a pronúncia incorreta do fonema
2. O usuário clica em “Tentar de Novo” (A3)
3. O sistema retorna ao passo 4 do fluxo principal

A3. Usuário clica no botão “Mudar de Fase”

1. Sistema retorna ao passo 3 do fluxo principal

A4. Usuário clica no botão “Desligar”

1. O sistema encerra o programa.
2. Caso de Uso é finalizado.

A5. Usuário clica no botão que representa outro fonema

1. Sistema retorna ao passo 3 do fluxo principal.

5.2.2.2.8. Regras de Negócio

R1. Critério de avaliação do fonema

Por correto, entende-se um fonema que seja exatamente igual à classe do fonema retornado pela rede neural após o processamento do som.

5.2.2.2.9. Data view

Todas as telas podem ser encontradas na seção **Anexos** deste documento.

5.2.2.3. UC002 – Manutenção do Som

5.2.2.3.1. Descrição

Este Caso de Uso serve para avaliar o som reproduzido pelo usuário.

5.2.2.3.2. Pré-Condições

Este Case de Uso inicia somente se:

1. Sistema tiver solicitado a verificação do som reproduzido pelo usuário no caso de uso UC001 – Iniciar Jogo
2. Sistema possuir os dados do fonema falado e os dados da classe desejada

5.2.2.3.3. Pós-Condições

1. Sistema deve retornar a classe do fonema da rede neural que mais se

assemelha com o reproduzido pelo usuário

5.2.2.3.4. *Ator* primário

Sistema

5.2.2.3.5. Ator secundário

Rede neural

5.2.2.3.6. Fluxo Principal de eventos

1. O sistema captura o som falado pelo usuário (R1)
2. O sistema desenha o espectrograma do som capturado (R2)
3. O sistema faz o tratamento do som (R3)
4. O sistema identifica as características do fonema(R4)
5. O sistema envia as características para a rede neural
6. O sistema retorna a classe do fonema (R5).
7. Caso de uso é finalizado

5.2.2.3.7. Fluxos Alternativos

Não aplicável

5.2.2.3.8. Regras de Negócio

R1. Captura do som

A captura do som deverá ser feita com um microfone com a sensibilidade de 2 a 58dB, o qual converterá as ondas sonoras em sinal analógico e conduzirá até a placa de som que transformará em sinal digital. Desta forma o sistema poderá manipular os dados conforme a sua necessidade.

R2. Desenho do espectrograma

O Sistema deverá desenhar o espectrograma do som utilizando os dados do sinal digital convertido anteriormente, no tamanho igual a 1000 pixels de altura por 2000 de largura, possibilitando assim a manipulação dos outros processos para identificação das características do fonema.

R3. Tratamento do som

O Sistema deverá pegar o som, já em formato digital, e aplicar uma técnica para cortar a parte representada por uma reta no espectrograma do fonema, o que, na prática, indica que não foi identificado barulho pelo microfone.

R4. Identificação das características do fonema

Sistema deverá utilizar a técnica conhecida como *Optical Character Recognition* (OCR), esta técnica está descrita na seção Glossário. Depois de aplicada, têm-se os valores de 4 características e mais outro valor retirado da divisão da altura pela largura da imagem, com isso temos as 5 características para enviar a rede neural.

R5. O sistema retorna o fonema

A classificação do fonema é feita de forma independente do sistema J-SOM, assim essa classificação é passível de falha, por se tratar de um sistema que faz utilização de redes neurais. O retorno da rede neural será a classe do fonema, o qual deverá ser comparado ao fonema selecionado pelo usuário. Caso sejam iguais deverá retornar verdadeiro, senão falso.

5.2.2.3.9. Data Views

Não aplicável

5.2.2.4. UC003 – Gerar Relatório

5.2.2.4.1. Descrição

Este Caso de Uso serve para gerar o relatório da sessão atual do usuário

5.2.2.4.2. Pré-Condições

Este Caso de Uso inicia somente se:

1. Usuário tiver clicado em “Sair do Jogo” no caso de uso UC001 – Iniciar Jogo

5.2.2.4.3. Pós-Condições

1. Sistema ter gerado um relatório com os dados da sessão jogada

5.2.2.4.4. Ator

Sistema

5.2.2.4.5. Fluxo Principal de eventos

1. Sistema recebe os dados da sessão realizada pelo usuário (R1)
2. Sistema gera um relatório em formato PDF(R2) (R3)

3. Caso de uso é finalizado

5.2.2.4.6. Regras de Negócio

R1. Validação da sessão

O sistema deve somente gerar o relatório se algum fonema foi clicado na tela de sumário. Caso não haja fonema clicado, o sistema não gera o relatório e finaliza o caso de uso.

R2. Nomeação do arquivo gerado

O sistema deve gerar um arquivo no formato PDF, relacionando os dados da sessão terminada pelo usuário. O nome do arquivo deverá seguir o padrão “dd_MM_yyyy HH_mm_ss”, onde “dd” é o dia, “MM” o mês, “yyyy” o ano, “HH” a hora, “mm” os minutos e “ss” os segundos, ambos referentes ao início do jogo. Por exemplo, um jogo iniciado no dia 26 de Novembro de 2008 às 15 horas, 42 minutos e 10 segundos, deveria ficar como “26_11_2008_15_42_10.pdf”.

R3. Caminho do arquivo gerado

O caminho padrão para o arquivamento do relatório será em C:\RelatoriosJSom\ e não há possibilidade de trocar o mesmo.

5.2.2.4.7. Data Views

Não aplicável

5.2.3. Diagrama de classes

5.2.3.1. FLEX (Camada de visão)

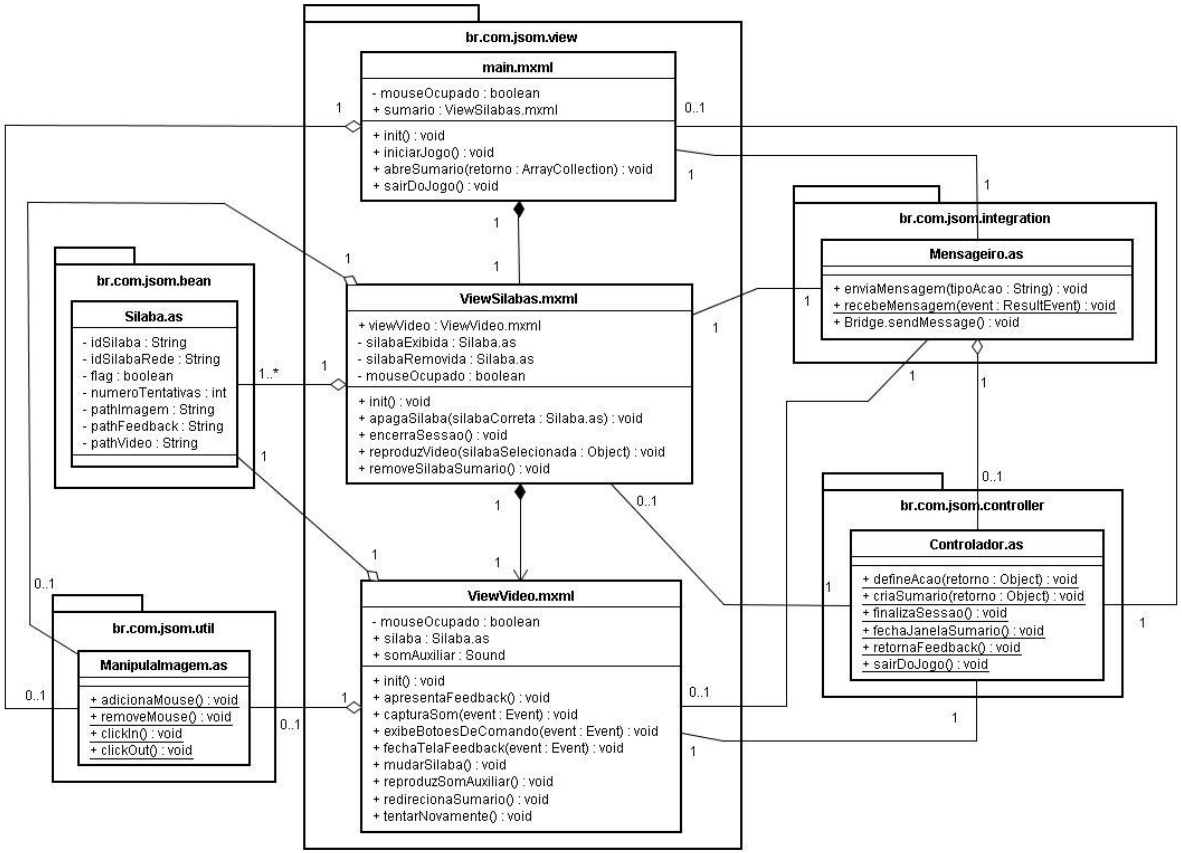


DIAGRAMA 3 – DIAGRAMA DE CLASSES FLEX

5.2.3.2. JAVA (Camada de Negócios)

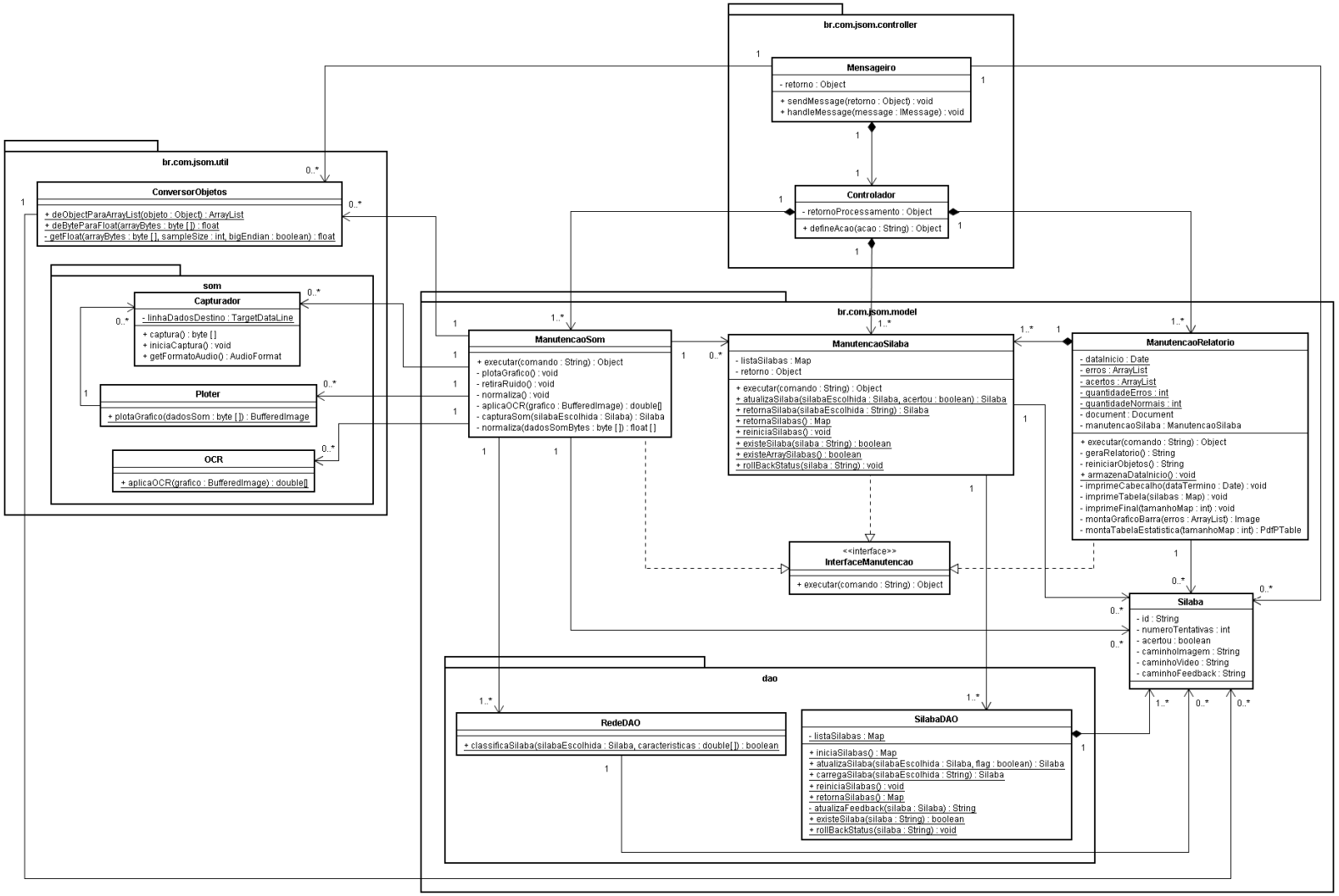


DIAGRAMA 4 – DIAGRAMA DE CLASSES JAVA

5.2.4. Diagramas de seqüência

5.2.4.1. Java - UC001 Iniciar Jogo – Fluxo Principal

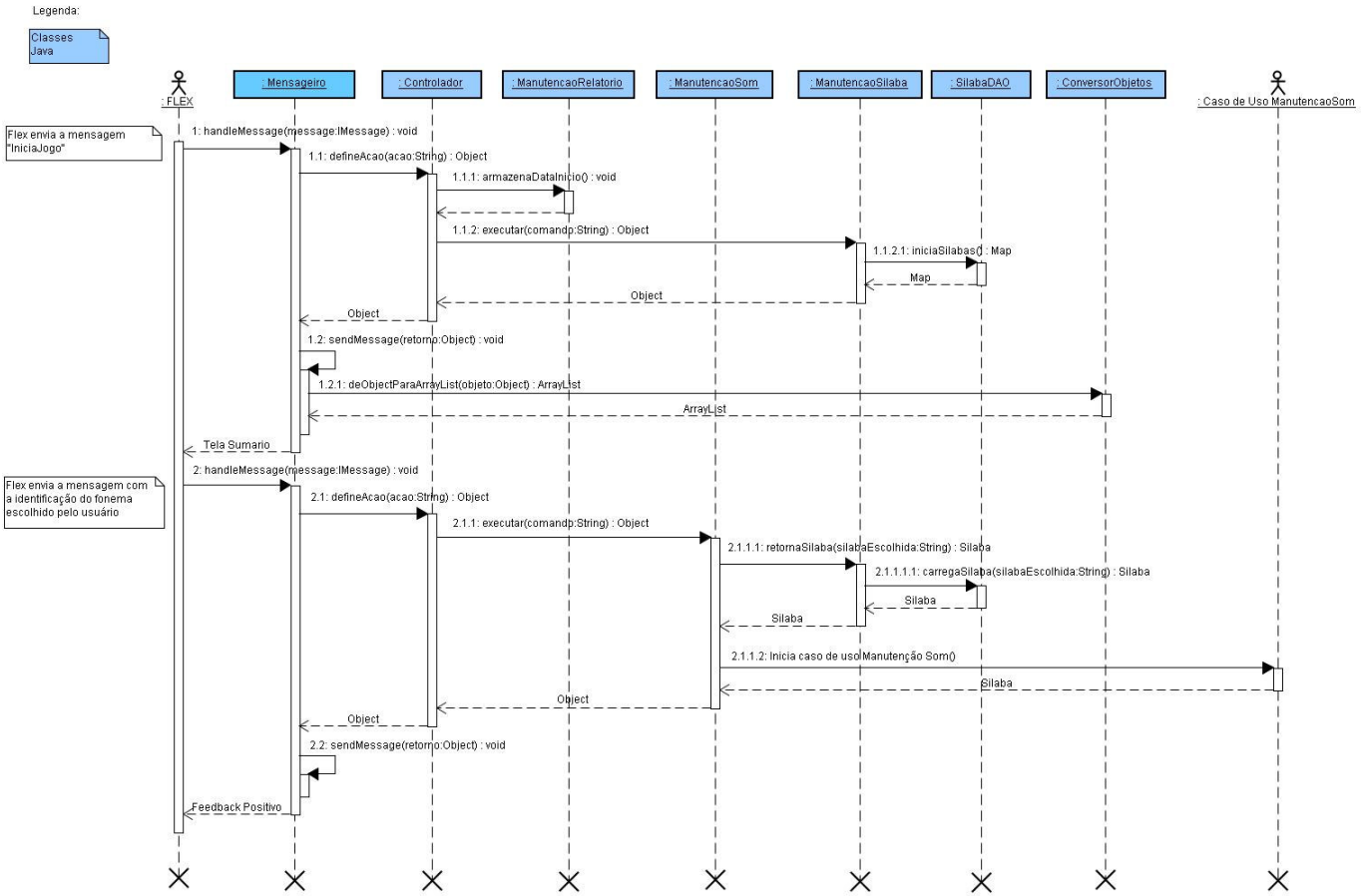


DIAGRAMA 5 – DIAGRAMA DE SEQÜÊNCIA JAVA UC001 - FLUXO PRINCIPAL

5.2.4.2. Java - UC001 Iniciar Jogo – Alternativo (A1)

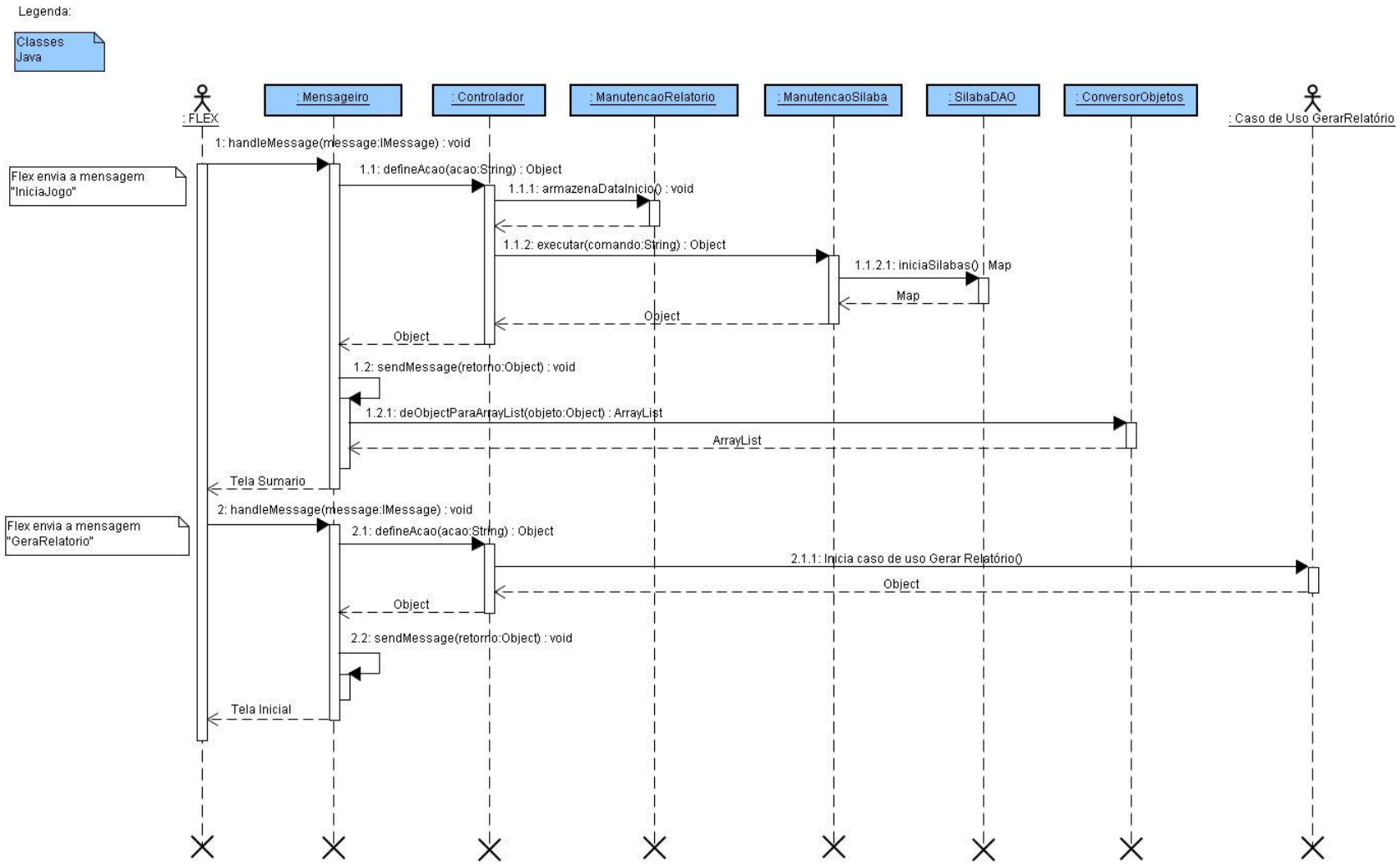


DIAGRAMA 6 – DIAGRAMA DE SEQÜÊNCIA JAVA UC001 - FLUXO (A1)

5.2.4.3. Java - UC002 Manutenção do Som – Fluxo Principal

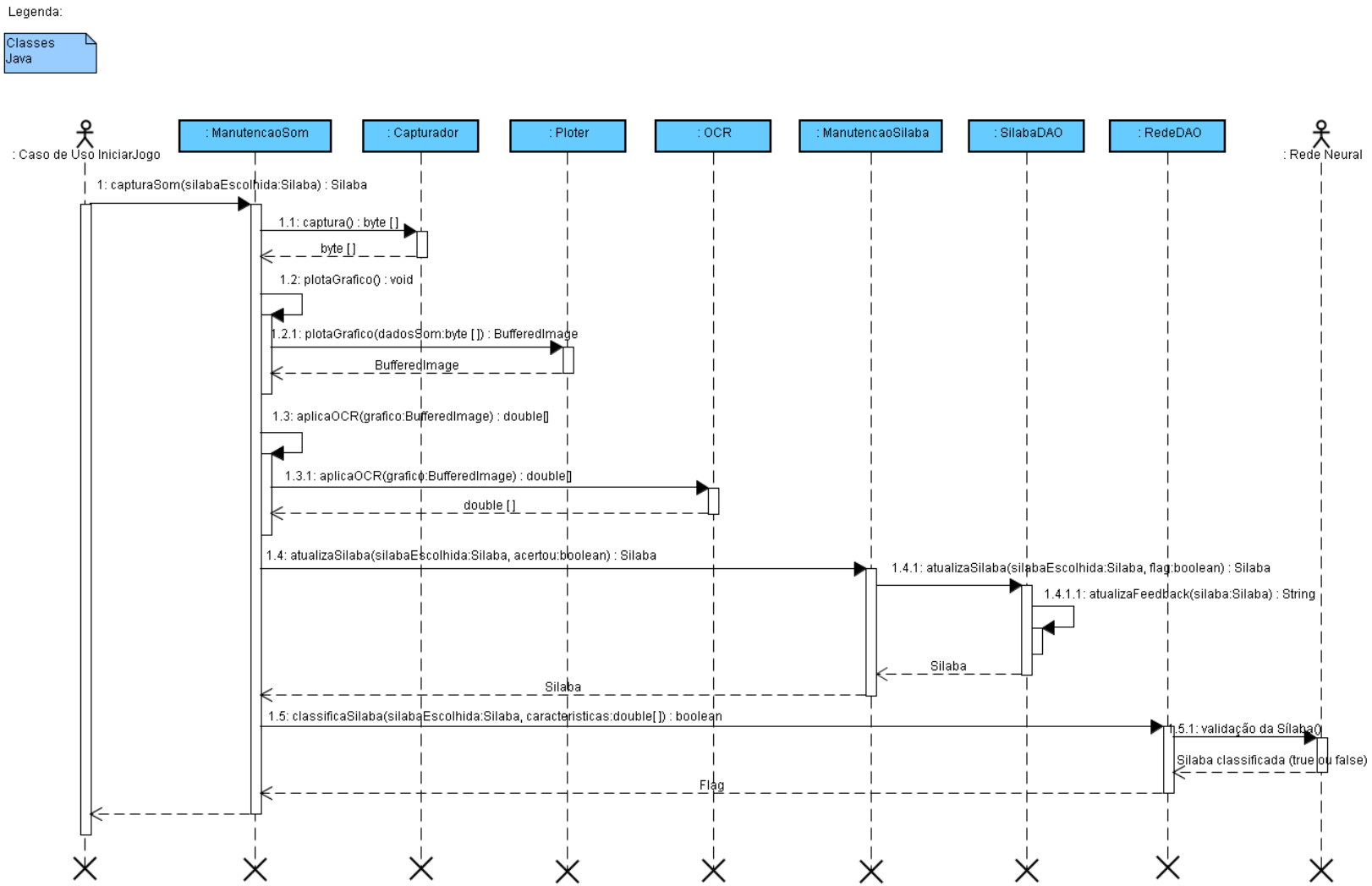


DIAGRAMA 7 – DIAGRAMA DE SEQÜÊNCIA JAVA UC002 - FLUXO PRINCIPAL

5.2.4.4. Java - UC003 Gerar Relatório – Fluxo Principal

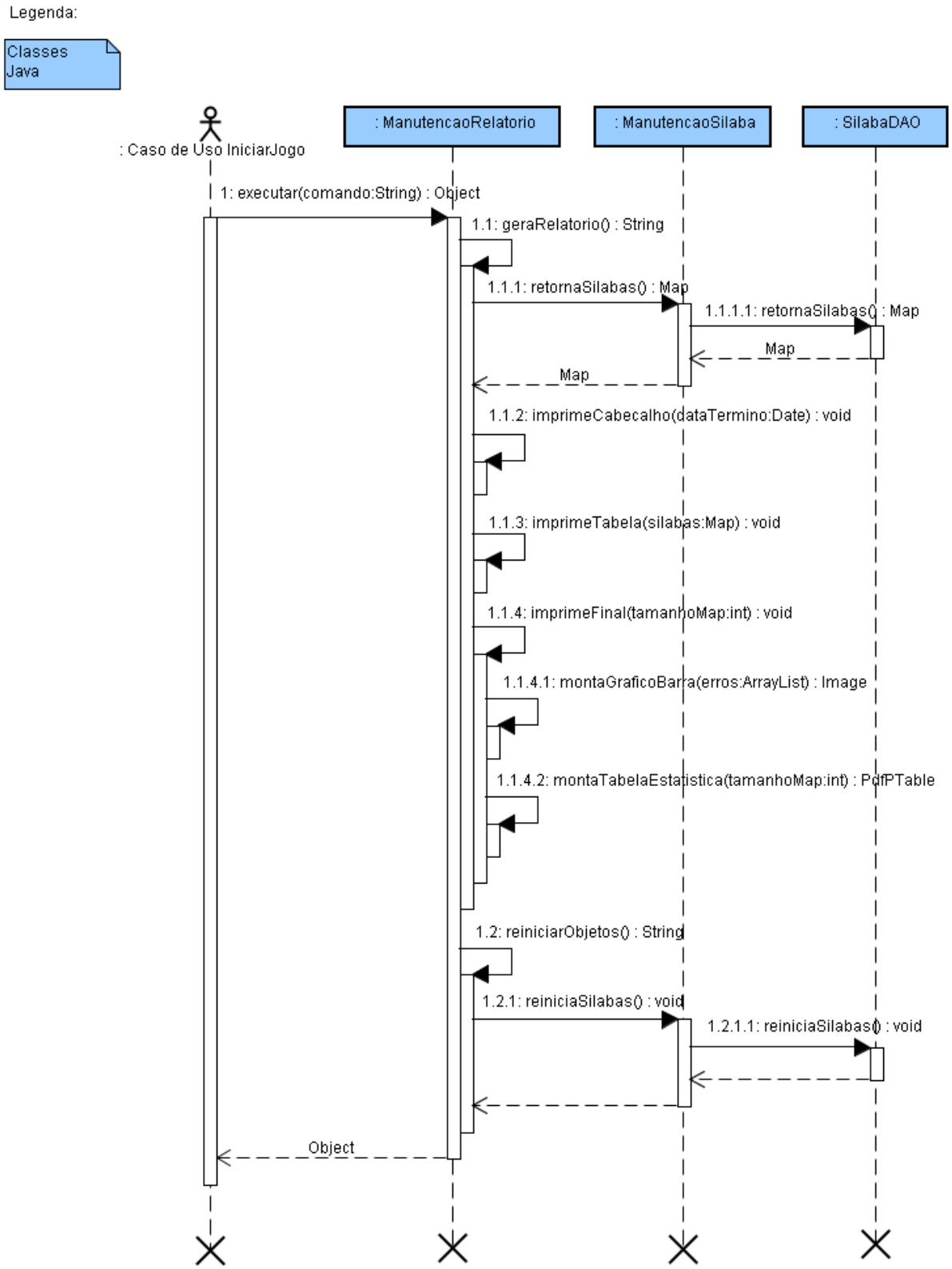


DIAGRAMA 8 – DIAGRAMA DE SEQÜÊNCIA JAVA UC003 - FLUXO PRINCIPAL

5.2.4.5. Flex - UC001 Iniciar Jogo – Fluxo Principal

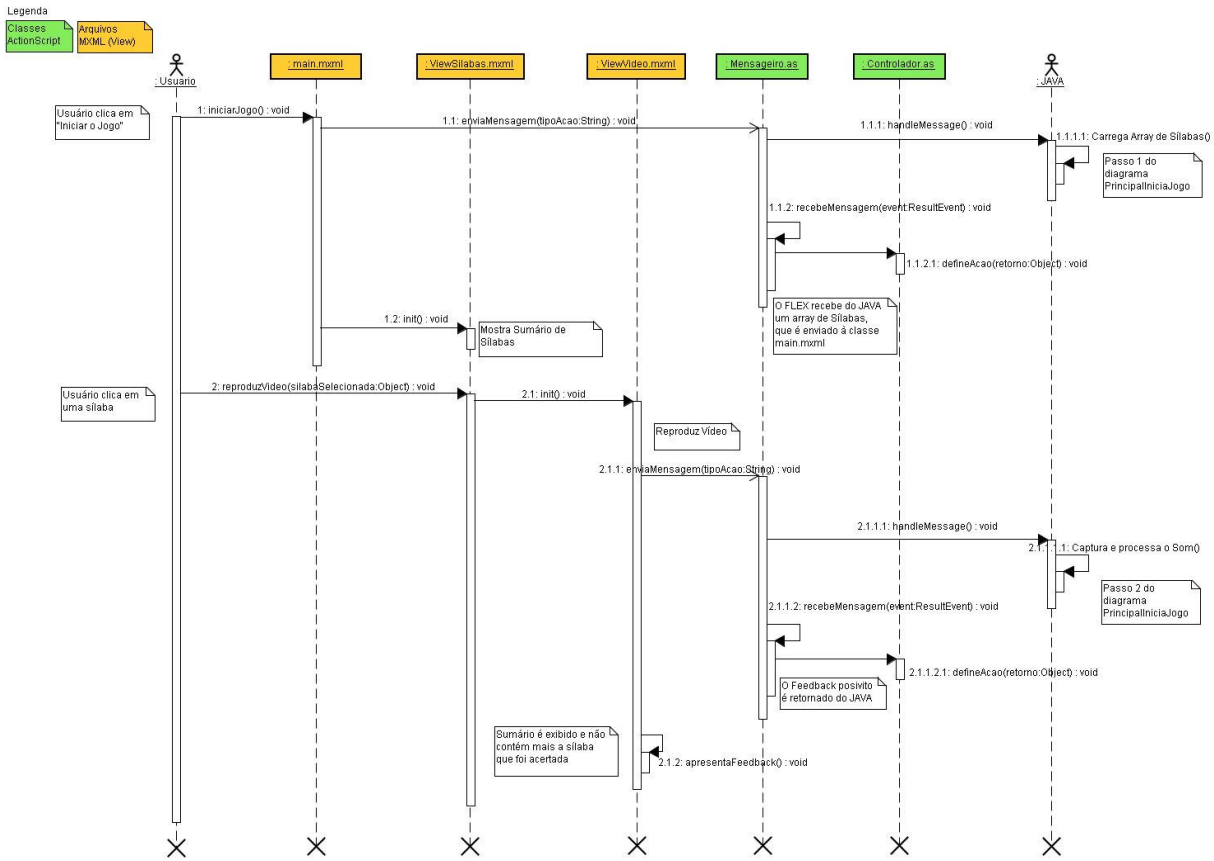


DIAGRAMA 9 – DIAGRAMA DE SEQÜÊNCIA FLEX UC001 - FLUXO PRINCIPAL

5.2.4.6. Flex - UC001 Iniciar Jogo – Alternativo (A1)

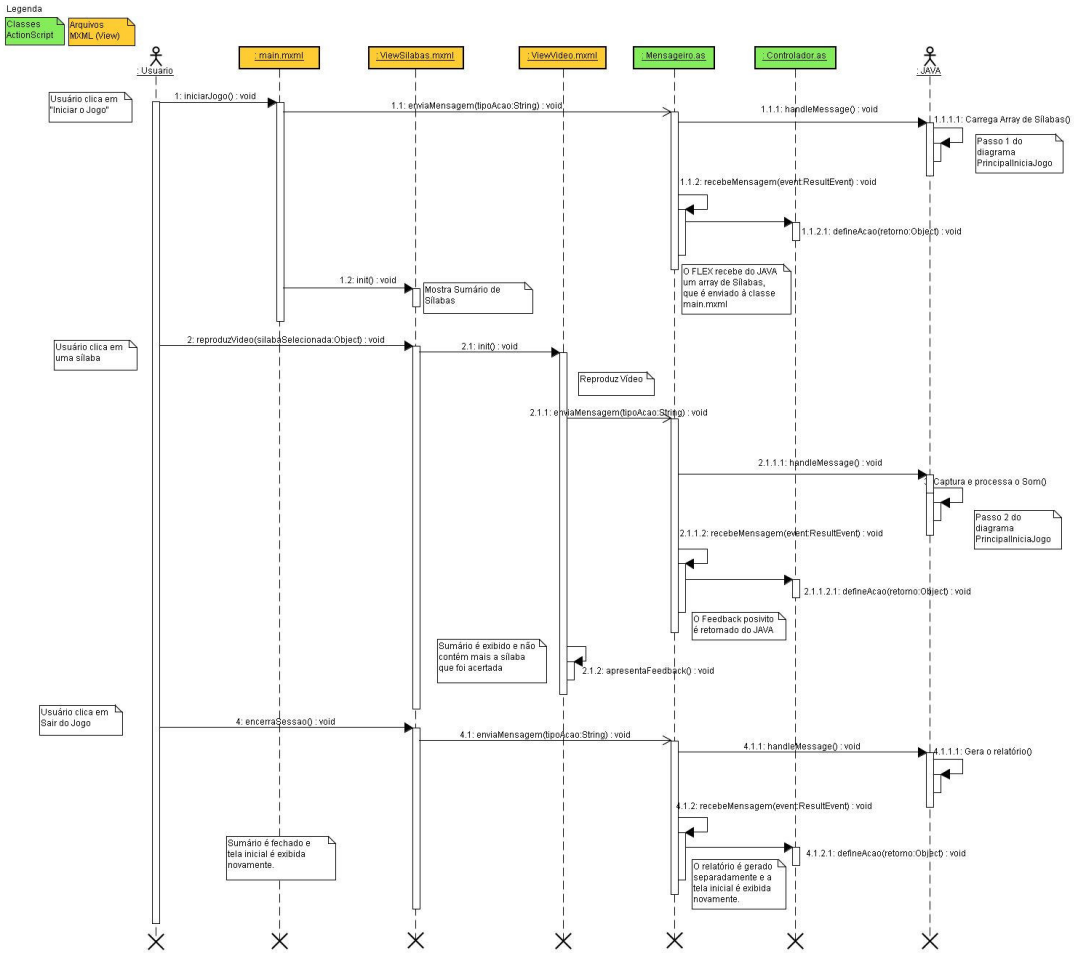


DIAGRAMA 10 – DIAGRAMA DE SEQÜÊNCIA FLEX UC001 - FLUXO (A1)

5.2.4.7. Flex - UC001 Iniciar Jogo – Alternativo (A2)

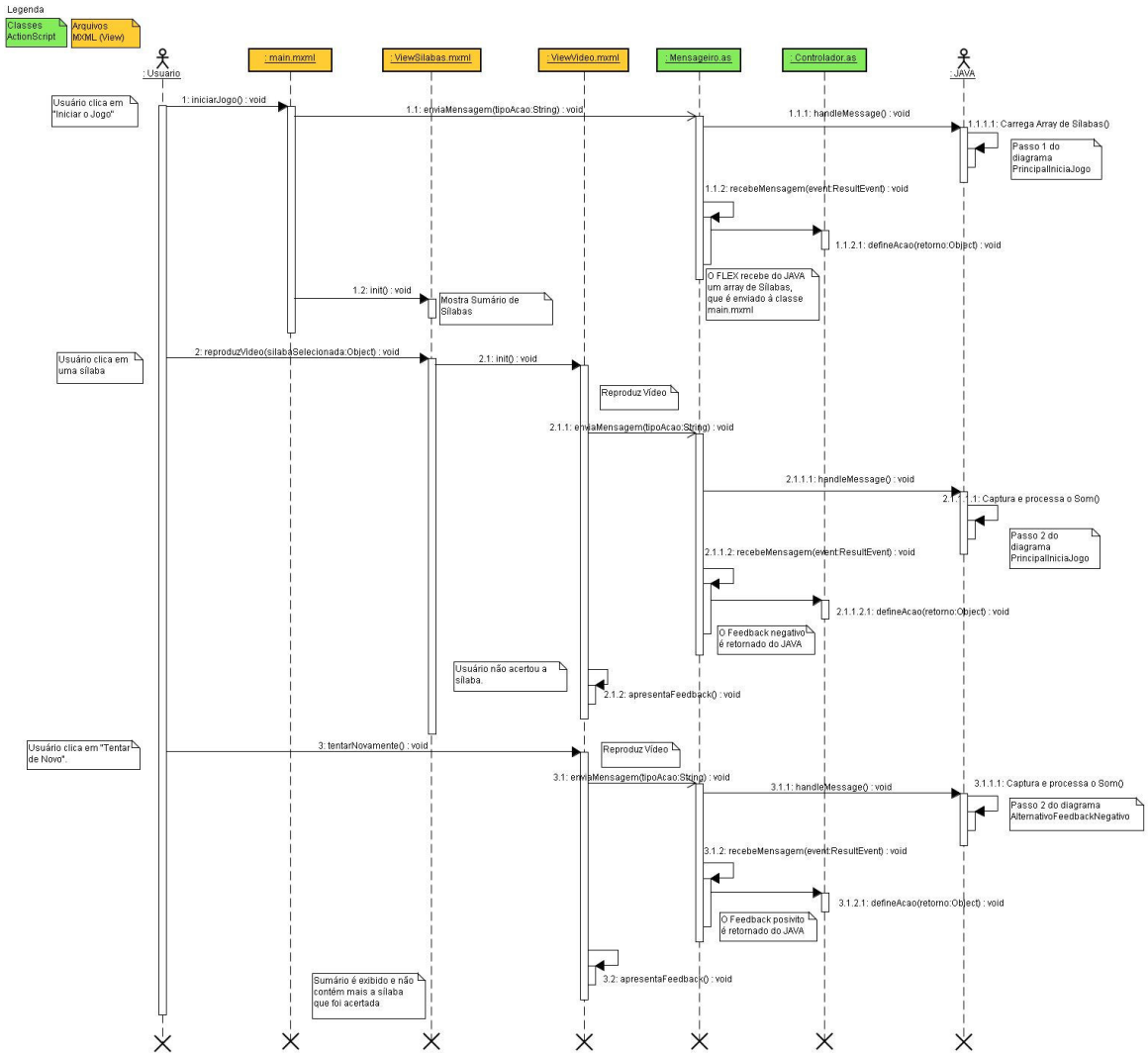


DIAGRAMA 11 – DIAGRAMA DE SEQÜÊNCIA FLEX UC001 - FLUXO (A2)

5.2.4.8. Flex - UC001 Iniciar Jogo – Alternativo (A3)

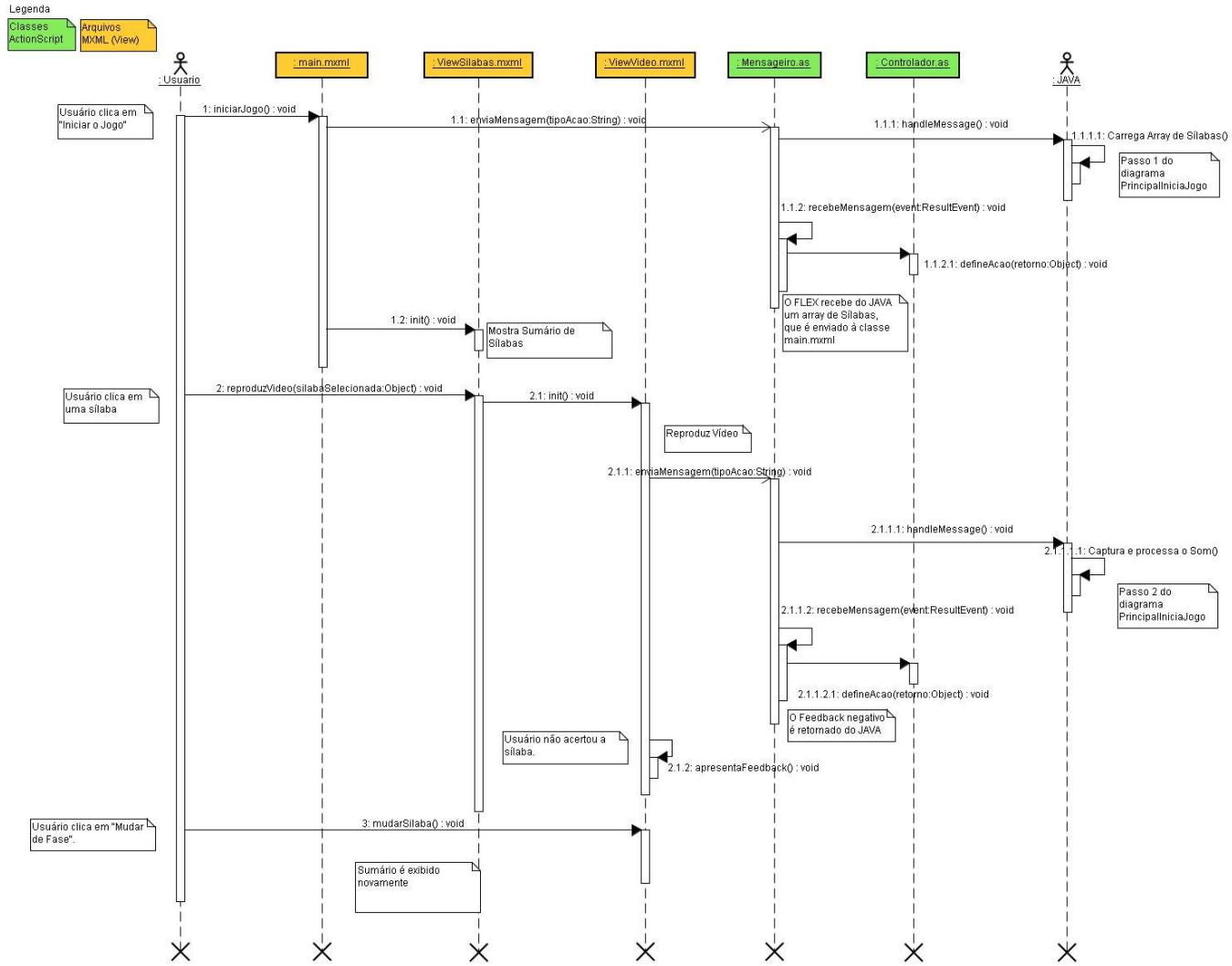


DIAGRAMA 12 – DIAGRAMA DE SEQÜÊNCIA FLEX UC001 - FLUXO (A3)

5.2.5. Diagrama de componentes

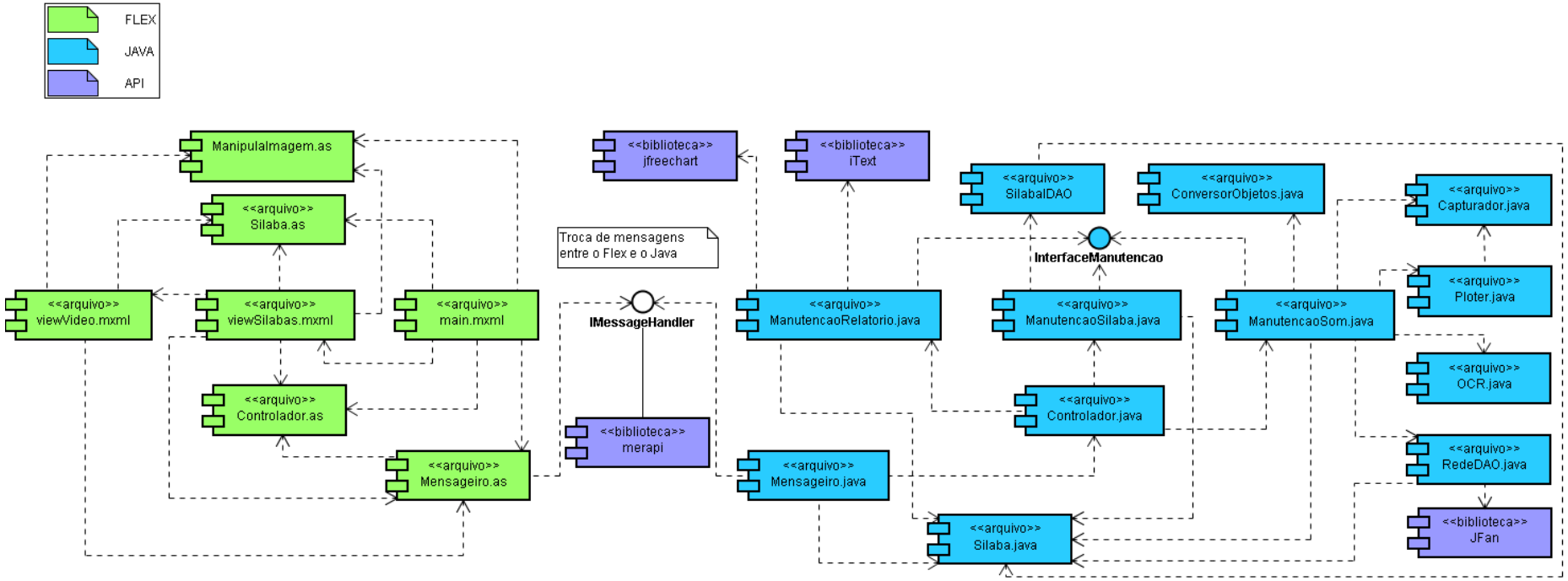


DIAGRAMA 14 – DIAGRAMA DE COMPONENTES

6. EXPERIMENTOS

Esta seção descreve todas as tentativas de se chegar a uma técnica para a extração de características para treino da Rede Neural. Cada experimento aqui descrito (com a exceção do experimento 6.5) foi descartado por não atingir resultados satisfatórios ou por decisão comum da equipe e orientadores. Deve-se notar que as porcentagens descritas nos resultados levam em consideração a média Harmônica observada no software EasyFan [12], o qual já foi descrito na seção de ferramentas.

6.1. Utilizando uma idéia já formulada

A idéia inicial era incrementar uma técnica implementada anteriormente por outra equipe, a qual consistia em extrair as amplitudes do som em um array, ordenar em ordem crescente e definir um ponto de corte em 70% do array. Desse ponto para trás todas as amplitudes seriam definidas como 0 e para frente todas seriam definidas como 1. A partir disso, voltariam os elementos do array para as suas posições originais (antes da ordenação). Com os elementos transformados em 1 e 0 nas posições originais verificam-se os pontos de mudança do array, ou seja, em qual posição muda o valor do elemento (de 1 para 0 ou o contrario). Tendo os pontos de mudança seria aplicado um cálculo em cada valor retornando números de 1 a 6. A quantidade de cada um desses números de 1 a 6 seria enviada para o treinamento da rede. Porém em reunião com os orientadores esta idéia foi descartada, pois foi resolvido utilizar a idéia atual, a qual compreende a utilização do gráfico do fonema.

6.1.1. Resultados

A técnica não chegou a ser aplicada.

6.2. Gráficos com Matlab e uma rede treinada

Após a mudança na técnica de extração de características, foram pesquisadas algumas formas de desenhar o gráfico das amplitudes no domínio do tempo, e a primeira forma encontrada foi utilizando o software Matlab. O som era capturado pelo módulo Java, passado ao Matlab para plotar o gráfico e retornado o gráfico ao módulo Java para aplicar as técnicas de extração de características em cima dele.

Esta técnica foi aplicada utilizando um universo de 318 indivíduos, o que representa a primeira coleta de vozes, na média de 12 indivíduos para cada classe.

Esta forma se mostrou ineficiente, e a mesma foi inviabilizada pelo fato de o Matlab ser uma ferramenta *Shareware* e de difícil integração com o Java.

6.2.1. Resultados

Com esta maneira de extração de características, foi obtida uma média de 40% acerto, o que não foi satisfatório em termos de teste da rede neural utilizando indivíduos do conjunto que foram incluídos no treinamento.

6.3. Gráficos rústicos com Java e uma rede treinada

Buscando uma alternativa para substituir o software Matlab, foi implementada uma solução para desenhar o gráfico das amplitudes do fonema utilizando o Java, através da API Java Sound [3] [4]. O gráfico foi desenhado de

forma correta, mas as linhas do gráfico eram pouco detalhadas, o que levou a equipe a concluir que se deveria tentar outras formas de plotar o gráfico.

Nesse ponto ainda continuava-se com 318 indivíduos e uma rede com todas as classes de fonemas.

Alguns testes foram aplicados nos arquivos gravados para observar os resultados obtidos, conforme lista a seguir:

- Aplicando a técnica do OCR na imagem dimensionada em 1000 pixels de altura por 1000 pixels de largura;
- Aplicando a técnica do OCR na imagem dimensionada em 50 pixels de altura por 50 pixels de largura;
- Aplicando a técnica de trimming em uma imagem de 1000 pixels de altura por 1000 pixels de largura;
- Normalizando os arquivos de som separados para o treinamento da rede e testando com arquivos não normalizados;
- Normalizando todos os arquivos, tanto pra treino da rede neural, quanto para testes.

6.3.1. Resultados

A taxa de acerto variou de 40% a 55%. Essa taxa ainda era baixa e não foi satisfatória, o que foi motivo para procurar uma nova maneira de gerar os gráficos.

6.4. Gráficos detalhados com Java e uma rede treinada

Após constatar que gráficos detalhados poderiam gerar taxas de acerto maiores, buscou-se uma nova forma de gerar os gráficos e foram encontrados alguns recursos da API Jmusic [6], a qual proporcionou uma nova forma de plotar o gráfico das amplitudes do fonema. Os gráficos gerados eram bem detalhados, mas

infelizmente essa hipótese foi descartada, pois a taxa de acerto diminuiu. Foram aplicados os seguintes testes:

- Aplicando a técnica do OCR na imagem dimensionada em 1000 pixels de altura por 1000 pixels de largura;
- Aplicando a técnica do OCR na imagem dimensionada em 50 pixels de altura por 50 pixels de largura;
- Aplicando a técnica de trimming em uma imagem de 1000 pixels de altura por 1000 pixels de largura;
- Normalizando os arquivos de som separados para o treinamento da rede e testando com arquivos não normalizados;
- Normalizando todos os arquivos, tanto pra treino da rede neural, quanto para testes.

6.4.1. Resultados

As taxas de acerto variaram entre 35% a 60%, pois dependiam de número de indivíduos por classe. O resultado continuou inaplicável.

6.5. Gráficos rústicos com Java e uma rede por fonema

Após a confirmação de que gráficos detalhados não trouxeram bons resultados, voltou-se a utilizar os gráficos rústicos, mas com dois diferenciais, uma nova coleta foi realizada, a qual adicionou mais 350 indivíduos ao conjunto e passou-se a utilizar uma rede treinada para cada classe de fonema. Desta forma, as taxas de acerto melhoraram consideravelmente.

Esta é a técnica utilizada atualmente, mas ainda não se mostrou muito eficiente na classificação de fonemas em tempo real.

6.5.1. Resultados

Segue abaixo a tabela com a porcentagem da taxa de acerto, as mesmas podem ser conferidas na seção Matrizes de confusão em Anexos.

Taxas de acerto com testes na rede neural (considerando a média harmônica):

TABELA 5 – TAXAS DE ACERTO CONSIDERANDO A MÉDIA HARMÔNICA

Á – 91%	MÊ - 93%
BÔ - 80%	NÊ - 94%
CHÁ - 94%	NHÊ - 86%
DÉ - 100%	Ô - 95%
Ê - 100%	Ó - 93%
É - 87%	PÊ – 87%
FÔ – 83%	RÊ - 100%
GÁ - 87%	RRÊ – 86%
I - 90%	SSÉ – 90%
JÁ - 75%	TÉ – 71%
KÁ - 78%	U – 92%
LÊ - 85%	VÁ – 70%
LHÊ - 70%	ZÁ – 67%

Embora as taxas de acerto na rede neural tenham sido satisfatórias, levando-se em conta o número relativamente pequeno de indivíduos por classe, o mesmo sucesso não se repetiu em testes reais da aplicação interagindo diretamente tanto com adultos como com crianças.

Pelo conhecimento obtido até o momento no desenvolvimento da técnica atual pode-se dizer que o problema estaria no fato de o gráfico gerado pela captura não estar parecido com os gráficos gerados a partir dos sons das coletas realizadas. Principalmente pelo fato de o *trimming* estar muito sensível a qualquer tipo de ruído, fazendo com que ainda exista silêncio não cortado no som, provocando a geração incorreta do gráfico.

Mas convém reavaliar as características utilizadas para a comparação de cada fonema, pois uma vez que não se tem características que conseguem

distinguir com clareza essa diferença elas acabam não sendo úteis para o objetivo do projeto, podendo impactar na taxa de acerto do teste real da aplicação.

A seguir é possível ver os relatórios reais gerados pelo próprio J-Som, sendo um deles gerado a partir de um teste realizado por uma pessoa adulta e o outro por uma criança da faixa etária que o J-Som cobre.

Data do teste: 04/12/2008
Horário de início: 12:54:47
Horário de término: 13:21:22

Resultado do Teste

É: 1	RÊ: 2	Ê: 2	DÉ: 2	FÔ: 2
PÊ: 3	ZÁ: 0	NÊ: 0	Á: 0	RRÊ: 0
I: 0	BÔ: 2	TÉ: 2	VÁ: 2	U: 3
KÁ: 2	Ô: 2	GÁ: 3	SSÉ: 4	Ó: 0
MÊ: 0	CHÁ: 2	LÊ: 2	NHÊ: 0	LHÊ: 3

Quantidade de Acertos	Quantidade de Erros	Não clicados
1	40	8

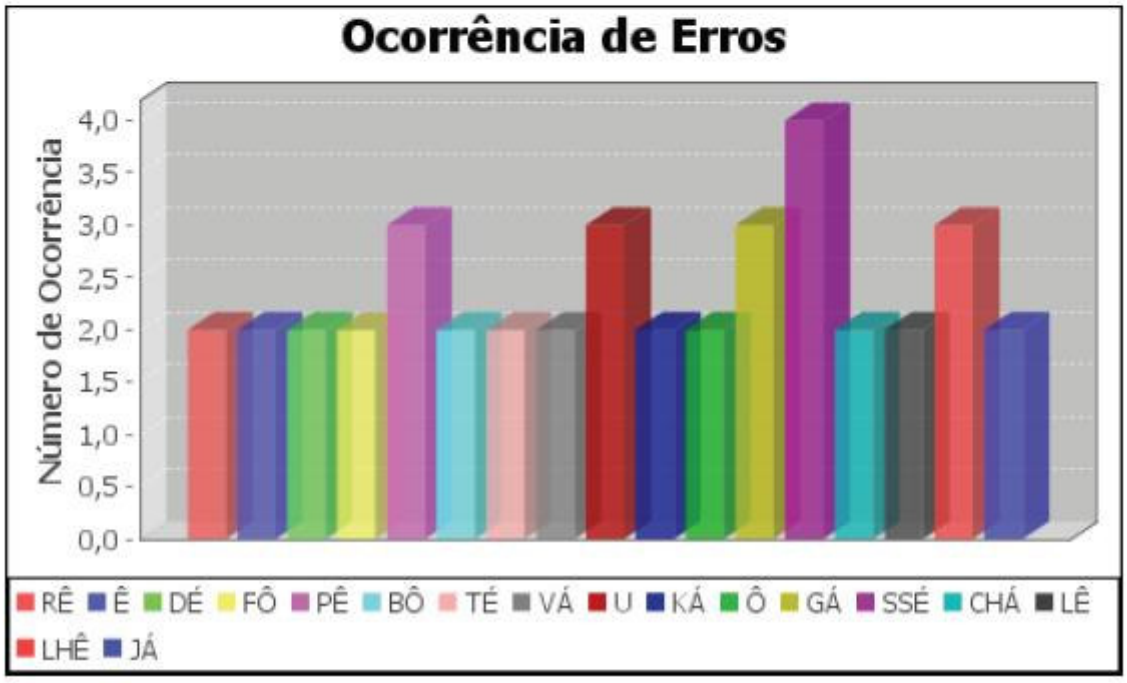


FIGURA 9 – RELATÓRIO DE TESTE REALIZADO POR UMA PESSOA ADULTA

Data do teste: 04/12/2008
Horário de início: 12:15:47
Horário de término: 12:51:36

Resultado do Teste

É: 1	RÊ: 2	Ê: 6	DÉ: 4	FÔ: 5
PÊ: 4	ZÁ: 2	NÊ: 1	Á: 3	RRÊ: 4
I: 3	BÔ: 4	TÉ: 2	VÁ: 3	U: 2
KÁ: 3	Ô: 1	GÁ: 2	SSÉ: 3	Ó: 2
MÊ: 6	CHÁ: 6	LÊ: 5	NHÊ: 1	LHÊ: 2

Quantidade de Acertos	Quantidade de Erros	Não clicados
7	74	0

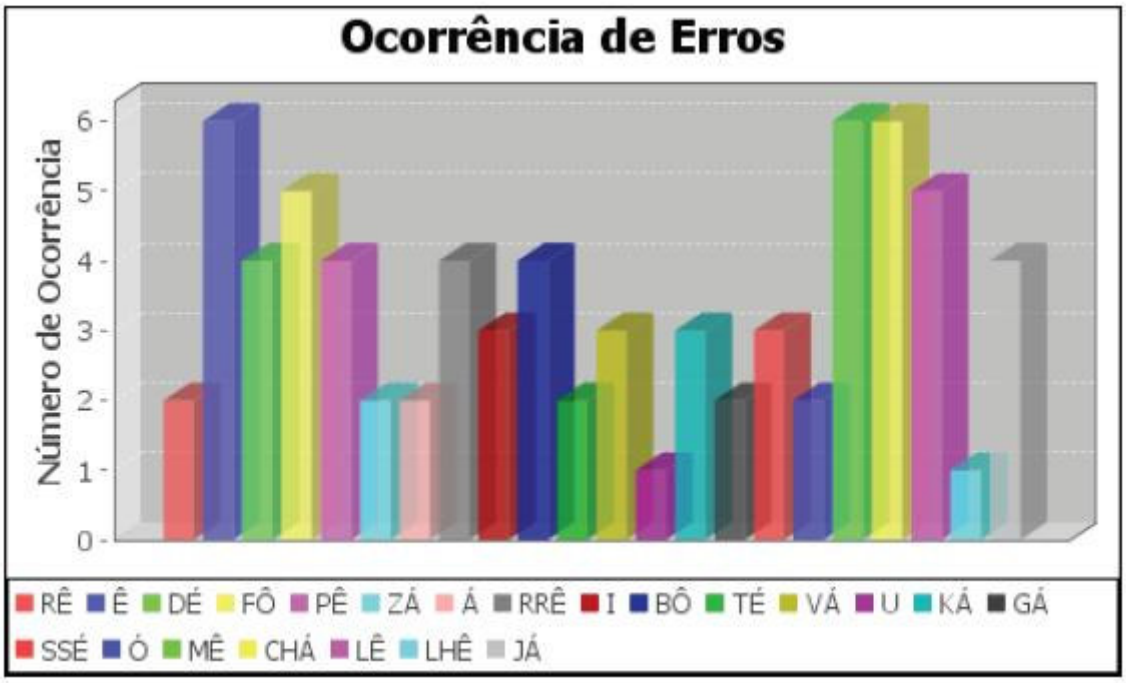


FIGURA 10 – RELATÓRIO DE TESTE REALIZADO POR UMA CRIANÇA

7. CONCLUSÃO E TRABALHOS FUTUROS

O desenvolvimento do projeto J-Som proporcionou à equipe do projeto um grande nível de aprendizado e conhecimento. Os objetivos estabelecidos no escopo do projeto foram alcançados de forma satisfatória, levando em consideração a complexidade do problema. A taxa média de acerto, no reconhecimento das sílabas faladas chegou a aproximadamente 90% em testes com a rede neural, além do quê, todas as funcionalidades planejadas foram implementadas.

Através da elaboração do sistema J-Som foi possível colocar em prática os conhecimentos adquiridos no período acadêmico, a busca por novos conhecimentos também foi necessária. Os integrantes da equipe adquiriram experiência trabalhando em equipe, compartilhando idéias, problemas e soluções, além de experiência em gerenciar um projeto, estabelecimento de uma metodologia capaz de controlar as atividades, os recursos e o tempo disponível. Desta forma foi possível acompanhar a evolução do projeto e manter sua rastreabilidade.

7.1. Dificuldades enfrentadas

Inúmeras dificuldades foram encontradas no andamento do projeto, dentre as quais é importante ressaltar algumas:

A maior dificuldade foi trabalhar com inteligência artificial; além de ser uma tecnologia complexa a equipe não tinha muita experiência no assunto, apenas a elaboração de alguns trabalhos acadêmicos. Para suprir esta deficiência foi preciso contar com a ajuda dos orientadores e ferramentas especializadas.

Trabalhar com reconhecimento de voz, principalmente porque a voz pode sofrer inúmeras alterações na captura e prejudicar o estabelecimento de padrões no processo de aprendizado da rede neural. Dentre as alterações podemos citar: ruídos

de fundo, estado físico, emocional e cultural do locutor, características do microfone, além de outras influências como sotaques e regionalismo.

Encontrar uma solução viável ao problema proposto, pois várias idéias foram testadas, no entanto muitas se tornaram inviáveis e foram descartadas, tomando o tempo hábil da equipe.

Para desenvolver o projeto foi necessário aprender novas ferramentas e tecnologias como o Easy Fan [12], Flex [17] e CVS, assim, até o grupo adquirir o conhecimento necessário para trabalhar com elas, levou certo tempo.

8. O FUTURO DO J-SOM

O projeto J-Som deixa uma grande janela para novas implementações e melhorias. As idéias relacionadas abaixo não fizeram parte do escopo do projeto, principalmente pelo pouco tempo disponível para sua conclusão.

- Aumentar a capacidade de reconhecimento do sistema, maximizando o conjunto de sílabas envolvidas e até módulos com reconhecimento de frases;
- Incorporar junto ao sistema de *feedback* um banco de dados para armazenar as informações dos usuários, desta forma, o profissional de Fonoaudiologia poderia acompanhar a evolução da criança recuperando informações sempre que necessário;
- Outras formas de extração de características também podem ser estudadas na intenção de conseguir melhores resultados.
- Apesar de todas as dificuldades e adversidades encontradas, a equipe do projeto aceitou o desafio, principalmente pelo retorno a nível de conhecimento que este projeto poderia, e pode, proporcionar. A equipe sofreu com a falta de experiência, mas ganhou aprendizado, superou suas limitações para alcançar os objetivos propostos. Os esforços valeram a pena considerando tudo que foi ganho, em conhecimento e experiência.

9. REFERÊNCIAS

[1] PONCE DE LEON F. DE CARVALHO, André. **Redes Neurais Artificiais.** Disponível em: http://www.icmc.usp.br/~andre/research/neural/. Acesso em: 15 novembro. 2008.

[2] GRABIANOVSKI, Edi. **Como funciona o reconhecimento de voz**. Disponível em: http://informatica.hsw.uol.com.br/reconhecimento-de-voz1.htm. Acesso em: 16 novembro 2008

[3] PFISTERER, Matthias; BOMERS, Florian. **Java Sound Resources**. Disponível em: http://www.jsresources.org/. Acessado em: 01 novembro 2008.

[4] Sun Microsystems, Inc. **Java Sound API**. Disponível em: http://java.sun.com/products/java-media/sound/. Acessado em: 29 outubro 2008.

[5] BRAGA, Petrônio. **Reconhecimento de voz dependente de locutor utilizando Redes Neurais Artificiais**. Universidade de Pernambuco, 2006. Dissertação, Engenharia da Computação, Recife 2006.

[6] SORENSEN, Andrew; BROWN, Andrew. **jMusic: Music composition in Java.** Disponível em: http://jmusic.ci.qut.edu.au/. Acessado em: 20 outubro 2008.

[7] Sun Microsystems, Inc. **Java™ 2 Platform Standard Edition 5.0**. Disponível em: http://java.sun.com/j2se/1.5.0/docs/api/. Acessado em: 10 setembro 2008.

[8] MOREIRA, Luís Filipe. **Reconhecimento Automático de Fala Contínua**. Disponível em: http://www.ipb.pt/~fmoreira/Itens/RecFalaCont.pdf. Acessado em: novembro 2008.

[9] Adobe System Incorporated, **Adobe Air**. Disponível em: http://www.adobe.com/products/air/. Acessado em: 05 setembro2008.

[10] BRAUN, Cláudia. **FonoFlex**. Disponível em: http://www.ctsinformatica.com.br/#fonoFlex.html{paginaProduto!8&1. Acessado em: 10 maio 2008.

[11] RAITTZ, R. T. **FAN 2002: um Modelo Neuro-Fuzzy para Reconhecimento de Padrões.** 85 f. Tese(Doutorado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 2002. T.

[12] GARRET, L. F. V.; IGNACIO, F. A.; KUSTER, C.W.; LENFERS, F. P.; ZOTTO S. P. **EASYFAN.** 165f. Trabalho de Conclusão de Curso de Tecnologia em Informática (Disciplina de Projetos) – Curso de Tecnologia em Informática, Setor da Escola Técnica, Universidade Federal do Paraná, Curitiba, 2006.

[13] TAFNER, Malcon Anderson. **RNA: Aprendizado e plasticidade**. Disponível em: http://www.cerebromente.org.br/n05/tecnologia/rna.htm. Acessado em 10 novembro 2008.

[14] Tecnológica, Inovação. **Programa transforma voz em linguagens de sinais**. Disponível em: http://www.inovacaotecnologica.com.br/noticias/noticia.php?artigo=010150070914 Acessado em: 3 novembro 2008.

[15] Telemar, assessoria. **Atendimento Telemar com reconhecimento de voz**. Disponível em; http://www.momentoeditorial.com.br/index.php?option=com_content&task=view&id=2785&Itemid=1. Acessado em: 10 novembro 2008.

[16] GRABIANOVSKI, Edi. **O futuro do reconhecimento de voz**. Disponível em; http://informatica.hsw.uol.com.br/reconhecimento-de-voz4.htm. Acessado em: 16 novembro 2008.

[17] Adobe System Incorporated, **Adobe Flex 3**. Disponível em: http://www.adobe.com/products/flex/?promoid=BPDEQ. Acessado em: 05 setembro2008.

10. GLOSSÁRIO

Amplitudes: Medida escalar não negativa da magnitude de oscilação de uma onda, a amplitude de uma onda pode ser constante ou variar com o tempo.

Arquivos Wave: Arquivos de armazenamento de áudio não comprimidos (sem perda), facilmente editado e manipulado através de softwares específicos

Cepstrum: é o resultado da Transformada de Fourier do espectro de decibéis como se fosse um sinal.

Espectro: é o conjunto de todas as ondas que compõem os sons audíveis e não audíveis pelo ser humano.

Fonema: Fonema é a menor unidade sonora (fonética) de uma língua que estabelece contraste de significado para diferenciar palavras.

Freeware: Programa de computador cuja utilização não implica no pagamento de licenças de uso ou *royalties*.

Glote: é uma estrutura anatômica localizada na porção final na laringofaringe com a função de impedir a entrada de alimentos facilitando a saída e a entrada de ar para os brônquios e pulmões, ajuda também na função fonatória uma vez que a prega vocal e vestibular localizam-se dentro dela.

IDE: do inglês Integrated Development Environment ou Ambiente Integrado de Desenvolvimento, é um programa de computador que reúne características e ferramentas de apoio ao desenvolvimento de software com o objetivo de agilizar este processo.

Lógica Difusa: A lógica difusa ou lógica fuzzy é uma generalização da lógica booleana que admite valores lógicos intermediários entre a falsidade e a verdade (como o talvez). Como existem várias formas de se implementar um modelo fuzzy, a lógica fuzzy deve ser vista mais como uma área de pesquisa sobre

tratamento da incerteza, ou uma família de modelos matemáticos dedicados ao tratamento da incerteza, do que uma lógica propriamente dita.

Matlab: é um software interativo de alto desempenho voltado para o cálculo numérico.

Normalização: é o ajuste das amplitudes a um valor padrão, desta forma, têm-se todas as amplitudes de todos os arquivos com a mesma proporção.

Optical Character Recognition **(OCR):** técnica que consiste em reconhecer caracteres a partir de um arquivo de imagem, ou mapa de bits. O objetivo desta técnica é identificar característica de um objeto através da varredura nas 4 direções possíveis (da direita para esquerda, de cima para baixo e seus inversos). Levando em consideração que a imagem de fundo em que está esse objeto, seja uma cor uniforme, como o branco, por exemplo, têm-se todos os pixels até chegar a uma "parte pintada" do objeto. Com isso é identificado o número total de pixels que não estão pintados e com a divisão pelo número total de pixels da imagem, têm-se uma característica. Aplicando isso as 4 direções, têm-se as 4 características principais.

Reconhecimento de Padrões: Classificar informações (padrões) baseadas ou em conhecimento a priori ou em estatísticas extraídas dos padrões. Essa área de atuação é estudada por vários campos, tais como psicologia, etologia e ciência da computação.

RNA (Rede Neural Artificial): Redes Neurais Artificiais, ou Redes Neuronais, são sistemas computacionais baseados numa aproximação à computação baseada em ligações. Nós simples (ou "neurões", "neurônios", "processadores" ou "unidades") são interligados para formar uma rede de nós - daí o termo "rede neuronal". A inspiração original para esta técnica advém do exame das estruturas do cérebro, em particular do exame de neurônios. Portanto, baseia-se no funcionamento de um neurônio biológico utilizado para treinamento de alguns sistemas que utilizam inteligência artificial.

Shareware: é um programa de computador disponibilizado gratuitamente, porém com algum tipo de limitação.

Silaba: Sílaba é o conjunto de um ou mais fonemas pronunciados numa única emissão de voz. Na língua portuguesa, o núcleo da sílaba é sempre uma vogal: não existe sílaba sem vogal e nunca há mais do que uma única vogal em cada sílaba

Som: é a propagação de uma frente de compressão mecânica ou onda longitudinal; esta onda se propaga de forma circuncêntrica, apenas em meios materiais que têm massa e elasticidade, como os sólidos, líquidos ou gasosos, quer dizer, não se propaga no vácuo.

Transformada Rápida de Fourier: é um algoritmo eficiente para se calcular a Transformada discreta de Fourier (DFT) e a sua inversa.

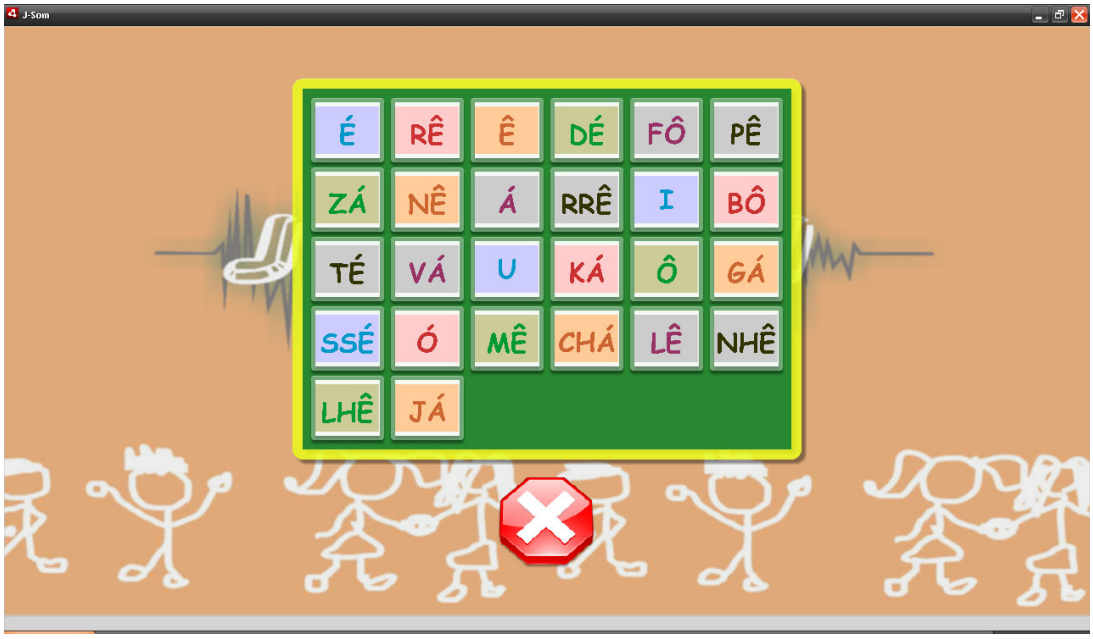
Matrizes de Confusão: são matrizes com as taxas de acerto da rede neural em um conjunto de indivíduos de teste após os treinamentos da mesma.

11. ANEXOS

11.1. Telas

11.1.1. DV001

11.1.2. DV002

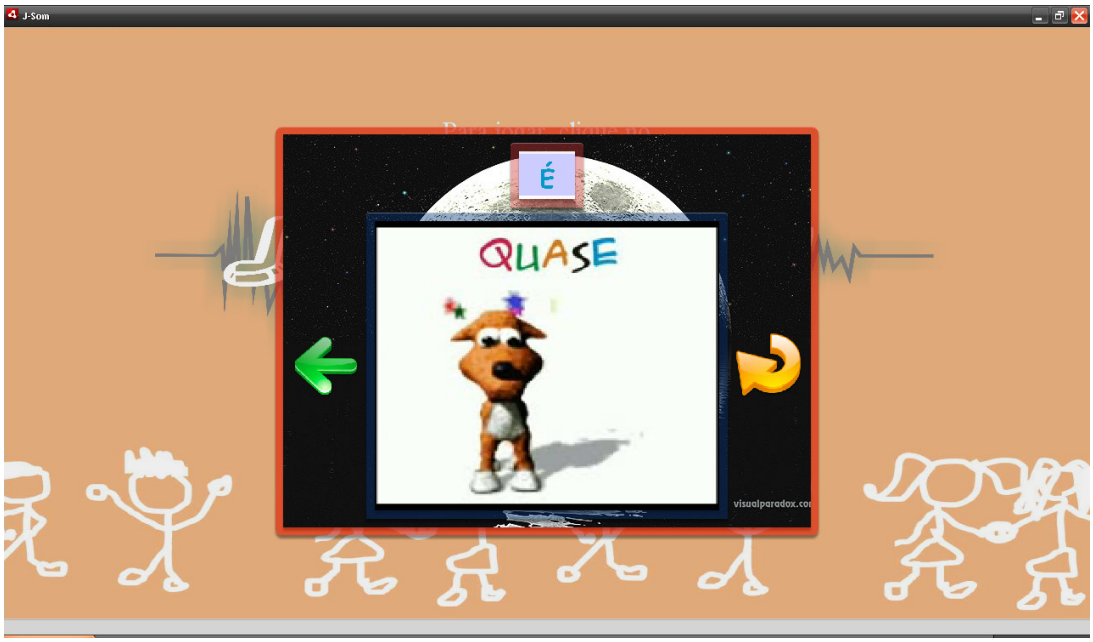


11.1.3. DV003

11.1.4. DV004

11.1.5. DV005

11.1.6. DV006



11.2. Guia de instalação

11.2.1. Requisitos Mínimos

Para o pleno funcionamento do sistema, é requerido um microcomputador com as seguintes configurações mínimas:

- Processador com velocidade superior a 1.0 GHz;
- Memória de 1 gigabyte;
- Caixas de som e microfone em pleno funcionamento;
- Sistema operacional Microsoft Windows XP ou superior;
- *Adobe Air* 1.1 [9] ou superior instalado.
- Java Runtime Environment 1.5 [7] instalado.

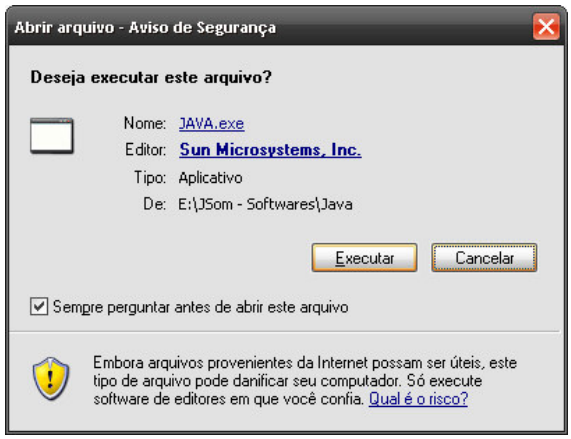
11.2.2. Passos para a Instalação

11.2.2.1. Java Runtime Environment

Este *software* é um dos necessários para executar o J-Som e qualquer programa que utilize tecnologia JAVA. Para instalá-lo, siga as instruções abaixo:

Abra o arquivo “JAVA.exe”.

Caso apareça uma janela como mostra a imagem abaixo, clique em “Executar”.



Em seguida, clique em “Accept”, como mostra a tela abaixo.

A instalação será realizada. Assim que acabar e a tela abaixo aparecer, clique em “Finish”.

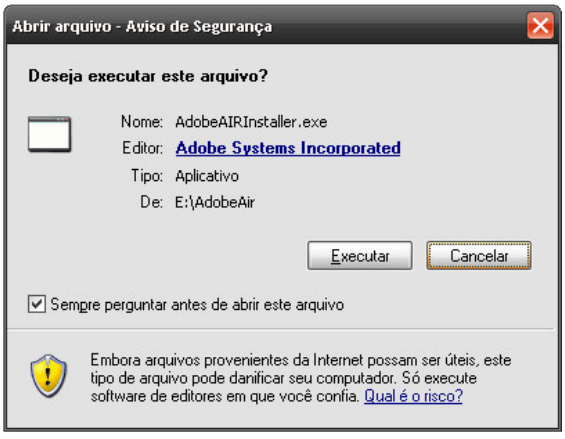
Depois de executar esses passos, você terá o Java Runtime Environment instalado no seu computador.

11.2.2.2. Adobe® Air™

Este *software* é um dos necessários para executar o J-Som e qualquer programa que utilize tecnologia FLEX. Para instalá-lo, siga as instruções abaixo:

Abra o arquivo “AdobeAIRInstaller.exe”.

Caso apareça uma janela como mostra a imagem abaixo, clique em “Executar”.



Em seguida, clique em “Concordo”.

A instalação será executada. Assim que a tela abaixo aparecer, clique em “Concluir”.

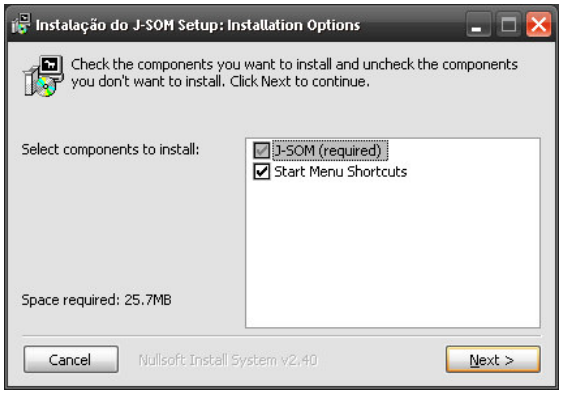
Depois de executar esses passos, você terá o *Adobe® Air™* instalado no seu computador!

11.2.2.3. J-Som

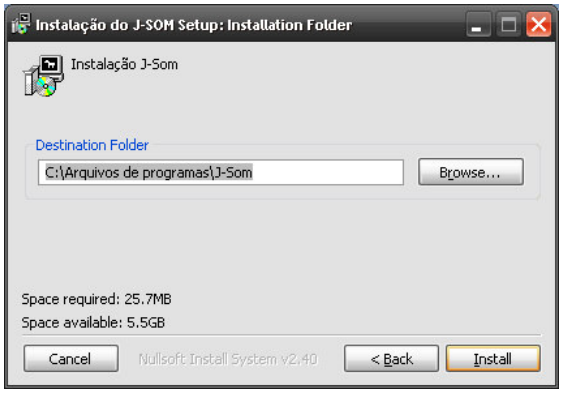
Após ter instalado os programas citados acima, você estará habilitado para instalar o próprio J-Som. Para isso, siga os passos abaixo:

Abra o arquivo “Instalar J-Som.exe”.

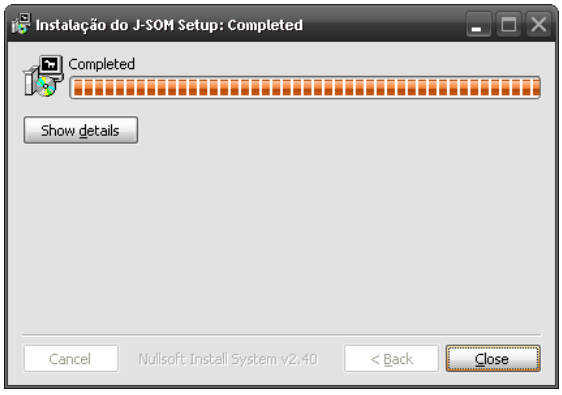
Na tela abaixo, clique em “Next”.



Na tela seguinte você poderá escolher a pasta onde os arquivos do J-Som serão copiados. A pasta “Arquivos de Programas” (onde normalmente são instalados os programas) estará previamente selecionada. Caso não tenha certeza de que pasta selecionar, não selecione outra pasta (lembre-se que se selecionar uma pasta que possa ser excluída por algum motivo o programa não funcionará mais). Clique em “Install”.

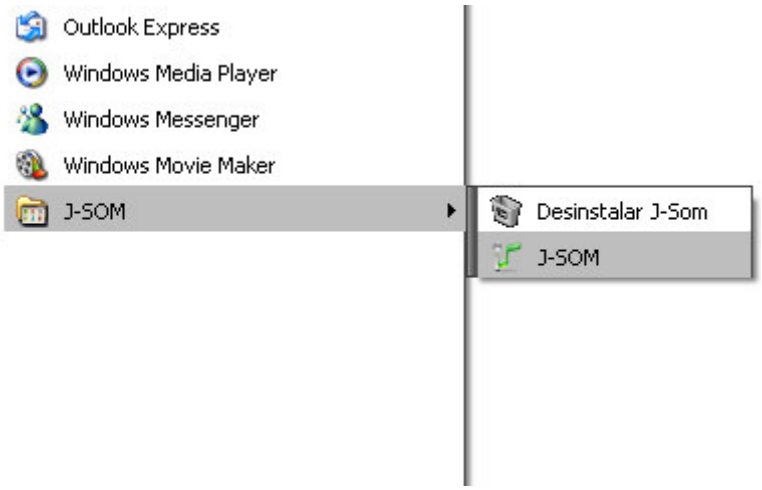


A instalação será executada. Assim que a tela abaixo aparecer, clique em “Close”.



Depois de executar esses passos, você terá o J-Som instalado no seu computador!

Para executar o programa, acesse na barra de tarefas o menu “Iniciar>Programas>J-SOM>J-SOM”, como você pode ver na figura abaixo.



Assim que clicar no ícone do J-Som você poderá ver a tela inicial do programa, como exibida abaixo, e poderá começar a jogar imediatamente!

11.3. Matrizes de confusão

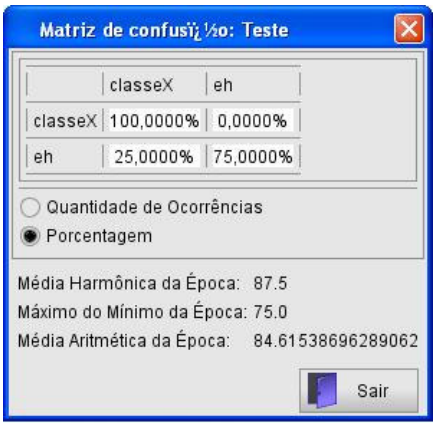
11.3.1. Á

11.3.2. BÔ

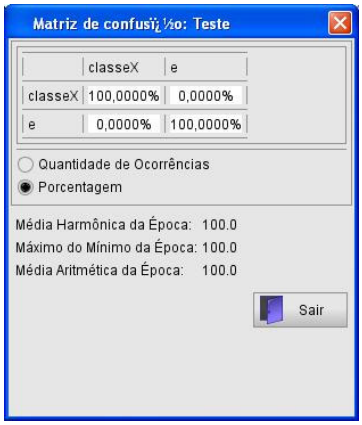
11.3.3. CHÁ

11.3.4. DÉ

11.3.5. É



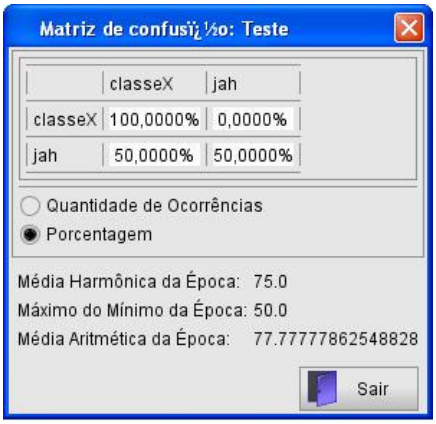
11.3.6. Ê



11.3.7. FÔ

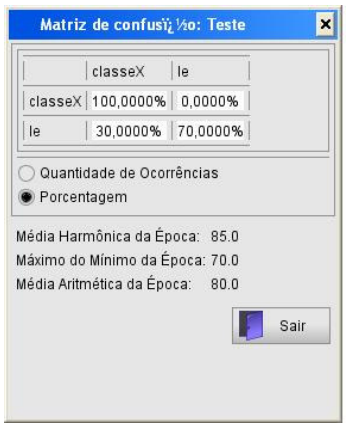
11.3.8. GÁ

11.3.9. JÁ

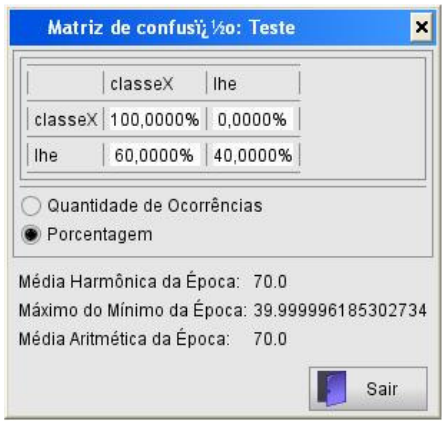


11.3.10. KÁ

11.3.11. LÊ



11.3.12. LHÊ

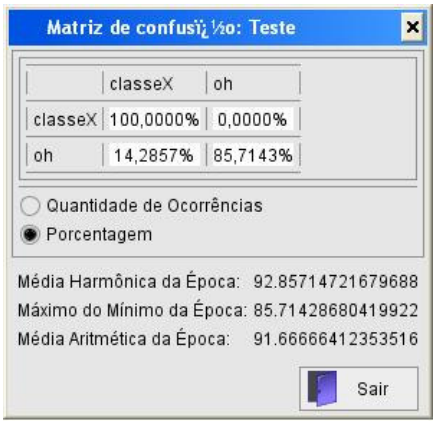


11.3.13. MÊ

11.3.14. NÊ

11.3.15. NHÊ

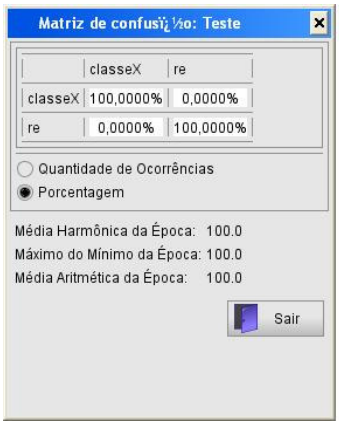
11.3.16. Ó



11.3.17. Ô

11.3.18. PÊ

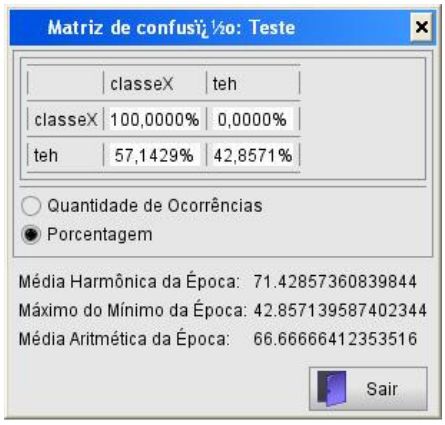
11.3.19. RÊ



11.3.20. RRÊ

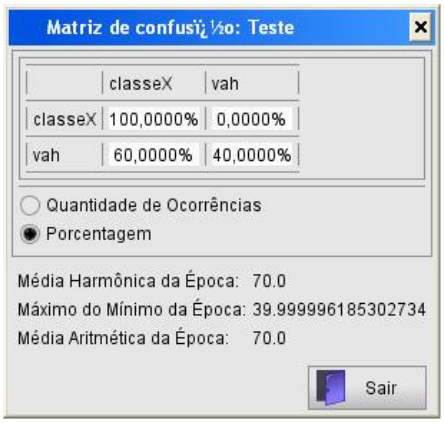
11.3.21. SSÉ

11.3.22. TÉ



11.3.23. U

11.3.24. VÁ



11.3.25. ZÁ