



Avaliação de *software* educativo com reconhecimento de fala em indivíduos com desenvolvimento normal e atraso de linguagem¹

Henildes José Carrer
henildescarrer@yahoo.com.br

Ednaldo Brigante Pizzolato
Universidade Federal de São Carlos
Departamento de Computação
ednaldo@dc.ufscar.br

Celso Goyos²
Universidade Federal de São Carlos
Departamento de Psicologia
celsogoyos@hotmail.com

Resumo *O objetivo geral deste estudo foi a avaliação do software MESTRE[®] capacitado com tecnologia de reconhecimento de fala, para verificar a sua viabilidade para finalidades educacionais. Participaram deste estudo 110 indivíduos, divididos em quatro grupos. Os estímulos experimentais foram apresentados através do programa informatizado MESTRE[®], constituindo-se em dois conjuntos: A e B, sendo o conjunto A formado por 51 palavras da língua portuguesa, e o conjunto B formado por 51 figuras correspondentes às palavras do conjunto A. Foram testadas as relações: palavra falada – produção oral (relação AE), figura – produção oral (relação BE). Os resultados foram analisados estatisticamente e mostraram que o procedimento é eficaz no reconhecimento da fala de adultos; para crianças pode ser eficaz, considerando-se a idade, pois quanto maior a idade, maior o índice de reconhecimento. À partir deste estudo pode-se concluir que o software educativo MESTRE[®] com a capacidade de reconhecimento automático de fala pode ser um instrumento de grande auxílio para os educadores no trabalho com indivíduos que apresentem desenvolvimento normal e com atrasos relacionados a problemas de cognição e linguagem.*

Palavras-Chave: MESTRE[®], reconhecimento de fala, atraso de linguagem, leitura.

Abstract *The general aim of this study was to evaluate the feasibility of a version of MESTRE[®] software coupled with speech-recognition technology, to meet educational objectives. Overall of 110 individuals participated in this study. These individuals were assigned to four groups according to their age range and language capabilities. The experimental stimuli were presented through the software MESTRE[®], constituted by two sets. Set A was consisted of 51 words of the Portuguese language and Set B constituted of 51 pictures corresponding to the words of Set A. The tested relations were: spoken word - oral production, where Set E were the corresponding words to the figures spoken by the participant (AE relations), and picture – oral production (BE relations). The results were statistically analyzed and showed that the procedure is effective for speech recognition in adults, to children it may be effective considering the age, because the older the children the larger the recognition rate. The conclusion one may draw from this study is that the educational software MESTRE[®] with the ability to automatically recognize speech may be a very useful instrument to aid educators in the work with individuals presenting special educational needs specially those related to language deficits.*

Keywords: MESTRE[®], speech recognition, language-delayed children, reading.

¹ Este trabalho teve como base a dissertação de mestrado da primeira autora, e o segundo e terceiro autores como coorientadores.

² Bolsista produtividade CNPq e Coordenador do Laboratório de Aprendizagem Humana, Multimídia Interativa e Ensino Informatizado (LAHMIEI).

1 Introdução

Nas ocorrências convencionalmente caracterizadas como fracasso escolar, é grande o número de crianças que apresentam dificuldades de leitura e escrita e em sala de aula não conseguem acompanhar as atividades acadêmicas. Parte significativa desse fracasso deve-se à falta de recursos para que o professor possa intervir de maneira eficaz. Entre as muitas causas dessas dificuldades encontra-se o transtorno fonológico que se caracteriza por substituições, omissões e distorções de fonemas, na ausência de fatores etiológicos conhecidos e detectáveis, tais como inadequações anatômicas do aparelho fonador, dificuldades gerais de aprendizagem, déficit intelectual, alterações neurológicas ou fatores ambientais [1].

Em crianças pré-escolares e escolares do ensino fundamental é alta a incidência da substituição entre fonemas oclusivos e fricativos sonoros e seus correspondentes surdos na fala, um dos tipos de transtorno fonológico, sendo que muitas dessas crianças, já no ensino fundamental, apresentam junto a essa alteração a troca dos grafemas (símbolos que representam os fonemas na escrita) e, na maioria dos casos, não superam essa dificuldade sem ajuda profissional especializada [2].

Treiman, Broderick, Tincof e Rodrigues [3], com base na importância da consciência fonêmica na aprendizagem da leitura e escrita, realizaram um trabalho, com três estudos, para examinar fatores lingüísticos que influenciavam o desempenho das crianças em tarefas de consciência fonêmica. O objetivo principal foi verificar se algumas classes de fonemas são mais fáceis para as crianças manipularem e identificarem do que outras. Um dos achados foi com relação a fonemas que diferem apenas no traço de sonoridade. Crianças em idade pré-escolar e de jardim de infância apresentaram mais dificuldades em diferenciar palavras e sílabas que diferem apenas em consoantes surdo/sonoras do que nas que diferem em ponto de articulação e traço de sonoridade.

Outra causa que muitas vezes leva ao fracasso escolar é a deficiência mental, cuja definição atual, proposta em 2002 pela AAMR, é a seguinte:

... uma incapacidade caracterizada por limitações significativas em ambos, funcionamento intelectual e comportamento adaptativo e está expresso nas habilidades sociais, conceituais e práticas. A incapacidade se origina antes da idade de 18 anos [4].

Todos os distúrbios de linguagem e de fala que ocorrem na criança com desenvolvimento normal podem ser observados na criança com deficiência mental, entretanto, nenhum deles lhe é particular [5].

As habilidades de comunicação dizem respeito à compreensão e expressão de informações por meios de

comportamentos simbólicos como palavras faladas, escritas, linguagem de sinais, ou ainda de compreender uma solicitação, uma emoção, um comentário, um questionamento, e escrever textos de maneira espontânea e criativa. Essas habilidades têm relação com as habilidades acadêmicas funcionais [5], que se referem à aprendizagem dos conteúdos curriculares propostos pelo sistema educacional, como ler, escrever, calcular, obter conhecimentos científicos e sociais relativos a diversos temas, que permitam maior autonomia na vida. Segundo Brauner [6], muitas crianças com deficiência mental possuem um vocabulário reduzido, considerável dificuldade em estruturar frases, em expressar e compreender uma idéia, trazendo dificuldades para o educador trabalhar com essa população em questões que envolvem esses domínios. Contudo, o autor esclarece, que nem todos os deficientes apresentam déficits de linguagem e sua deficiência de linguagem é desigual. Yavas et al. [7] relataram que a maioria das crianças que apresentam alterações importantes de comunicação tem pelo menos alguma dificuldade no nível fonológico, no domínio de regras e segmentos fonéticos, tendo como consequência, em muitos casos, a ocorrência de problemas de fala e linguagem. O comprometimento das habilidades de comunicação pode dificultar a aquisição da leitura, da escrita e do entendimento de conteúdos acadêmicos gerais, levando ao fracasso escolar.

As crianças com deficiência mental, geralmente, são lentas quanto à aquisição de habilidades de linguagem, o que afeta a forma com que aprendem conceitos importantes de seu contexto de vida. Em geral, essas crianças são lentas em adquirir habilidades de comunicação, mas com orientação podem ter melhor desenvolvimento nessa área.

Um problema que surge é a questão do acompanhamento individualizado. O ideal é o acompanhamento de um profissional altamente especializado, capaz de discriminar variações sutis nas vocalizações. Sem esse requisito há o risco do estabelecimento de comportamentos verbais inadequados e o enfraquecimento daqueles adequados. Entretanto, a dependência a esses profissionais, que não fazem parte da realidade da rede pública de ensino, torna restrita a intervenção terapêutica. Uma alternativa ao acompanhamento humano individualizado é um acompanhamento humano intermediado por ferramentas computacionais.

Este artigo aborda a questão do software educativo combinado com a tecnologia de reconhecimento de fala aplicado com grande potencial de aplicação na área de educação especial e está organizado da seguinte forma: na seção 2 é apresentado o software educativo MESTRE[®]; na seção 3 é apresentado o paradigma da equivalência de estímulos, base de sustentação do MESTRE[®] e sua relação com o desenvolvimento de linguagem; na

seção 4 é apresentada a teoria de reconhecimento de fala e também são apresentados alguns exemplos de aplicações de softwares educativos que utilizam a tecnologia de reconhecimento de fala; a seção 5 descreve os experimentos e os resultados obtidos e, na seção 6, são apresentadas as conclusões deste trabalho.

2. Software educativo MESTRE[®] e suas aplicações no desenvolvimento da linguagem

O propósito de um software educacional é oferecer meios que facilitem o aprendizado do indivíduo, além disso, deve ter uma interface apropriada e em sintonia com a faixa etária do público alvo, além de ser confiável, consistente e de oferecer mecanismos de auxílio caso o usuário apresente dificuldades ou diante do previsto [12]. Adicionalmente, o software deve apresentar validade científica e social. A escolha do software MESTRE[®] justifica-se na medida em que número grande de pesquisas sobre aquisição de habilidades lingüísticas variadas foi realizado com o mesmo.

2.1 O software MESTRE[®] – Automatização do Paradigma de Equivalência de Estímulos

O software MESTRE[®] [14] é um programa automatizado para a apresentação de estímulos visuais e auditivos, organização das sessões, registro e análise preliminar de respostas e ensino de relações conceituais.

Em um estudo realizado com quatro softwares educacionais, sendo dois de ensino e pesquisa e dois dos programas comerciais mais vendidos, Abreu [16] (2001) concluiu que o MESTRE[®] foi o software que preencheu com os maiores valores os testes de avaliação aos quais os programas foram submetidos.

O MESTRE[®] é baseado em tarefas de escolha de acordo com o modelo (MTS¹) e no paradigma da equivalência de estímulos (descritos abaixo). Este procedimento consiste na apresentação de um estímulo modelo seguido pela apresentação de dois ou mais estímulos de comparação, sendo a escolha correta aquela arbitrariamente relacionada ao modelo, se o emparelhamento for arbitrário, ou o estímulo igual, se o emparelhamento for de identidade. Em pesquisas da área de análise do comportamento, durante a fase de ensino, as respostas de escolha consideradas corretas são seguidas de um estímulo reforçador e as incorretas apenas do intervalo inter-tentativas. Em situações de testes para emergência de novos repertórios, usualmente não se apresentam conseqüências. No

MESTRE[®] o usuário pode escolher se apresenta ou não conseqüências diferenciais para escolhas corretas e incorretas. Ao final de uma seqüência de tentativas o aprendiz é levado a responder condicionalmente diante das escolhas. O estímulo modelo se torna um estímulo condicional e os estímulos comparação se tornam estímulos discriminativos para a resposta de escolha. Os estímulos podem ser escolhidos a critério do educador, dentre aqueles que melhor possam alcançar os objetivos da sua programação de ensino. As aplicações praticas do MESTRE[®] já foram realizadas na área de ensino de várias habilidades, dentre as quais, leitura e escrita, matemática e LIBRAS (Língua Brasileira de Sinais) [13,17,18].

2.2 Vantagens do uso de softwares educativos em aplicações pedagógicas

Goyos e Freire [11], destacam aspectos considerados vantajosos no uso do computador para o desenvolvimento da linguagem em ambiente escolar:

- Precisão. Tanto o material apresentado, quanto as respostas do aprendiz, se desejável, podem ser mantidos constantes para o uso de diferentes educadores e para quaisquer assuntos, ou aulas. Para que isso seja possível, é preciso que os elementos componentes da aprendizagem sejam claramente especificados pelos educadores.
- Eficiência. Apresentações sucessivas de exercícios/tarefas. Em uma única tela o educador pode programar lições/tarefas para uma ou mais sessões de ensino. O registro da interação do aprendiz com o programa é feito automaticamente, sem que o educador tenha que se envolver diretamente com isso. Assim, o tempo do educador pode ser gasto atendendo a outras necessidades do aluno, ou de outros alunos. Os alunos podem, com alguma experiência, trabalhar independentemente. Os resultados do trabalho do aluno podem ser impressos imediatamente após a conclusão das atividades, eliminando muitas tarefas para o professor. A análise e interpretação dos resultados podem também ser facilitada.
- Eliminação de variáveis irrelevantes. Em qualquer tarefa, quando utilizada para fins de avaliação do conhecimento do repertório comportamental ou conhecimento do aluno, o resultado deve refletir que o aluno esteja sob a influência do conteúdo da tarefa. Outras possíveis fontes potencialmente indesejáveis de influência tais como postura do educador; variações temporais ou espaciais devem ser eliminadas.

3. A Linguagem e o Paradigma de Equivalência de Estímulos

Uma das abordagens da Análise Experimental do Comportamento que tentam explicar a origem e o desen-

¹ Do inglês *matching-to-sample*

volvimento da linguagem é conhecida como o paradigma de equivalência de estímulos. Uma das razões pelas quais esse paradigma tem despertado o interesse dos pesquisadores diz respeito à análise do significado. Segundo de Rose [8]:

... dizer que uma palavra tem um significado implica em que esta palavra é um estímulo equivalente a um conjunto de estímulos, que correspondem a objetos, eventos, qualidades ou ações. Esta classe de estímulos a que a palavra se tornou equivalente é o seu significado. (p. 294).

A formação da linguagem pode ser vista como relações entre classes de estímulos e classes de respostas. Classes de estímulos podem ser formadas por similaridade física ou podem ser estabelecidas arbitrariamente entre estímulos fisicamente diferentes. As classes de estímulos arbitrariamente estabelecidas, são formadas através das relações que elementos dessas classes mantêm entre si. No universo da linguagem essas relações podem ser estabelecidas entre a palavra oral, a palavra impressa, e o objeto/ação/pessoa. O fenômeno de equivalência de estímulos é observado quando o ensino de relações entre conjuntos de estímulos resulta na aprendizagem não somente dessas relações, mas também na de outras relações não diretamente ensinadas. Assim, o ensino das relações entre palavra oral e figura e palavra oral e palavra impressa, as relações entre figura e palavra impressa emergem sem necessidade de ensino adicional.

Segundo Sidman e Tailby [9], para se observar a relação de equivalência de estímulos é necessário que se tenha três propriedades da matemática: reflexividade, simetria e transitividade. Por exemplo: considere o conjunto A de estímulos como sendo palavras orais, o conjunto B, como sendo figuras e, o conjunto C, como sendo palavras impressas. Reflexividade: é observada se o estímulo mantém uma consigo mesmo. o estímulo amostra A se relaciona ao estímulo A, o B ao B e o C ao C. Simetria: é observada quando emerge, sem ensino direto, a relação do estímulo amostra B com o estímulo A, após o ensino direto da relação do estímulo A com o B. Transitividade: é observada se após o treino das relações AB e AC, imediatamente ocorre a emergência das relações BC e CB. A metodologia nos estudos sobre equivalência de estímulos envolve um conjunto de relações condicionais treinadas diretamente, com conseqüências distintas para escolhas certas e erradas, e em seguida, a aplicação de testes para verificar a emergência de novas relações condicionais. Portanto, uma característica relevante da formação de classes de equivalência é a economia que traz para o planejamento de ensino, já que o ensino de algumas relações possibilita a emergência de outras sem a necessidade de treino [9]. Aos conjuntos de estímulos referentes a palavra oral, figura e palavra impressa, podem também ser acrescentados conjuntos de respostas

como produção oral, construção de anagramas e soletração, compõem o repertório básico para a aprendizagem de leitura e de escrita, e possibilitam o ensino de relações condicionais, representadas na Figura 1 [10,11]. Essas relações são fundamentais para englobar todas as relações básicas presentes no ensino de linguagem. A leitura, por exemplo, é entendida não só como a relação entre a palavra oral e a palavra impressa (leitura receptiva oral), mas também como a relação entre a palavra impressa e a produção oral (leitura expressiva oral). Mas esta relação não está completa se as relações entre palavra oral e figura (compreensão oral) e figura e produção oral (reconhecimento e nomeação) estiverem ausentes. Mas, sob o ponto de vista da leitura, é também fundamental que as relações simétricas entre figura e palavra impressa (leitura com compreensão) estejam presentes.

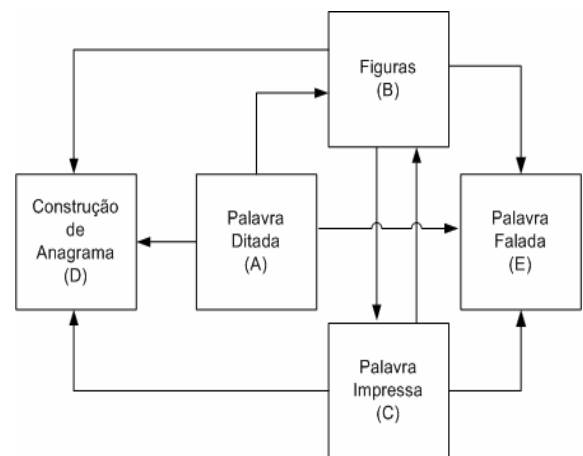


Figura 1- Relações presentes no ensino de habilidades de escrita e leitura (baseado em [11])

As relações são usualmente examinadas através de tarefas de MTS. As setas ligando os quadros apontam os estímulos usados como modelo para aqueles usados como comparação. A tarefa A-B refere-se às relações entre os nomes (modelo) e as figuras (comparação) que representam os nomes. A tarefa A-C refere-se às relações palavra (modelo) – nome (comparação), também designada leitura receptiva. A tarefa A-E, reconhecimento oral-auditivo, refere-se às relações entre palavra falada (modelo) – produção oral (resposta). A tarefa B-E refere-se às relações entre figura (modelo) – palavra falada (resposta), chamada nomeação. A tarefa C-E refere-se à relação entre palavra impressa (modelo) – produção oral (resposta) é designada leitura expressiva oral. As tarefas B-C, que referem-se às relações entre figura (modelo) – palavra impressa (escolha) e C-B, que referem-se às relações entre palavra impressa (modelo) – figuras (comparação), são indicativas de leitura com compreensão. As relações que compõem os repertórios da leitura e da escrita têm características distintas e são analisadas como unidades funcionais, podendo existir independentemente

das demais, sendo, no entanto, passíveis de integração. Se trabalhadas de forma integrada, podem facilitar a aquisição de repertórios em leitura e escrita [10,11].

Pesquisas têm sido realizadas para avaliar programas computadorizados nas questões relacionadas à linguagem, principalmente aquelas que podem comprometer a aprendizagem da leitura e escrita. As pesquisas que se seguem são ilustrativas.

Zuliani [13] utilizou o MESTRE[®][14] em um estudo com o objetivo de verificar se o procedimento de escolha de acordo com o modelo com resposta construída (CRMETS) com emissão de ecóico facilita a aprendizagem de palavras complexas, isto é, que têm em sua composição sílabas com dificuldades da língua, como dígrafos e grupos consonantais, e se o tempo decorrente entre os treinos e os testes comportamentais influencia na aprendizagem das palavras. Participaram do estudo seis crianças com dificuldades de aprendizagem e uma com diagnóstico de deficiência mental. Em todas as relações condicionais, na presença de estímulos modelos visuais o participante “via – ouvia – repetia” e na presença de estímulos auditivos, “ouvia – repetia”, seguindo-se a esse comportamento a apresentação dos estímulos comparação. A emissão de operante ecóico neste estudo pode ter favorecido o controle pelo estímulo auditivo no treino de ditado, a emergência das relações condicionais previstas e a aprendizagem das palavras com dificuldades da língua.

Silva [15] realizou um estudo utilizando o procedimento CRMETS para melhorar a produção da fala em crianças com transtornos fonológicos com o objetivo de avaliar a eficácia e funcionalidade desse método na área da fonoaudiologia. Participaram três alunos do ensino fundamental com idade de sete a nove anos, com quadro de transtorno fonológico e distúrbio de leitura e escrita, com substituição do grafema sonoro pelo opositor surdo. Os estímulos experimentais, apresentados pelo software de pesquisa MESTRE[®] [14], foram palavras ditadas, figuras, palavras impressas e conjuntos de letras, referentes às oposições fonêmicas /p/ /b/; /t/ /d/; /k/ /g/; /f/ /v/; /s/ /z/; /ʃ/ /ʒ/, que compunham respectivamente os conjuntos A, B, C e D. Inicialmente testaram-se as relações palavra ditada - figura, palavra ditada – palavra impressa, palavra ditada – produção oral (ecóico), figura – produção oral (nomeação) e palavra impressa – produção oral (textual), para escolha da oposição com maior quantidade de acertos. Em seguida aos testes das relações foi aplicado o CRMETS. Nas relações onde as respostas eram a produção oral do participante (ecóico, nomeação, textual), os registros foram realizados manualmente. O estímulo discriminativo foi apresentado simultaneamente ao aparecimento de um quadrado branco na parte central e superior da tela do monitor do computador. O partici-

pante tocava o quadrado branco como sinal de observação e emitia a resposta oral. O experimentador, a partir do reconhecimento da fala do participante, fazia o registro acionando o cursor na parte inferior da tela. As respostas dessas relações testadas foram gravadas em fita cassete e ouvidas, em momentos diferentes, pelo experimentador e por outro profissional fonoaudiólogo que desconhecia os objetivos do estudo, para julgarem se as respostas eram corretas ou incorretas. Esses registros foram utilizados para o teste de fidedignidade. O desempenho do Participante 1 foi de 100% de acertos, com exceção da oposição t/d na relação AE e p/b na BE, que foi de 90% de acertos. O Participante 2 atingiu a média de 75% e o Participante 3 de 80%. Para todas as outras relações de treino os participantes atingiram 91% de acertos em três sessões consecutivas, ou 95% em duas consecutivas ou uma com 100%, demonstrando a eficácia do procedimento para intervenção fonoaudiológica. As sucessivas exposições a tarefas que levaram à produção oral (ecóico, nomeação, textual), associadas às de escolha de acordo com o modelo com resposta construída podem ter sido importantes para a obtenção dos resultados desejáveis.

4. Reconhecimento Automático de Fala e Softwares Educativos

O reconhecimento automático da fala (RAF) por máquinas tem sido um objetivo de muitas pesquisas, por mais de cinco décadas.

O processo de reconhecimento de fala baseia-se nos princípios da fonética, e fundamenta-se na análise e identificação de sons. Basicamente, a fala humana divide-se em pequenas partes denominadas fonemas. De maneira simplificada, os fonemas seriam os menores componentes da fala que podem afetar o significado de uma palavra, que têm sua manifestação oral através do som. Existem vários tipos de sons, e dentro dos diversos tipos, existem várias formas de utilização, que caracterizam os diversos sotaques e sons peculiares de cada idioma. De uma maneira geral, podemos identificar e classificar os sons como: vocálicos, fricativos, plosivos (oclusivos) e nasais. Os fonemas ou sons vocálicos são aqueles emitidos durante a pronúncia de uma vogal. As vogais são caracterizadas essencialmente pela ausência de obstrução no trato vocal, o ar passa livre pelo trato vocal, seja com entonação aberta ou não. Os sons fricativos se caracterizam por um estreitamento quando da passagem do ar, que produz uma fricção ao passar pela pequena abertura formada pelo órgão articulante. Os sons fricativos da língua portuguesa são: /v/, /z/, /ʒ/, que são sonoros e /f/, /s/ e /ʃ/, que são surdos. Os fricativos sonoros diferem dos seus surdos correspondentes apenas no traço de sonoridade: para os sonoros as pregas vocais são vibradas e para os

surdos não há vibração das pregas vocais. Os sons plosivos ou oclusivos recebem essa terminologia porque a fase mais importante de sua formação é a oclusão momentânea da passagem de ar. As consoantes oclusivas são /p/, /t/, /k/, que são surdas e as consoantes /b/, /d/ e /g/, que são sonoras. Para o /b/, o local onde a área vocal é fechada é nos lábios; para o /d/ é atrás dos dentes e para o /g/ é perto da úvula. Durante o período em que a área vocal é totalmente fechada, nenhum som é emitido pelos lábios. Nas consoantes nasais se combina o fechamento do canal bucal com uma posição rebaixada do véu palatino e uma passagem livre do ar pelo nariz. As nasais são /m/, /n/ e /ŋ/.

Um dos processos de reconhecimento de fala consiste em analisar e identificar os sinais ou formas de ondas produzidos pelos sons. Cada um dos sons descritos anteriormente possui uma forma de onda própria e característica. Uma onda é criada por uma vibração que pode ser periódica ou não periódica, dependendo de sua frequência, ou seja, do número de vibrações por unidade de tempo. Os sons vocálicos possuem uma característica de quase-periodicidade, o que facilita a sua compreensão e análise. Entretanto, os sons das consoantes fricativas e plosivas são extremamente complexos, pois são produzidos de uma maneira aleatória. Assim, ao observarmos (através de um diagrama da frequência x tempo) a forma de onda de um som fricativo, não é possível notar nenhuma periodicidade ou mesmo coerência no sinal, o que acarreta maior dificuldade na tarefa de analisá-lo.

A tarefa de reconhecimento automático de fala é bastante difícil e complexa e sofre influência de vários fatores, tais como:

- Traços dos fonemas: os fonemas de vogais e ditongos parecem ser mais fáceis de serem reconhecidos do que as consoantes fricativas e plosivas;

- Influências geográficas: são responsáveis pelos chamados regionalismos, provenientes dos falares ou dialetos locais. Suas manifestações são, geralmente, compreendidas e aceitas, contribuindo para o nivelamento das diferenças regionais. Os fonemas e as palavras podem sofrer influências regionais como diferenças de sotaques, fazendo com que sejam articulados de forma mais longa ou com modificações nas frequências. Um exemplo é o R do interior paulista comparado com o R do carioca;

- Ruídos: ruídos do ambiente como conversas, barulhos de carros, etc.;

- Fatores socioculturais: raça, profissão, idade, posição social, grau de escolaridade, podem também influenciar na fala, interferindo, por exemplo, na concordância nominal ou verbal de frases ou com relação ao uso de plural (as casas são bonita).

- Fatores de saúde: rouquidão, gripe, resfriados podem interferir na pronúncia, por exemplo, a pessoa resfriada faz com que toda a pronúncia perca a parte nasal.

A melhoria nos sistemas de reconhecimento automático da fala vem sendo alvo de muitos trabalhos na área. Os pesquisadores, influenciados pelo trabalho de Lee [19], passaram a utilizar HMMs para resolver o problema de reconhecimento automático de fala. Após este trabalho foi possível construir softwares com capacidade de reconhecimento de fala contínua e para grandes vocabulários sem a necessidade de especialização para um determinado indivíduo ou grupo de indivíduos.

Do ponto de vista de software educativo, existem várias aplicações de reconhecimento de fala em educação. As mais simples estão relacionadas com o aprendizado de idiomas estrangeiros (aprendizado de pronúncias). Do ponto de vista funcional, para aquisição de primeira língua, existe um grande potencial de aplicação de softwares educativos com reconhecimento de fala.

4.1 Software Mestre com RAF

O MESTRE[®] [14] utiliza os conceitos derivados da pesquisa em equivalência de estímulos. Assim, quando uma palavra falada é apresentada, o aluno deve escolher uma figura ou uma palavra impressa que esteja relacionada com o estímulo apresentado. O software apresenta uma interface gráfica bem adequada ao público ao qual se destina. Outro diferencial do MESTRE[®] é a possibilidade que ele oferece ao professor de criar tarefas (figura 2).



Figura 2 - Tela de criação de tarefas do MESTRE®

Através de tarefas, o professor cria situações onde o aluno deverá ser estimulado. Sua resposta será armazenada para, posteriormente, poder ser analisada. Desta forma, é possível aplicar, simultaneamente um conjunto de estímulos a vários alunos e armazenar suas respostas. Cada aluno terá seu desempenho acompanhado e poderá obter um conjunto particular de tarefas para aprimorar seus conhecimentos. A padronização na aplicação dos estímulos e a correta captura das informações da tarefa executada pelo aluno são características importantes do MESTRE®. Apesar do potencial de aplicação prática do paradigma de equivalência de estímulos para aquisição de linguagem e, principalmente, para o ensino de leitura, ainda nenhum programa específico com essas características foi desenvolvido. Praticamente todas as vantagens de desenvolvimento de sistemas informatizados para o ensino de linguagem se perdem quando a avaliação da leitura, por exemplo, deve ser feita pessoalmente, sem o recurso informatizado. Em contraste, o potencial de uso prático de sistemas informatizados para o ensino de leitura aumenta exponencialmente se o recurso de reconhecimento de fala for introduzido.

Com alguma adaptação foi possível fazer com que o MESTRE® incorporasse também a captura de outro tipo de resposta: a resposta falada. Assim, agora é possível apresentar um estímulo (a figura abacaxi, por exemplo) e aguardar que o aluno emita o som correspondente.

5. Experimentos e Resultados

O objetivo dos experimentos foi avaliar a possibilidade de fazer com que o software MESTRE® fosse capaz de atuar na equivalência de estímulos também para os casos em que os aprendizes emitissem sons. No presente estudo, para efeito de avaliação do sistema, foram introduzidos somente os estímulos representados por figura e som correspondente, o que caracteriza, respectivamente, reconhecimento de figura e de som, mas não de leitura. A ausência do teste de leitura, caracterizado pela apresentação do estímulo constituído pela palavra escrita, justificase, no caso do presente estudo, por evitar a introdução de uma variável adicional que poderia confundir os efeitos do reconhecimento de fala.

O MESTRE® foi, portanto, adaptado para atuar em conjunto com o software de reconhecimento de fala, na plataforma Windows XP, com atualização de Xtras, Script e Interface. Foi necessário acrescentar a função da barra de espaço para Amostra Som e um arquivo "Reconhecer" que mostra o som que deve ser reconhecido. O software de reconhecimento de fala utilizado é um software comercial, desenvolvido pela empresa *Nuance*

Communications Inc [21], que concedeu sua utilização neste trabalho. Esse software é um sistema baseado em Hidden Markov Models (ferramenta estatística/estocástica) que, em fase de treinamento é exposto a um grande conjunto de evidências acústicas de fonemas de um determinado idioma, normalmente com variações regionais e de gênero, que possibilita que o sistema "aprenda" as distribuições estatísticas/estocásticas dos sons dada a população da "fase de treinamento" em questão. Como o propósito do software é comercial, o público alvo do treinamento foram homens e mulheres adultos.

Concluída as adaptações dos programas, os indivíduos foram selecionados, todos residentes em um município de pequeno porte do interior paulista. Participaram 110 indivíduos, distribuídos em quatro grupos:

Grupo 1 (G1): Dez estudantes universitários, sendo 5 do sexo masculino e 5 do sexo feminino. Os critérios de inclusão para esse grupo foram de terem dezoito anos ou mais, com voz e pronúncia desenvolvidas, estarem cursando a universidade, e emitirem a articulação e pronúncia das palavras de forma correta. Quatro estudantes eram do curso de administração de empresa, dois de pedagogia, dois de psicologia, um de música e um de fisioterapia.

Grupo 2 (G2): Sessenta indivíduos com desenvolvimento normal, sendo 20 na faixa etária de 4 anos a 4 anos e 11 meses, sendo nove do sexo masculino e onze do sexo feminino, 20 na faixa etária de 5 anos a 6 anos e 11 meses, sendo dez do sexo masculino e dez do sexo feminino, que freqüentam escola de educação infantil e 20 sujeitos de 7 anos a 7 anos e 11 meses, sendo onze do sexo masculino e nove do sexo feminino, que freqüentam escola de ensino fundamental. A escolha dessas idades se deve ao fato de que a partir dos 4 anos a fala da criança normalmente já é inteligível para pessoas que não fazem parte do seu ambiente social imediato [1], e aproximadamente aos 7 anos de idade a criança, normalmente, já adquiriu todos os fonemas de uma língua. Os indivíduos foram selecionados por suas professoras, que utilizaram como critério fala bem articulada, bom nível de comunicação e socialização e desempenho pedagógico plenamente satisfatório para suas idades.

Grupo 3 (G3): Vinte indivíduos com diagnóstico de deficiência mental, onze do sexo masculino e nove do sexo feminino, na faixa etária de 7 a 13 anos 6 meses, sendo que 16 freqüentavam uma escola especial e 4 incluídos em escola regular. Essa faixa etária foi escolhida por ser geralmente dentro desse período que se trabalha a alfabetização dessas crianças que apresentam um atraso das etapas de desenvolvimento cognitivo. Todos os indivíduos desse grupo passaram por avaliação de nível mental (Escala de Maturidade Mental Colúmbia). Todos os indivíduos, além do déficit intelectual, também apresentaram defasagens na área adaptativa de habilidades acadê-

micas e de comunicação: nenhum deles sabe ler e escrever.

Grupo 4 (G4): Vinte indivíduos com diagnóstico de transtorno fonológico, onze do sexo masculino e nove do sexo feminino, na faixa etária de 7 a 10 anos, que frequentam escolas de ensino fundamental no município de São Manuel e/ou são atendidos pelo setor de fonoaudiologia do Centro de Saúde Municipal. Todos os indivíduos passaram por avaliações fonoaudiológicas e foram diagnosticados como tendo transtorno fonológico, ou por substituições, omissões ou distorções de fonemas, ou ainda a combinação desses transtornos. Dezesete indivíduos deste grupo estão no ensino fundamental, de 1ª a 4ª séries, e, segundo suas professoras e/ou fonoaudiólogas que os atendem, todos apresentam problemas pedagógicos relacionados à escrita.

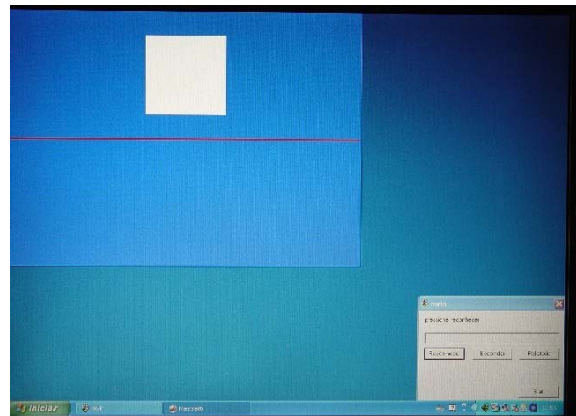
Todos os participantes menores de 18 anos apresentaram autorização dos pais ou de seus responsáveis. A escolha de diferentes populações objetivou maior generalização dos resultados, e conseqüentemente, maior fidelidade.

O trabalho foi desenvolvido em quatro diferentes salas, mantendo características semelhantes entre elas, como silêncio, boa iluminação, mesa, cadeiras e um computador, que foi o mesmo para todos os indivíduos. As salas normalmente tinham medidas próximas a 5m x 4m. Os equipamentos foram um microcomputador tipo notebook com sistema Microsoft Windows XP, CPU 2.00 GHz, 240 MB de RAM, incluindo teclado, monitor 15", impressora, mouse, o software MESTRE® (adaptado), uma lâmpada com interruptor, montada sobre base de madeira no formato de cubo, medindo aproximadamente 10 cm de lado, situada ao lado do notebook, microfone e fones de ouvidos acoplados. Foram usados no trabalho estímulos auditivos (palavra falada) e visuais (figuras), escolhidos considerando-se a aquisição de fonemas esperada para cada faixa etária dos participantes, tendo por base o teste ABFW de Linguagem Infantil. Os estímulos auditivos foram as palavras ditadas e os visuais foram figuras correspondentes a essas palavras, que por convenção foram respectivamente denominados A e B, e a resposta do indivíduo, produção oral, foi denominada E. Os estímulos foram divididos pelos Grupos de Estímulos I, II e III, respectivamente, GE I, GE II e GE III. O GE I foi composto por palavras cujos fonemas são normalmente presumíveis de aquisição até os quatro anos de idade, como pato, bola, dente, vaso. O GE II foi composto por palavras cujos fonemas são normalmente presumíveis de aquisição até os cinco anos de idade, como chinelo, peixe e jacaré. O GE III foi formado por palavras compostas por fonemas cujas discriminações são normalmente esperadas até os sete anos de idade, como gato, gado, faca e vaca (Tabela 1).

Os estímulos auditivos (palavras faladas) foram gravados por uma fonoaudióloga. As figuras usadas fazem parte do acervo de imagens do software MESTRE®, pelo qual os estímulos foram apresentados, em um quadrado azul, medindo aproximadamente 20 cm x 15 cm, que ficou na parte superior esquerda da tela do computador.

A resposta do indivíduo, uma elocução, foi reconhecida como certa ou errada pelo software de reconhecimento de fala, cujo ícone, medindo aproximadamente 8 cm x 6 cm, ficou localizado na parte inferior direita da tela (Figura 3).

Figura 3. Representação gráfica da tela do computador tal qual vista pelo participante no momento da apresentação de um estímulo auditivo na posição de estímulo modelo.



dor tal qual vista pelo participante no momento da apresentação de um estímulo auditivo na posição de estímulo modelo.

Tabela 1. Palavras utilizadas no estudo, que foram atribuídas a três grupos, de acordo com a faixa etária de sua presumível aquisição.

Grupos de estímulos					
Grupo I	Grupo II	Grupo III			
(GI)	(GII)	(GIII)			
pato,	bola,	osso,	pássaro,	bote,	pote,
dente,	caracol,	sapato,	chinelo,	gato,	gado,
galinha,	régua,	calça,	peixe,	cola,	gola,
sofá,	violão,	jacaré,	queijo,	faca,	vaca,
vaso,	abelha,	televisão,	tesou-	lousa,	louça,
banana,	mala,	ra, bolsa,	serrote,	xis,	giz,
caminhão,	rato,	sino,	avião	flor,	zebra,
leão,	arara,			livro,	abajur,
aquário				biscoito,	lápiz,
				árvore	

Cada um dos grupos de indivíduos foi exposto a um total de oitenta repetições por palavra, sendo quarenta em tarefas da relação AE e quarenta na relação BE. As tare-

fas tinham os números de tentativas referentes à quantidade de palavras (relação AE) ou quantidade de figuras (relação BE) do grupo de estímulos correspondente.

O número de tentativas para as tarefas referentes às relações AE e BE de GE I, GE II, e GE III foi, respectivamente, 16, 15 e 20. A ordem de apresentação dos estímulos foi randomizada. Para todos os indivíduos iniciou-se o trabalho por uma tarefa da relação AE, do GE I.

Os dez participantes do G1 foram expostos a oito repetições por palavras cada um, quatro nas relações AE, e quatro nas relações BE, em um total de vinte e quatro tarefas, sendo: na relação AE quatro do grupo de palavras I, quatro do grupo de palavras II, quatro do grupo de palavras III; na relação BE quatro do grupo de figuras I, quatro do grupo de figuras II e quatro do grupo de figuras III.

Os vinte participantes do G2, na faixa etária de 7 anos a 7 anos e 11 meses, foram expostos a quatro repetições por palavras cada um, duas nas relações AE, e duas nas relações BE, em um total de doze tarefas, sendo: na relação AE duas do GE I, duas do GE II e duas do GE III; na relação BE duas GE I, duas do GE II e duas do GE III.

Os vinte participantes do G2, na faixa etária de 5 anos a 6 anos e 11 meses, foram expostos a quatro repetições por palavras cada um, duas nas relações AE, e duas nas relações BE, em um total de oito tarefas, sendo: na relação AE duas do GE I e duas do GE II; na relação BE duas GE I e duas do GE II. Estes indivíduos não foram expostos às tarefas do grupo de estímulos III, por serem compostos de palavras cujas aquisições não são esperadas para esta faixa etária.

Os vinte participantes do G2, na faixa etária de 4 anos a 4 anos e 11 meses, foram expostos a quatro repetições por palavras cada um, duas nas relações AE, e duas nas relações BE, em um total de quatro tarefas, sendo: na relação AE duas do GE I e na relação BE duas do GE I. Estes sujeitos não foram expostos às tarefas do GE II e do GE III, por serem compostos de palavras cujas aquisições não são esperadas para esta faixa etária.

Os vinte participantes do G3 foram expostos a quatro repetições por palavras cada um, duas em cada uma das relações, em um total de doze tarefas, sendo: na relação AE duas do GE I, duas do GE II, duas GE III; na relação BE duas do GE I, duas do GE II e duas do GE III.

Os vinte participantes do G4 foram expostos a quatro repetições por palavras cada um, duas nas relações AE, e duas nas relações BE, em um total de doze tarefas, sendo: na relação AE, duas GE I, duas do GE II, duas do GE III; na relação BE, duas GE I, duas do GE II e duas do GE III.

Em todas as tarefas foram intercaladas, uma da rela-

ção AE e uma da relação BE, do GE I, começando com uma tarefa da relação AE.

Os indivíduos do G2, receberam carimbos com motivos infantis e os do G3 e G4, canetas perfumadas, após a participação.

5.1 Resultados do G1

A média de elocuições reconhecidas como pronunciadas corretamente, entre os grupos de estímulos, foi de 90,51% na relação AE e de 90,49% na relação BE ou de 90,50% no geral. Separados por sexo obteve-se, na média geral, 90,61% para o sexo masculino, e 89,39% para o sexo feminino. Os resultados mostraram um alto índice de elocuições reconhecidas como corretas no GE I, tanto para estímulos auditivos como para os visuais. As tarefas desse grupo de estímulos eram compostas de dezesseis tentativas cada, totalizando 128 elocuições por sujeito e 1280 no grupo. A porcentagem de elocuições reconhecidas como pronunciadas corretamente foi de 94,55% nas tarefas das relações AE e 94,8% nas tarefas das relações BE. As palavras com menores índices de elocuições reconhecidas como pronunciadas corretamente no GE I foram “bola”, que testou o fonema /b/, que é oclusivo e sonoro, com reconhecimento de 86,25%, com 69 elocuições reconhecidas como certas e 11 como erradas e a palavra “violão”, que testou o fonema /v/, fricativo e sonoro, com o mesmo índice de elocuições reconhecidas como pronunciadas corretamente da palavra “bola”. Todas as outras palavras tiveram índice de reconhecimento acima de 90%, isto é, foram reconhecidas como elocuições corretas.

No GE II, o índice de elocuições reconhecidas como corretas foi de 94,1% nas tarefas das relações AE e 93,8% nas relações BE. As tarefas desse grupo de estímulos eram compostas de quinze tentativas cada, totalizando 120 elocuições por indivíduo e 1200 no grupo. A palavra que apresentou a menor porcentagem de reconhecimento foi “sino”, que avaliou o fonema /i/, nasal e sonoro, com 73,75% de acertos. Todas as outras palavras tiveram índice de elocuições reconhecidas como pronunciadas corretamente acima de 90%.

No GE III, a margem de elocuições reconhecidas como pronunciadas corretamente foi menor em relação aos outros grupos de estímulos, 82,87% de elocuições reconhecidas como corretas nas relações AE e nas relações BE. Neste grupo algumas palavras tiveram uma baixa porcentagem de reconhecimento; “bote”, que testou o fonema /b/, que é oclusivo e sonoro, foi reconhecida em 62,5% das elocuições como pronunciadas corretamente, “pote”, que testou o fonema /p/, oclusivo e surdo, foi reconhecida em 45% das elocuições como tendo sido pronunciadas corretamente, “gola” com 35% de elocuições reconhecidas como pronunciadas corretamente; “vaca” que testou o fonema /v/ obteve 70% de elocuições

reconhecidas como pronunciadas corretamente e “livro” que avalia o /r/ no grupo consonantal, com 77,5% de elocuições reconhecidas como pronunciadas corretamente. Todas as outras palavras tiveram média de elocuições reconhecidas como pronunciadas corretamente acima de 87%. As tarefas do GE I e GE II tiveram maior número de elocuições consideradas corretas do que as do GE III, embora a maioria das palavras do GE III tivessem alto índice de elocuições reconhecidas como pronunciadas corretamente. No total das tarefas, cada um dos indivíduos emitiu 408 elocuições como respostas aos estímulos apresentados nas diferentes tarefas. Neste grupo o número total de respostas (elocuições) foi de 4080.

5.2 Resultados do G2

Os participantes do G2 foram divididos em três grupos, por faixas etárias. A apresentação dos resultados será iniciada pelo grupo de 7 anos a 7 anos e 11 meses. Esses participantes foram expostos a quatro repetições de cada estímulo, duas na relação AE e duas na relação BE, o que totalizou 80 elocuições de cada estímulo. Este grupo foi exposto a todos os grupos de estímulos (GE I, GE II e GE III).

As tarefas do GE I eram compostas de dezesseis tentativas cada, totalizando 64 elocuições por indivíduo e 1280 no grupo. A porcentagem de elocuições reconhecidas como corretas foi de 79,97% nas tarefas das relações AE e 78,25% nas tarefas das relações BE. As palavras com menores índices de elocuições reconhecidas como pronunciadas corretamente no GE I foram “bola”, que testou o fonema /b/, que é oclusivo e sonoro, com porcentagem de 45% de elocuições reconhecidas como pronunciadas corretamente, com 36 elocuições reconhecidas como certas e 44 como erradas; a palavra “dente”, que testou o fonema /d/, oclusivo e sonoro, com a porcentagem de 47,5% de elocuições reconhecidas como pronunciadas corretamente, com 38 reconhecidas como certas e 42 como erradas e a palavra “arara”, que testou o fonema /r/, sonoro e vibrante, com porcentagem de 66,25% de elocuições reconhecidas como pronunciadas corretamente. Todas as outras palavras tiveram índice de elocuições reconhecidas como pronunciadas corretamente acima de 70%.

No GE II, a porcentagem de elocuições reconhecidas como corretas foi de 85,12% nas tarefas das relações AE e de 81,75% nas relações BE. As tarefas desse grupo de estímulos eram compostas de quinze tentativas cada, totalizando 60 elocuições por indivíduo e 1200 no grupo. As palavras que apresentaram a menor porcentagem de reconhecimento foram “sino” e “serrote”. “Sino” avaliou o fonema /i/, nasal e sonoro, com porcentagem de 61,19% de acertos e “serrote” com porcentagem de 54,26% de acertos. Todas as outras palavras tiveram

índice de reconhecimento acima de 77%.

No GE III, a margem de elocuições reconhecidas como corretas foi menor em relação aos outros grupos de estímulos: 67,75 % nas relações AE e 68% BE. Neste grupo algumas palavras tiveram uma baixa porcentagem de elocuições reconhecidas como pronunciadas corretamente; “bote”, que testou o fonema /b/, que é oclusivo e sonoro, foi reconhecida em 48 elocuições, ou em 60% das elocuições; “pote”, que testou o fonema /p/, oclusivo e surdo, foi reconhecida em 22 elocuições, ou 27,5% de elocuições reconhecidas como pronunciadas corretamente; “vaca” que testou o fonema /v/ obteve 40% de elocuições reconhecidas como pronunciadas corretamente; “livro” que avalia o grupo consonantal com /r/ intercalado, 36,25% e a palavra “zebra”, avalia o grupo consonantal com /r/ intercalado, com 41,25% de elocuições reconhecidas como pronunciadas corretamente. Todas as outras palavras tiveram média de elocuições reconhecidas como pronunciadas corretamente acima de 70%. O grupo de estímulos que teve melhor índice de acertos foi o GE II, tanto na relação AE, como na relação BE. No total das tarefas, cada um dos indivíduos emitiu 204 elocuições como respostas aos estímulos apresentados nas diferentes tarefas. Neste grupo, o número total de respostas (elocuições) foi de 4080.

Os participantes do G2, na faixa etária de 5 anos a 6 anos e 11 meses, realizaram as tarefas dos GE I e GE II e foram expostos a quatro repetições de cada estímulo, duas na relação AE e duas na relação BE, o que totalizou 80 elocuições de cada estímulo. As tarefas foram intercaladas, uma AE e uma BE do mesmo grupo de estímulos.

As tarefas do GE I eram compostas de dezesseis tentativas cada, totalizando 64 elocuições por indivíduo e 1280 no grupo. A porcentagem de elocuições reconhecidas como corretas foi de 67,72% nas tarefas das relações AE e 66,82% nas tarefas das relações BE. As palavras com menores índices de reconhecimento no GE I foram “bola”, que testou o fonema /b/, que é oclusivo e sonoro, com porcentagem de 41,25% de elocuições reconhecidas como pronunciadas corretamente, 33 elocuições reconhecidas como certas e 42 como erradas; a palavra “dente”, que testou o fonema /d/, oclusivo e sonoro, com a porcentagem de 47,5% de elocuições reconhecidas como pronunciadas corretamente, com 38 reconhecidas como certas e 42 como erradas; a palavra “pato” testou o fonema /p/, oclusivo e surdo, com 47,5 % de elocuições reconhecidas como pronunciadas corretamente, sendo 38 reconhecidas como certas e 42 como erradas. Todas as outras palavras tiveram índice de elocuições reconhecidas como pronunciadas corretamente acima de 63%.

No GE II, nas tarefas das relações AE a porcentagem de elocuições reconhecidas como corretas foi de 66,7% e nas relações BE foi de 70,7%. As tarefas desse grupo de

estímulos eram compostas de quinze tentativas cada, totalizando 60 elocuições por indivíduo e 1200 no grupo. As palavras que apresentaram as menores porcentagens de elocuições reconhecidas como pronunciadas corretamente foram “sino”, que avalia o fonema /i/, nasal e sonoro, com porcentagem de 55% de elocuições reconhecidas como pronunciadas corretamente, “serrote”, que avalia o fonema /ʃ/, sonoro e vibrante, com porcentagem de 60% de elocuições reconhecidas como pronunciadas corretamente e “tesoura”, que avalia o fonema /t/, oclusivo e surdo, com 53,75% de elocuições reconhecidas como pronunciadas corretamente. Todas as outras palavras tiveram índice de elocuições reconhecidas como pronunciadas corretamente acima de 67%. O grupo de estímulos que teve melhor índice de elocuições reconhecidas como pronunciadas corretamente foi o GE II, na relação BE. No total das tarefas, cada um dos indivíduos emitiu 124 elocuições como respostas aos estímulos apresentados nas diferentes tarefas. Neste grupo de estímulos, o número total de respostas (elocuições) foi de 2480.

Os participantes do G2, na faixa etária de 4 anos a 4 anos e 11 meses, realizaram apenas as tarefas do GE I. Os indivíduos foram expostos a quatro repetições de cada estímulo, duas na relação AE e duas na relação BE, o que totalizou 80 elocuições de cada estímulo.

As tarefas do GE I eram compostas de dezesseis tentativas cada, totalizando 64 elocuições por indivíduo e 1280 no grupo. A porcentagem de elocuições reconhecidas como corretas foi de 54,4% nas tarefas das relações AE e 52,65% nas tarefas das relações BE. As palavras com menores índices de elocuições reconhecidas como pronunciadas corretamente no GE I foram “sofá”, que testou o fonema /s/, fricativo e surdo, com 37,5% de elocuições reconhecidas como pronunciadas corretamente, com 30 elocuições reconhecidas como certas e 50 como erradas; “bola”, que testou o fonema /b/, que é oclusivo e sonoro, com porcentagem de 35% de elocuições reconhecidas como pronunciadas corretamente, com 28 elocuições reconhecidas como certas e 52 como erradas; a palavra “dente”, que testou o fonema /d/, oclusivo e sonoro, com a porcentagem de 47,5% de elocuições reconhecidas como pronunciadas corretamente, sendo 38 elocuições reconhecidas como certas e 42 como erradas; a palavra “pato”, que testou o fonema /p/, oclusivo e surdo, com 48,75% de elocuições reconhecidas como pronunciadas corretamente, sendo 39 elocuições reconhecidas como certas e 41 como erradas; “abelha”, que testou o fonema /ʎ/, lateral e sonoro, com 46,25% de elocuições reconhecidas como pronunciadas corretamente, com 37 elocuições reconhecidas como certas e 43 como erradas. Todas as outras palavras tiveram índice de elocuições reconhecidas como pronunciadas corretamente a partir de 50%. No total das tarefas, cada um dos indivíduos emitiu 62 elocuições como respostas aos estímulos apresentados

nas diferentes tarefas. Neste grupo, o número total de respostas (elocuições) foi de 1240.

5.3 Resultados do G3

Os participantes do G3 foram expostos a oito repetições de cada palavra, quatro na relação AE e quatro na relação BE, o que totalizou 80 elocuições de cada palavra. A coleta foi iniciada pelo GE I. As tarefas desse grupo de estímulos eram compostas de dezesseis tentativas cada, totalizando 128 elocuições por indivíduo e 1280 no grupo. A porcentagem de elocuições reconhecidas como corretas foi de 65,92% nas tarefas das relações AE e 64,32% nas tarefas das relações BE.

As palavras com menores índices de elocuições reconhecidas como pronunciadas corretamente no GE I foram “bola”, que testou o fonema /b/, que é oclusivo e sonoro, com porcentagem de 57,5% de elocuições reconhecidas como pronunciadas corretamente, com 46 elocuições reconhecidas como certas e 34 como erradas; a palavra “dente”, que testou o fonema /d/, oclusivo e sonoro, com a porcentagem de 45% de elocuições reconhecidas como pronunciadas corretamente, sendo 36 elocuições reconhecidas como certas e 44 como erradas; a palavra “abelha” que testou o fonema /ʎ/, lateral e sonoro, com 52,5% de elocuições reconhecidas como pronunciadas corretamente, sendo 42 elocuições corretas e 38 incorretas; a palavra “arara”, que testou o fonema /r/, sonoro e vibrante, com porcentagem de 60% de elocuições reconhecidas como pronunciadas corretamente, sendo 48 elocuições corretas e 32 incorretas e “vaso”, que avaliou o fonema sonoro e fricativo /z/, com 61,25% de elocuições reconhecidas como pronunciadas corretamente, sendo 49 elocuições corretas e 31 incorretas. Todas as outras palavras tiveram índice de elocuições reconhecidas como pronunciadas corretamente acima de 63%.

No GE II, nas tarefas da relação AE, o índice de elocuições reconhecidas como corretas foi de 67,65% e nas da relação BE foi de 63,4%. As tarefas desse grupo de estímulos eram compostas de quinze tentativas cada, totalizando 60 elocuições por indivíduo e 1200 no grupo. As palavras que apresentaram as menores porcentagens de elocuições reconhecidas como pronunciadas corretamente foram: “pássaro”, que avalia o fonema /s/, surdo e fricativo, com 42,5% das elocuições reconhecidas como pronunciadas corretamente, 34 reconhecidas como corretas e 46 como incorretas; “tesoura”, que avalia o fonema /t/, oclusivo e surdo, com 46,25% de elocuições reconhecidas como pronunciadas corretamente, sendo 37 elocuições corretas e 43 incorretas; “televisão”, que avalia o fonema /t/, surdo e oclusivo, com 61,25% elocuições reconhecidas como pronunciadas corretamente, 49 elocuições corretas e 31 incorretas; “sino”, que avalia o fonema /i/, nasal e sonoro, e “bolsa” que avalia o arquifonema

lateral /L/, tiveram 58,75% elocuições reconhecidas como pronunciadas corretamente, e “serrote”, que avalia o fonema /ʃ/, sonoro e vibrante com 47,5% elocuições reconhecidas como pronunciadas corretamente. Todas as outras palavras tiveram índice de reconhecimento a partir de 65%.

No GE III, a porcentagem de elocuições reconhecidas como corretas foi menor em relação aos outros grupos de estímulos: 52,4% nas relações AE e 51,5% nas relações BE. Neste grupo, algumas palavras tiveram uma baixa porcentagem de elocuições reconhecidas como pronunciadas corretamente. Entre elas se destacaram as palavras “livro”, com 20% e “zebra”, com 25% de elocuições reconhecidas como pronunciadas corretamente, sendo que ambas avaliam o /t/ no grupo consonantal; “giz” com 26,25% de elocuições reconhecidas como pronunciadas corretamente; “bote”, que testou o fonema /b/, que é oclusivo e sonoro, foi reconhecida como pronunciada corretamente em 37 elocuições, ou 46,25%; “pote”, que testou o fonema /p/, oclusivo e surdo, foi reconhecida em 30 elocuições, ou 37,5% de elocuições reconhecidas; “vaca” que testou o fonema /v/, surdo e fricativo, obteve 33,75% de elocuições reconhecidas. Todas as outras palavras tiveram média de elocuições reconhecidas a partir de 50%.

O fato de este grupo de participantes apresentar índice de elocuições reconhecidas como pronunciadas incorretamente, bem maior que o do G2, apesar da média de idade ser mais elevada, se deve a vários fatores. Alguns dos indivíduos tiveram dificuldade em esperar o sinal para falar, falavam antes da hora, errando a tentativa. Outro ponto é que esses indivíduos articulam mal muitas palavras, por exemplo, “vião”, em vez de “avião”, “arve”, no lugar de “árvore”, “flô”, “fror” ou “for” em vez de “flor”, “borsa”, em vez de “bolsa”, “zeba”, em vez de “zebra”, “abeia” ou “belha” no lugar de “abelha”, “tevisão” em vez de “televisão”, entre outras. As palavras com encontro consonantais, que fazem parte do GE III, foram as que a maioria dos participantes erraram ao pronunciar, o que justifica a baixa porcentagem de acertos neste grupo de estímulos. No total das tarefas, cada um dos indivíduos emitiu 408 elocuições como respostas aos estímulos apresentados nas diferentes tarefas. Neste grupo, o número total de respostas (elocuições) foi de 4080.

5.4 Resultados do G4

Os participantes do G4, foram expostos a oito repetições de cada palavra, dos três grupos de estímulos, quatro na relação AE e quatro na relação BE, o que totalizou 80 elocuições de cada palavra.

A coleta foi iniciada pelo GE I. As tarefas desse grupo de estímulos eram compostas de dezesseis tentativas cada, totalizando 128 elocuições por indivíduo e 1280 no

grupo. A porcentagem de elocuições reconhecidas como corretas foi de 50,75% nas tarefas das relações AE e 52,35% nas tarefas das relações BE. Muitas palavras tiveram baixo índice de elocuições reconhecidas como corretas, provavelmente pela dificuldade de alguns dos indivíduos em articularem essas palavras. Entretanto, algumas das palavras menos reconhecidas, coincidiram com as também menos reconhecidas dos demais grupos de indivíduos, como “bola”, que testou o fonema /b/, que é oclusivo e sonoro, com reconhecimento de 21,25%, com 17 elocuições reconhecidas como certas e 63 como erradas; a palavra “dente”, que testou o fonema /d/, oclusivo e sonoro, com reconhecimento de 31,25%, com 25 elocuições reconhecidas como certas e 55 como erradas; a palavra “abelha” que testou o fonema /λ/, lateral e sonoro, com 26,25% de elocuições reconhecidas como pronunciadas corretamente, 21 como certas e 59 erradas, e a palavra “arara”, que testou o fonema /r/, sonoro e vibrante, com porcentagem de elocuições reconhecidas como pronunciadas corretamente de 60%, 48 como corretas e 32 como incorretas, e “vaso”, que avaliou o fonema sonoro e fricativo /z/, com 33,75% de elocuições reconhecidas como pronunciadas corretamente, 27 como corretas e 53 como incorretas. Todas as outras palavras tiveram índice de elocuições reconhecidas como tendo sido pronunciadas corretamente acima de 38,75%.

No GE II, nas tarefas das relações AE, a porcentagem de elocuições reconhecidas como corretas foi de 60,35% e nas relações BE foi de 58,42%. As tarefas desse grupo de estímulos eram compostas de quinze tentativas cada, totalizando 60 elocuições por indivíduo e 1200 no grupo. As palavras que apresentaram as menores porcentagens de elocuições reconhecidas foram: “pássaro”, que avalia o fonema /s/, surdo e fricativo, com 46,25% de acertos, 37 elocuições corretas e 43 incorretas; “tesoura”, com 51,25% de elocuições reconhecidas, sendo 41 elocuições corretas e 39 incorretas; “televisão”, que avalia o fonema /t/, surdo e oclusivo, com 48,75% de elocuições reconhecidas como pronunciadas corretamente, 39 elocuições corretas e 41 incorretas; “bolsa” que avalia o arquifonema lateral /L/, teve 42,25% de elocuições reconhecidas como pronunciadas corretamente, e “serrote”, que avalia o fonema /ʃ/, sonoro e vibrante com 41,25% de elocuições reconhecidas como pronunciadas corretamente. Todas as outras palavras tiveram índice de reconhecimento a partir de 60%.

No GE III, a porcentagem de elocuições reconhecidas como corretas foi menor em relação aos outros grupos de estímulos, 37,87% nas relações AE e 37,87% nas relações BE. No GE III algumas palavras tiveram uma baixa porcentagem elocuições reconhecidas como pronunciadas corretamente. Entre elas se destacaram as palavras: “gola”, que testou o fonema /g/, sonoro e oclusivo, com 11,25% de elocuições reconhecidas como pronunciadas

corretamente, o que corresponde a 9 elocuições reconhecidas como certas e 71 como erradas; “vaca” que testou o fonema /v/, surdo e fricativo, obteve 11,25% de elocuições reconhecidas como pronunciadas corretamente, sendo 9 elocuições reconhecidas como certas e 71 como erradas; “giz” com 17,5% de elocuições reconhecidas como pronunciadas corretamente; “bote”, que testou o fonema /b/, que é oclusivo e sonoro, foi reconhecida em 15 elocuições, ou 18,75% de elocuições reconhecidas; “pote”, que testou o fonema /p/, oclusivo e surdo, foi reconhecida em 23 elocuições, ou 28,75% de elocuições reconhecidas; “livro”, com apenas 22,5% de elocuições reconhecidas e “zebra”, com 13,75% de elocuições reconhecidas como pronunciadas corretamente, sendo que ambas avaliam o /r/ no grupo consonantal. Neste grupo de estímulos o índice de elocuições reconhecidas como pronunciadas corretamente foi muito baixo para a maioria das palavras, provavelmente por ser neste grupo que se encontram as palavras cujos fonemas a maior parte dos indivíduos apresentam dificuldades de articulação e/ou de discriminação, como nos casos das palavras nas quais os fonemas diferem apenas pelo traço de sonoridade. Um fato importante foi de os indivíduos realizarem as tarefas com entusiasmo, principalmente por tratarem de crianças tímidas, que resistem em falar na presença de pessoas que não fazem parte do seu cotidiano. Entretanto, todos realizaram as tarefas sem a menor resistência. No total das tarefas, cada um dos indivíduos emitiu 408 elocuições como respostas aos estímulos apresentados nas diferentes tarefas. Neste grupo o número total de respostas (elocuições) foi de 4080.

6. Conclusões

No total, os 110 indivíduos que participaram deste estudo, juntos, foram expostos a 20.040 estímulos, emitindo o mesmo número de elocuições como respostas aos estímulos.

Os resultados mostram que as palavras com menores índices de reconhecimento em todos os grupos são, na maioria, as mesmas. Na análise por grupos de indivíduos, as médias gerais, nas relações AE e BE juntas, de elocuições reconhecidas pelo software de reconhecimento automático de fala como corretas foram: 90,5% no G1; no G2, na faixa etária de 7 anos a 7 anos e 11 meses, foi de 76,81%, na faixa etária de 5 anos a 6 anos e 11 meses foi de 67,99% e na faixa etária de 4 anos a 4 anos e 11 meses foi de 53,53%; no G3 foi de 60,87% e no G4 foi de 49,60%.

Por ser composto por universitários, credita-se ao G1 o papel de referência para a confiabilidade da tecnologia avaliada. Por pronunciarem as palavras corretamente, infere-se que a porcentagem de erros no reconhecimento de elocuições deve-se a problemas na tecnologia de re-

conhecimento de fala, ou seja, é a margem de erro que o software apresenta no reconhecimento da fala em populações adultas. Sendo assim, pode-se dizer que para a maioria dos estímulos testados neste estudo, a tecnologia é confiável. Conseguiu captar as diferenças sutis de pronúncia que as pessoas apresentam normalmente nas suas elocuições, que são consideradas normais pela sociedade, como por exemplo, a palavra peixe, algumas pessoas pronunciam peixe, pêxe e outras peixi. Os grupos de estímulos I e II apresentaram índices de acertos elevados e maiores que o do G III. Estes resultados sugerem uma dificuldade maior da tecnologia em reconhecer algumas palavras do GE III. Entretanto, as palavras com maiores índices de erros dos GE I e GE II também sugerem dificuldades de reconhecimento da tecnologia. No reconhecimento das palavras, os resultados demonstraram que as palavras com menores índices de reconhecimento foram as oclusivas “bola”, com reconhecimento de 86,25%, “bote”, que testou o fonema /b/, que é oclusivo e sonoro, foi reconhecida em 50 elocuições, ou 62,5%; “pote”, que testou o fonema /p/, oclusivo e surdo, foi reconhecida em 36 elocuições, ou 45% de reconhecimento erradas e as fricativas “violão”, 86,25%, “sino”, com 73,75% de acertos, “vaca” que testou o fonema /v/ obteve 70% de reconhecimento. Esses resultados apoiam trabalhos na área com grupo de adultos, que demonstram índices altos de acertos [22].

O G2, por ser composto por crianças consideradas normais tanto no aspecto de desenvolvimento intelectual quanto no da fala, tem, assim como o G1, a função de parâmetro para as faixas etárias nele avaliadas. No grupo de indivíduos de 7 anos a 7 anos e 11 meses, o maior índice de acertos aconteceu no GE II, seguido do GE I e, com o menor porcentagem de acertos, o GE III. A média geral de acertos destes indivíduos foi abaixo das alcançadas pelo G1. Este resultado sugere dificuldades da tecnologia no reconhecimento das elocuições dessa faixa etária. Porém, algumas palavras tiveram alto índice de reconhecimento, o que demonstra que a dificuldade no reconhecimento para essa faixa etária do G2, não se generaliza para todas as palavras, muitas tiveram alto índice de reconhecimento.

O G3 foi exposto aos três grupos de estímulos, apresentando maior índice de acertos no GE II, seguido do GE I e do GE III, com o menor número de acertos. Em todos os grupos de estímulos a maior porcentagem de acertos foi nas relações AE, sugerindo uma melhor pronúncia das palavras na presença de um modelo auditivo adequado, demonstrando a importância que essa relação pode ter em procedimentos de ensino de fala nesta população, que, apesar de não ter transtorno fonológico, apresentou dificuldades em pronunciar palavras mais complexas.

Uma questão que pode merecer uma investigação

mais detalhada é com relação às palavras pronunciadas de forma incorreta. A dúvida é se alguns “erros” na pronúncia são por questões de ambiente sócio-cultural, ou se por imprecisão articulatória. O mais provável é que haja uma combinação desses fatores. As palavras “borsa”, “fror”, “frô”, “arve”, parecem ser pronúncias do ambiente cultural, enquanto “tevisão”, “zeba”, “igu”, “livo”, parecem mais um problema de origem articulatória ou fonológica, como dificuldade em manipular os segmentos fonéticos de algumas palavras, o que pode ocorrer em indivíduos com problemas nas habilidades de comunicação [7]. Outro fator importante está nos resultados individuais do G3. Nos grupos de estímulos em que alguns participantes não tinham dificuldades para pronunciar as palavras, as porcentagens de acertos foram bem altas, e relacionando esse resultado à idade média do grupo, que é de 9 anos e 8 meses, nota-se novamente a importância do fator idade cronológica na eficácia da tecnologia de reconhecimento da fala.

O G4 realizou todas as tarefas dos três grupos de estímulos. A média de idade dos indivíduos é de 8 anos e 7 meses. Assim como no G3, os resultados individuais do G4 são muito importantes. Demonstrem que o índice de acertos foi maior nos grupos compostos pelos estímulos com os fonemas que os indivíduos não apresentavam dificuldades para articular, sugerindo eficácia no reconhecimento da fala dessa população, na faixa etária avaliada neste estudo. Esse achado é muito importante na elaboração de tarefas de ensino para essa população, principalmente por educadores, que não possuem o mesmo treino auditivo dos profissionais da área de fonoaudiologia para discriminar alterações sutis nas elocuições dessas crianças, e precisam, constantemente, trabalhar com a questão da leitura e escrita, geralmente comprometida com muitos erros ortográficos, acarretando dificuldades pedagógicas relacionadas a atividades que exigem escrita e, muitas vezes, na compreensão da leitura e escrita.

Do ponto de vista pedagógico, um fator positivo e extremamente influente na questão da aprendizagem, foi a motivação de todos os indivíduos na realização das tarefas. Todos, dos G2, G3 e G4, sem exceção, participaram com muito entusiasmo e ao término perguntaram quando voltariam novamente. Os que participaram do estudo nas escolas, durante as aulas, contaram da experiência para os colegas nas salas de aula, como tendo sido muito agradável, o que fez com que os que não foram selecionados manifestassem interesse em participar. Dentre os principais avanços que o programa apresenta, encontra-se o de poder avaliar o comportamento do falante juntamente com o do ouvinte, fechando por completo o episódio verbal, dentro de uma sólida perspectiva analítico-experimental [20].

Do ponto de vista da tecnologia de reconhecimento de

fala e educacional, alguns fatores parecem ser relevantes: a) analisando os grupos de estímulos, observa-se que o GE II, independente de o estímulo ser auditivo ou visual, foi que teve maior porcentagem de acertos, o que indica maior confiabilidade para procedimentos educacionais; b) as palavras que tiveram menor índice de reconhecimento foram, a maioria delas, as mesmas em todos os grupos, o que indica problemas de ordem tecnológica no reconhecimento dessas palavras, portanto é necessário cautela para usá-las em procedimentos de ensino; c) as palavras com alto índice de acertos podem servir de referencial para escolhas de palavras em trabalhos de alfabetização com indivíduos que tenham dificuldades de leitura e escrita; d) a idade parece, pelos resultados, estar diretamente ligada à questão do reconhecimento. Uma hipótese que pode justificar este fator é a questão da frequência de voz, que é diferente em cada fase do desenvolvimento do ser humano. As frequências fundamentais das vozes masculinas de adultos podem variar de 80 a 150 Hz, as femininas de 150 a 250 Hz e as infantis encontram-se acima de 250Hz. Como o software NU-ANCE foi desenvolvido visando atender populações adultas, fica esclarecida a questão do reconhecimento de fala relacionada à idade cronológica. Os altos índices de reconhecimento no G1 nos permite sugerir que a tecnologia de reconhecimento de fala pode ser uma ferramenta que poderia ser utilizada em programas educativos direcionados à alfabetização de jovens e adultos.

7. AGRADECIMENTOS

Agradecemos a Nuance (através do Sr. Rodrigo Damhem) por ter nos cedido uma versão do software para a condução da pesquisa antes que uma licença fosse adquirida pela UFSCar. Também agradecemos aos ex-alunos Thalma e Guilherme que propiciaram as devidas alterações no software MESTRE, às fonoaudiólogas Leia Carmo e Rosa Acerra pela avaliação dos indivíduos com transtornos fonológicos e à psicóloga Cláudia Parenti pelas avaliações dos indivíduos com deficiência mental.

Referências

- [1] Mota, H.B. (2001). *Terapia Fonoaudiológica para os desvios fonológicos*. Rio de Janeiro: Editora Revinter Ltda.
- [2] Brasolotto, A.G. (1993). *Investigação diagnóstica de trocas entre fonemas sonoros e surdos e entre os grafemas correspondentes*. Dissertação de Mestrado, Programa de Pós-Graduação em Educação Especial, São Carlos: UFSCar.
- [3] Treiman, R., Broderick, V; Tincoff, R. & Rodri-

- guez, K. (1998). *Children's Phonological Awareness: Confusions between Phonemes that Differ Only in Voicing*. *Journal of the Experimental Child Psychology*, 68.3-21.
- [4] Luckasson, R. Borthwick-Duffy, S. Buntinx, W. H. E., Coulter, D. L., Craig, E. M.; Reeve, A.; Schalock, R. I., Snell, M. E., Spitalnik, D.M. E, Spreat, S., & Tassé, M. J. (2002). *Mental Retardation – Definition, Classification, and Systems of Supports*. 10ª ed. Washington (DC): American Association on Mental Retardation.
- [5] Almeida, M.A (2004). *Apresentação e Análise das Definições de Deficiência Mental Propostas pela AAMR – Associação Americana de Retardo Mental - no Período De 1908 A 2002*. *Revista de Educação (Campinas)*, Vol. 6, 33-48.
- [6] Brauner, R. A. & Brauner, F. (1989). *Distúrbio da fala e da linguagem dos deficientes mentais*. In: Launy, L.; Borel – Maissonny, S. *Distúrbios da linguagem, da fala e da voz na infância*. 2. ed. São Paulo: Roca. 121 a 131.
- [7] Yavas, M. Hernandorena, C.L.M. Lamprecht, R.R. (1992). *Avaliação Fonológica da Criança – Reeducação e Terapia*. Porto Alegre: Artes Médicas.
- [8] de Rose, J. C. (1993). *Classes de Estímulos: Implicações para uma Análise Comportamental da Cognição*. *Psicologia Teoria e Pesquisa*. Vol. 9, 283-303.
- [9] Sidman, M. & Tailby, W. (1982). *Conditional Discrimination VS. Matching to Sample: An expansion of the testing paradigm*. *Journal of the Experimental Analysis of Behavior*, 37. 5-22.
- [10] Stromer, R, Mackay, H. A, & Stoddard, L.T. (1992). *Classroom Applications of Stimulus Equivalence Technology*. *Journal of Behavioral Education*, 2, 225-256
- [11] Goyos, C., & Freire, A.F. (2000). *Programando Ensino Informatizado para Indivíduos Deficientes Mentais*. In: Manzini, E.J. (org.). *Educação Especial: Temas Atuais*. Marília: Unesp Marília Publicações. p. 57-73.
- [12] Squires, D. & Preece, J. – Usability and learning: Evaluating the potential of educational software. *Computers and Education*, vol. 27, n 1, pp 15-22, 1996.
- [13] Zuliani, G. (2007). *Aquisição e manutenção de comportamentos de leitura e fluência através de contingências de repetição oral e velocidade nos procedimentos de equivalência de estímulos*. Tese de Doutorado não publicada. Programa de Pós-graduação em Educação Especial, São Carlos: UFSCar.
- [14] Goyos, C., & Almeida, J.C.B. (1996). *MESTRE® (Version 1.0) [Computer software]*. São Carlos, Brasil: MESTRE® Software.
- [15] Silva, A.M.R.C. (2001). *O Efeito do Uso do CRMTS para Produção dos Fonemas Sonoros*. Dissertação de Mestrado, Programa de Pós-Graduação em Educação Especial, São Carlos: UFSCar.
- [16] Abreu, M.A.F.G. (2001). *Análise de recursos computacionais aplicados a pesquisa e ensino de leitura no Brasil*. Dissertação de mestrado não publicada. Universidade Mackenzie, São Paulo, Brasil.
- [17] Escobal, G., Rossit, R.A.S., Goyos, C. (2009). *Aquisição do conceito de número por pessoas com atraso no desenvolvimento intelectual*. *Psicologia em Estudo (Maringá)*.
- [18] Elias, N.C., Goyos, C., Saunders, M.D., & Saunders, R.R. (2008). *Teaching manual signs to adults with mental retardation using matching-to-sample procedures and stimulus equivalence*. *The Analysis of Verbal Behavior*, 24, 1-13.
- [19] Lee, K.F. (1988) *Large-Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System*. Ph.D. Thesis, CMU.
- [20] Horne, P. J., & Lowe, C. F. (2000). *Putting the Naming Account to the Test: Preview of an Experimental Program*. In *Experimental and Applied Analysis of Human Behavior*, Eds. Julian C. Leslie and Derek Blackman. Context Press, Reno: Nevada.
- [21] Nuance Communications Ltda – www.nuance.com
- [22] Ynoguti, C. A. ; Violaro, F. (2001) *Desenvolvimento de um conjunto de ferramentas para pesquisas em reconhecimento de fala*. *Revista Telecomunicações, Santa Rita do Sapucaí - MG*, v. 4, n. 2, p. 36-43.