

Abordagem colaborativa para apoio à avaliação do ensino de português

Luiz Carlos Carchedi¹, Eduardo Barrére¹, Jairo Francisco de Souza¹

¹Pós-graduação em Ciência da Computação
Universidade Federal de Juiz de Fora (UFJF) – MG – Brasil

{carchedi, eduardo.barrere, jairo.souza}@ice.ufjf.br

Abstract. *During the literacy process, orality is as important as written language. The large-scale evaluation of this process is fundamental for public bodies responsible for education. With the aid of automatic speech recognition (ASR) tools, it is possible to automate the evaluation of orality, which allows its use on a large scale; however, when automated, there are cases where the process can not define the quality of the reading performed. In such cases, it requires human involvement. The goal of this work is present a computational solution to support a orality evaluation in large scale. This work presents a collaborative solution to complement the automatic large scale evaluation of orality. The results presented reinforce the viability of the solution.*

Resumo. *Durante o processo de alfabetização, a fluência na fala é tão importante quanto a linguagem escrita. A avaliação em larga escala desse processo é fundamental para os órgãos públicos responsáveis pela educação. Com o auxílio de ferramentas de reconhecimento automático de fala (ASR), é possível automatizar a avaliação da fluência na fala, o que permite o seu uso em larga escala. Contudo, há casos em que o processo não consegue realizar uma classificação adequada da leitura e, nesses casos, é preciso o envolvimento humano. O objetivo desse trabalho é apresentar uma solução colaborativa para complementar a avaliação automática da fluência na fala em larga escala. Os resultados apresentados reforçam a viabilidade da solução.*

1. Introdução

A linguagem oral é determinante na vida do aluno. Toda a produção do conhecimento parte dessa linguagem e, por isso, a fluência na fala é um fator essencial para crianças em idade de alfabetização. Quando bem desenvolvida, esta permitirá que o aluno interprete o que ouve, tenha pensamento organizado e responda perguntas com lógica e clareza [Chaer and Guimarães 2012]. Assim, sua avaliação é fundamental e uma ferramenta que auxilie nesse processo possibilita acompanhar o desenvolvimento do aluno.

Uma ferramenta para a avaliação da fluência na fala buscará oferecer um conjunto de facilidades para agilizar a avaliação de determinados aspectos que podem ser automatizados, gerando um conjunto de informações úteis que poderá ser utilizado como apoio para a avaliação do aluno e da turma [Araújo et al. 2003]. Tal ferramenta tem como principal vantagem a otimização do processo, com rapidez na coleta dos dados, o uso de grandes amostras, menor custo de administração e taxas de retorno mais altas [Dixon 2001]. Assim, os avaliadores podem encontrar de maneira mais fácil e rápida os pontos a serem

corrigidos e o professor poderá voltar sua atenção a esses pontos, o que o ajudará a obter um resultado melhor e mais específico poupando tempo e esforço.

Considerando o volume de avaliações realizadas pelos órgãos responsáveis em âmbito nacional, se faz necessária a automatização do processo, uma vez que os métodos de avaliação comumente utilizados dentro de sala de aula são ineficazes quando aplicados em escalas maiores. Na avaliação da fluência na fala, a qual avalia também a capacidade de leitura da criança, a economia gerada com a automatização é significativa e, sem ela, o custo de contratação de avaliadores humanos pode inviabilizar o processo.

O presente trabalho apresenta uma abordagem que tem por objetivo apoiar a avaliação em larga escala da leitura infantil através de uma classificação automática de áudios de leituras e, nos casos em que a classificação automática não for possível, é aplicada uma abordagem colaborativa para chegada de consenso na avaliação. A abordagem foi criada a partir de entrevistas com avaliadores com prática em avaliação em larga escala e os resultados apresentados sinalizam a viabilidade da proposta. O trabalho está organizado da seguinte maneira: na Seção 2 é apresentada uma revisão bibliográfica sobre a avaliação da fluência na fala; na Seção 3 é apresentada a abordagem convencional da avaliação da leitura, na Seção 4 é detalhada a proposta deste trabalho. Em seguida, a Seção 5 apresenta os experimentos realizados para mostrar a viabilidade da proposta. Por fim, a Seção 6 apresenta as conclusões dos autores e trabalhos futuros.

2. Avaliação da Fluência na Fala

A partir da avaliação do desempenho dos alunos, educadores podem obter informações suficientes sobre turmas, escolas e até mesmo todo o país. Dessa maneira, é possível o planejamento de um modelo de ensino mais eficaz, melhorando os indicadores de interesse [Araújo et al. 2003]. Contudo, a avaliação da aprendizagem é uma questão complexa, com diferentes vertentes e posicionamentos. Embora a qualidade seja um conceito socialmente construído, é possível atingir consensualmente um patamar mínimo sobre o qual deve se assentar o processo educacional de qualquer cidadão [Gomes 2009]. A relação entre o custo anual por aluno e a qualidade da educação no ensino fundamental é estabelecida a partir de coeficientes de referência para distribuição dos recursos financeiros entre os diferentes níveis e modalidades de ensino [Gomes 2009], e a automatização de parte do processo de avaliação traz consigo a possibilidade de uma redução significativa de tempo e dos recursos necessários.

A automatização de avaliação da fluência da fala pressupõe um sistema que possa avaliar a leitura humana presente em um áudio gravado. Nesse cenário, o uso de técnicas de reconhecimento automático de fala (ASR) se faz necessário. Avanços significativos já foram alcançados no reconhecimento de fala para adultos e muitas vezes é assumido que esses avanços serão transferidos para o domínio com crianças. Essa transferência não é tão automática e, quando o sistema é treinado especificamente para a criança, tende a ter resultados significativamente melhores [Wöllmer et al. 2011]. Por esse motivo, é adequado que o sistema leve em consideração a faixa etária do falante [Hämäläinen et al. 2014], por conta das propriedades acústicas e linguísticas da fala infantil, que são diferentes para leitura e para a fala espontânea. Ainda, o efeito de mudanças advindas em crianças a partir dos 6 anos de idade em relação ao desenvolvimento e variabilidade das características acústicas de suas vozes pode afetar o sistema [Gerosa et al. 2009].

A definição de fluidez na leitura é relativamente estreita [Kuhn et al. 2010] e diretamente relacionada com os resultados obtidos na avaliação da leitura. Diversos autores utilizam como métrica a quantidade total de palavras lidas, o tempo gasto na leitura e a quantidade de acertos para classificar as crianças em grupos que expressam sua capacidade de leitura [Duchateau et al. 2007]. A partir dessas métricas, é possível encontrar anomalias, como hesitações, sussurros, alongamentos, entonação de perguntas, entre outras, o que permite avaliar de uma leitura como disfluente [Black et al. 2007]. As medidas de fluidez de um minuto podem ser utilizadas como classificadoras e são tecnicamente interessantes por demandar de pouco espaço de armazenamento dos áudios e pouco custo de transmissão dos arquivos, o que faz a métrica ser utilizada por vários autores [Valencia et al. 2010, Deeney 2010]. Para esses autores, esta métrica é útil para identificar estudantes em risco de dificuldade de leitura e definir intervenção do professor ou nova política para escola.

Contudo, sistemas para avaliação automática de fluência implementam modelos probabilísticos, o que são passíveis de gerar falso-positivos ou falso-negativos. Isto ocorrerá, principalmente, quando houver existência de interferências no áudio que dificultem a avaliação automática. Em avaliações em larga escala, por sua vez, é custoso controlar todas as variáveis que possam dificultar a correta captação do áudio, como uso padronizado de equipamentos de captação, interrupção do professor, sala sem isolamento acústico, etc. Para aumentar a confiabilidade de sistemas para avaliação, é necessário definir parâmetros que minimizem a existência de falso-positivos e falso-negativos. Diferente de [Black et al. 2007], onde os resultados falso-positivos contribuíram para que o resultado geral ficasse pior, ou de [Black et al. 2008], onde os casos de falso-positivos e falsos-negativos são detectados para que não sejam levados em conta nos resultados, no presente trabalho é identificado a região de incerteza do sistema de classificação para dividir as avaliações automática com alta probabilidade de certeza e aquelas que necessitam passar por uma avaliação manual.

3. Abordagem automática e colaborativa para avaliação em larga escala

O presente trabalho propõe uma ferramenta para apoio à avaliação da fluência na fala de crianças em idade de alfabetização. Para guiar o desenvolvimento da solução, foi realizada uma pesquisa com avaliadores com experiência em avaliação de fluência de crianças e que participaram de avaliações que utilizaram áudio gravado por professores. Foi distribuído um questionário a esses avaliadores, os quais 17 responderam. Entre as respostas, destacam-se as situações na leitura que não haviam sido previstas no protocolo de avaliação. Também foi frequente o relato de situações nas quais os próprios avaliadores não tinham certeza de como avaliar alguns áudios. Nesses casos, o próprio avaliador deveria decidir sobre o resultado da avaliação e submetê-lo para supervisores do processo avaliativo, os quais geralmente são poucos e não são tão céleres na resposta. Ainda, os colaboradores também puderam expressar o quanto facilitaria o trabalho caso o mesmo pudesse ser feito em ambiente remoto, visto que o trabalho de avaliação geralmente não é a principal profissão do avaliador. Por fim, 88.2% dos participantes responderam que, havendo a possibilidade de comunicação entre os colaboradores, isso poderia auxiliar o trabalho.

Assim, foi projetada uma solução para auxiliar o trabalho desses avaliadores. Embora a ferramenta não substitua o trabalho humano, através dela é possível diminuir a

carga de trabalho em avaliações em larga escala e fornecer suporte para que profissionais da área realizem a avaliação de maneira mais eficiente. A ferramenta possui três interfaces: a API para o envio de áudios, o *dashboard* para o gerenciamento dos dados obtidos e as ferramentas para análise a serem utilizadas pelos colaboradores. O funcionamento da ferramenta é dividido em duas etapas: um processamento automatizado dos áudios para classificação da leitura e uma análise manual colaborativa dos áudios que não possuem garantia de correteza pelo processo de classificação automática. O fluxo de trabalho da ferramenta (Figura 1) será descrito em detalhes.

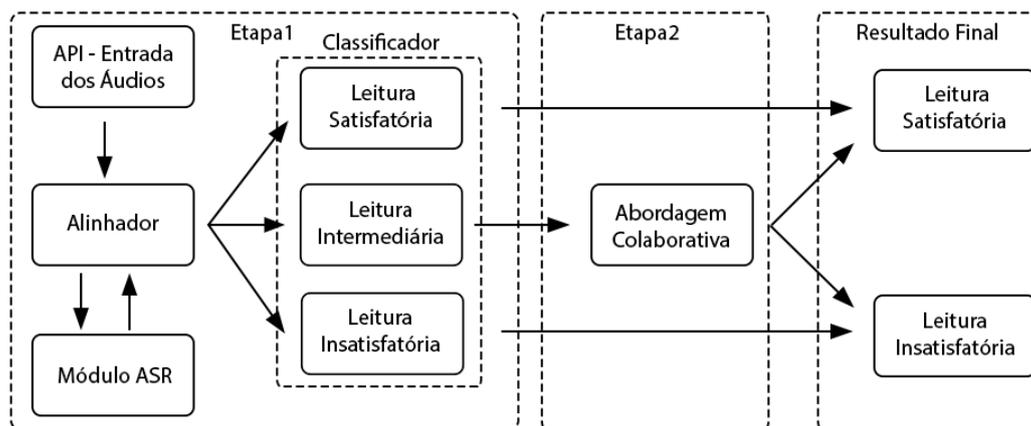


Figura 1. Fluxo do funcionamento da ferramenta

3.1. Etapa 1: Avaliação automática da leitura

O processamento automatizado é responsável pela classificação do sinal de áudio em categorias de qualidade segundo as leituras gravadas. Os áudios de leituras são transmitidos através de uma API REST para a ferramenta de avaliação. Para realizar a classificação automática desses áudios, foi utilizado (1) um texto de entrada, (2) um *dicionário léxico* reduzido com palavras presentes no texto e (3) um *modelo acústico*, responsável por transformar o sinal de áudio em fonemas.

O processo de classificação é realizado através da análise do áudio por um algoritmo de alinhamento temporal forçado, o qual é utilizado para determinar o tempo de ocorrência e a duração de cada fonema contido no áudio. Para isso, o alinhador utiliza o modelo acústico treinado com áudios em língua portuguesa e com características próximas aos áudios que serão avaliados. O dicionário léxico é utilizado para auxiliar na identificação da sequência de fonemas contidas no áudio para a formação de palavras. O alinhador forçado associa cada fonema encontrado com trechos do texto passado como referência. Na última etapa, o alinhador utiliza o texto de leitura como base e constrói um modelo de linguagem. Com o modelo de linguagem, o reconhecimento de palavras fica restrito às palavras que ocorrem no texto e, dessa maneira, força o alinhamento do áudio e do texto aumentando as chances de um reconhecimento mais preciso, ou seja, diminuindo a chance de que o reconhecimento seja prejudicado pela baixa qualidade do áudio.

Por meio do alinhador é possível obter a informação do momento em que os fonemas/palavras do texto base ocorreram nos áudios e qual foi sua duração. Com isso, é calculada a quantidade de palavras por intervalo de tempo que foram lidas

corretamente no áudio. Essa medida é utilizada para mensurar a velocidade com a qual o leitor consegue pronunciar corretamente as palavras do texto. Essa métrica demonstrou ser uma das mais eficientes para avaliação automática de fluência de leitores [Valencia et al. 2010, Deeney 2010]. A métrica aplicada no trabalho leva em consideração não só a capacidade da criança pronunciar corretamente as palavras, mas também a duração da leitura. Indivíduos de séries iniciais com dificuldades na leitura frequentemente têm a fala marcada pela ocorrência de muitas pausas. Assim, esta métrica é um bom classificador, pois um bom leitor tende a ter, além de uma pronúncia foneticamente mais correta, um ritmo de leitura mais adequado, o que é refletido positivamente nesta métrica.

A partir da avaliação automática do áudio, as leituras são classificadas como: uma leitura satisfatória, uma leitura insatisfatória ou uma leitura intermediária. Leituras intermediárias são aquelas que se encontram dentro de uma zona de incerteza do algoritmo, com maior probabilidade de presença de falso-positivos e falso-negativos. De acordo com experimentos realizados no laboratório, com uma base de áudio previamente classificada manualmente, foi determinado que uma taxa de 110 palavras/minuto é suficiente para garantir classificações de leituras satisfatórias com baixa incidência de falso-positivos. Da mesma forma, uma taxa de até de 30 palavras/minuto é suficiente para garantir a classificação de áudios como leituras insatisfatórias com baixa incidência de falso-negativos. Os áudios que obtiveram taxa entre 30 e 110 palavras/minuto são considerados intermediários e são enviados para a avaliação manual.

Um dos fatores que podem influenciar uma boa leitura ser classificada como intermediária é a presença de interferências no áudio. Interferências podem ocorrer por falta de equipamento adequado ou presença de ruídos durante a aplicação do teste (vozes ao fundo, sons de veículos, etc), o que dificulta o sistema em reconhecer corretamente a sequência de fonemas.

3.2. Etapa 2: Avaliação manual dos áudios

Todas as leituras que até esse estágio ainda não houverem sido classificadas automaticamente, é porque a ferramenta encontrou dificuldades motivadas pela baixa qualidade do áudio ou mesmo da baixa qualidade da leitura. Desse modo, é comum que os avaliadores também tenham dificuldades ao avaliarem parte dessas leituras. De acordo com a necessidade dos avaliadores entrevistados para essa pesquisa, a solução proposta contempla as seguintes funcionalidades principais:

1. **Abordagem colaborativa:** permitir a comunicação entre os avaliadores, de forma a incentivar a troca de conhecimento entre os avaliadores e que, nas situações de dúvidas, a avaliação da leitura não dependa apenas da opinião pessoal de um avaliador, mas de uma decisão da equipe, o que tende a ser uma decisão mais acertada.
2. **Marcação de trechos do áudio:** a ferramenta permite que se destaque no espectro do áudio regiões com características de interesse do avaliador, como dúvidas, erro de pronúncia, interferências, etc. Através das marcações manuais, essa informação é indexada e é possível adquirir, de maneira automática, informações variadas a respeito do áudio.

A avaliação é realizada com a apresentação do espectro do áudio a ser avaliado (Figura 2), permitindo que o avaliador realize marcações ao longo desse espectro. Cada

marcação é representada por um retângulo sobre a área do espectro que representa a região na qual o avaliador registra alguma dúvida ou consideração. Cada região marcada pode ser identificada por uma cor, a qual representa um tipo de marcação.



Figura 2. Ferramenta para avaliação de áudios

Para os áudios que possuem a presença de marcações de dúvidas, a avaliação ficará sujeita a uma decisão do grupo. Outros avaliadores, ao acessá-los, analisarão os trechos marcados como dúvidas e poderão interagir através de troca de mensagens para chegar em um consenso quanto a avaliação da leitura naquele áudio. Dessa maneira, os colaboradores podem melhorar o seu processo de avaliação e esse processo tende a gerar resultados mais confiáveis, uma vez que vêm da discussão de várias opiniões. Além disso, como a abordagem colaborativa ainda envolve a construção de conhecimento, à medida que os colaboradores interagem entre si, é esperado que os mesmos tenham menos dúvidas e gerem avaliações mais precisas [Ferreira and Fernandes 2017].

3.3. Geração de dados para tomada de decisões

A ferramenta apresenta relatórios e gráficos em um *dashboard* (Figura 3) para acompanhamento do processo de avaliação e para auxiliar os responsáveis pela avaliação na análise dos dados. Os gestores podem obter uma série de informações a respeito da avaliação, tais como quais projetos estão em funcionamento, seus prazos, quantos dos áudios ainda existem pra serem processados, quantos já o foram, quantos já foram avaliados por colaboradores, como estão os acessos ao sistema, entre outras. Essas informações também podem ser buscadas por data, região, tamanho do projeto, situação do projeto, assim como exibiadas de forma individual pra cada projeto cadastrado na ferramenta.

Para auxiliar no processo de análise, os resultados das avaliações podem ser dispostos em um mapa das regiões que participaram da avaliação. Os dados são exibidos de forma agregada por região do país, estado ou cidade e é possível identificar locais com índices insuficientes de avaliação ou locais com grande presença de áudios com ruídos, o que indica uma dificuldade na aplicação dos testes naqueles locais, seja por falta de equipamento ou treinamento inadequado dos aplicadores.

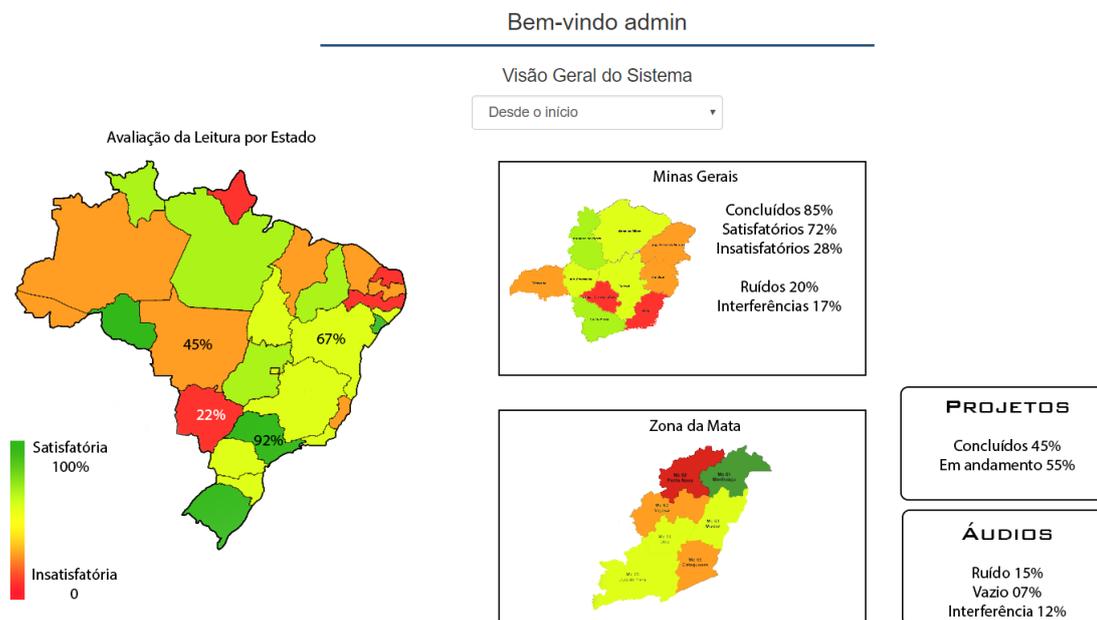


Figura 3. Recorte de uma das telas do *dashboard*

4. Experimentos

Foi realizado um experimento piloto para avaliar o comportamento da solução em um cenário real, embora numa escala reduzida. O experimento utilizou-se de uma base real de áudios de leituras de crianças em fase de alfabetização. A base contém 471 áudios com 1 minuto de duração e foram captados por professores de escolas públicas através de telefones celulares. A base engloba a leitura de 9 textos distintos comumente utilizados para avaliação de leitura nas escolas. A cada criança foi solicitada a leitura de três desses textos.

A base foi submetida à ferramenta, a qual classificou automaticamente 201 áudios, sendo 141 como leituras satisfatórias e 60 como leituras insatisfatórias. O montante corresponde a aproximadamente 43% do volume da base. Os demais, 270 áudios, foram classificados como leituras intermediárias e submetidos à avaliação manual.

A avaliação manual foi realizada por dois grupos distintos de avaliadores. Os avaliadores do primeiro grupo utilizaram a ferramenta para avaliação dos áudios (grupo 1), enquanto o segundo grupo realizou a avaliação dos mesmos áudios de forma convencional (grupo 2). Os grupos não se conheciam, não tiveram qualquer tipo de contato e realizaram a tarefa em momentos distintos. A todos os colaboradores, foi realizado um treinamento sobre o protocolo de avaliação com os critérios utilizados para classificação dos áudios, demonstradas algumas das situações comumente encontradas durante a avaliação e a forma que deveriam proceder perante a elas. Cada grupo realizou as avaliações durante o período de uma hora.

Durante a avaliação, cada palavra lida de maneira errada ao longo da leitura deveria ser contabilizada como erro pelo avaliador e, em caso de dúvida, em alguma situação não prevista no protocolo ou que o avaliador não se sinta confortável em decidir pelo erro

ou acerto, este deveria especificar o trecho de dúvida. Os participantes do grupo 1 deveriam marcar o trecho de áudio diretamente na ferramenta e submeter a dúvida para os outros avaliadores de forma anônima. Os participantes do grupo 2 deveriam especificar o intervalo do áudio que gerou dúvida em uma planilha eletrônica. No grupo 1, a dúvida seria exposta aos demais colaboradores para que ajudassem na tomada de decisão; no grupo 2, porém, a dúvida é registrada para que seja verificado posteriormente por um supervisor mas a o avaliador deve decidir sozinho quanto ao resultado da avaliação.

Como resultado, durante o período do experimento foram avaliados 112 áudios pelo grupo 2 e 133 áudios pelo grupo 1. O tempo médio gasto para avaliação de cada áudio foi de 3:48 minutos no grupo 2 e 3:21 min no grupo 1.

Dos 133 áudios avaliados pelo grupo 1, 42 áudios tiveram trechos que geraram algum tipo de dúvida ao avaliador. Dessas, 37 dúvidas foram respondidas pelos demais avaliadores, com discussões que, no total, geraram 146 respostas. Cada participante tinha acesso ao trecho marcado como dúvida por algum avaliador, poderia ouvi-lo e, sentindo-se à vontade, responder com a sua interpretação daquele trecho, justificando com base no protocolo. Com o fim do experimento, 40 áudios com dúvidas terminaram o experimento classificados pelos avaliadores, onde 27 foram avaliados como leituras satisfatórias e 13 como leituras insatisfatórias. Ainda, apenas 5 dúvidas não foram respondidas por conta do término da sessão.

Em comparação com a avaliação sem a ferramenta, houveram 12 áudios avaliados em ambos os grupos cujo o resultado da avaliação foi diferente e foram marcados com dúvidas ao serem avaliados pelo grupo 1. Como a decisão sobre o resultado foi discutida entre os colaboradores do grupo 1, nessa abordagem o resultado tende a ser mais confiável. Percebe-se, também, que a média de dúvidas por avaliador é maior quando utilizada a ferramenta, o que mostra que quando a abordagem é colaborativa, há uma tendência maior de expor as dúvidas e, dessa maneira, uma chance menor de ocorrer decisões precipitadas.

Ao final do experimento, os participantes dos dois grupos preencheram um questionário com perguntas sobre o trabalho realizado. Os participantes do grupo 2 foram questionados sobre critério que utilizaram para dar um parecer final sobre a leitura nos casos em que haviam dúvidas. Em geral, estes responderam que desconsideraram o trecho do áudio com a dúvida. Com isso, no final são contabilizadas menos palavras do texto, o que pode interferir diretamente na classificação da leitura. Por fim, 62.5% dos participantes desse grupo responderam que, podendo ser auxiliado por outros avaliadores nesses casos, é possível que suas decisões fossem diferentes. Os participantes do grupo 1, por sua vez, responderam se a troca de informações com outros participantes fez com que avaliassem a qualidade da leitura de maneira diferente da que faria caso tivesse que tomar a decisão sozinho. Foram registradas 87.5% das respostas afirmativas a essa pergunta. Isso mostra que a ferramenta pode ter auxiliado na construção do conhecimento dos avaliadores e, principalmente, em garantir uma avaliação mais correta desses áudios avaliados.

5. Conclusões e Trabalhos Futuros

Neste trabalho foi apresentado o problema de avaliação em larga escala da leitura infantil e como a avaliação automática da fluência pode gerar resultados equivocados. Para dimi-

nuir os erros de classificação, o sistema de avaliação automática foi treinado para classificar a qualidade da leitura em três classes: leituras satisfatórias, leituras insatisfatórias e leituras intermediárias. A classificação nas duas primeiras classes possuem limiares definidos de forma a diminuir a presença de falso-positivos e falso-negativos. As leituras intermediárias estão numa região que não garante-se uma alta precisão da classificação automática e, por isso, são enviadas para a avaliação manual. Para apoiar a avaliação manual, foi desenvolvido um sistema que permite a colaboração entre avaliadores nos casos de dúvidas no processo avaliativo.

Os experimentos realizados com áudios gravados por professores de escolas públicas mostraram que a abordagem é capaz de reduzir o trabalho manual, tendo, na base utilizada nos experimentos, classificado aproximadamente 43% dos áudios como leituras satisfatórias ou insatisfatórias. Caso essa proporção se confirme em uma avaliação em escala nacional, tem-se a redução de mais de 40% do custo com contratação de avaliadores. Após o processo de avaliação manual, a maioria dos avaliadores considerou que a troca de mensagens com outros avaliadores gerou uma reavaliação por parte do avaliador.

Alguns desafios, contudo, ainda necessitam ser estudados para melhoria desse processo. A maioria dos áudios avaliados foram classificados como leituras intermediárias por conter ruídos que dificultaram a classificação automática. Métodos para identificação de situações anômalas no áudio, como áudio muito baixo ou áudio cortado, devem ser explorados. Ainda, técnicas para identificação de presença de múltiplos falantes podem ser utilizadas para diferenciar vozes de fundo ou a fala do professor, caso ele venha a interromper o aluno durante a gravação. Por fim, ao aplicar o sistema para outro conjunto de áudios, é possível que a proporção de áudios classificados automaticamente não corresponda à proporção encontrada neste experimento, sendo necessárias novas avaliações para validar a proposta antes do seu uso efetivo.

6. Agradecimentos

O autor agradece ao CAEd/UFJF pelo material fornecido para a pesquisa; e também agradece à CAPES pela concessão de bolsa de mestrado, que possibilitou a pesquisa.

Referências

- Araújo, A. M. T., de Menezes, C. S., and Cury, D. (2003). Apoio automatizado à avaliação da aprendizagem utilizando mapas conceituais. In *Simpósio Brasileiro de Informática na Educação*, volume 1, pages 287–296.
- Black, M., Tepperman, J., Kazemzadeh, A., Lee, S., and Narayanan, S. S. (2008). Pronunciation verification of english letter-sounds in preliterate children. In *Ninth Annual Conference of the International Speech Communication Association*.
- Black, M., Tepperman, J., Lee, S., Price, P., and Narayanan, S. S. (2007). Automatic detection and classification of disfluent reading miscues in young children's speech for the purpose of assessment. In *Eighth Annual Conference of the International Speech Communication Association*.
- Chaer, M. R. and Guimarães, E. d. G. A. (2012). A importância da oralidade: educação infantil e séries iniciais do ensino fundamental. Disponível em: <http://pergaminho.unipam.edu.br/documents/43440/43870/a-importancia.pdf>. Acesso em: 04 abril 2018.

- Deeney, T. A. (2010). One-minute fluency measures: Mixed messages in assessment and instruction. *The Reading Teacher*, 63(6):440–450.
- Dixon, J. (2001). Evaluation tools for flexible delivery (workshop version). *Melbourne: TAFE frontiers*.
- Duchateau, J., Cleuren, L., Ghesquière, P., et al. (2007). Automatic assessment of children's reading level. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 1210–1213.
- Ferreira, T. and Fernandes, M. (2017). Detecção de traços de personalidade em textos para apoiar a formação de grupos para colaboração. In *Simpósio Brasileiro de Informática na Educação*, volume 28, page 1627.
- Gerosa, M., Giuliani, D., Narayanan, S., and Potamianos, A. (2009). A review of asr technologies for children's speech. In *Proceedings of the 2nd Workshop on Child, Computer and Interaction*, page 7. ACM.
- Gomes, A. V. A. (2009). Custo aluno qualidade. *Brasília: Câmara dos Deputados*.
- Hämäläinen, A., Meinedo, H., Tjalve, M., Pellegrini, T., Trancoso, I., and Dias, M. S. (2014). Improving speech recognition through automatic selection of age group-specific acoustic models. In *International Conference on Computational Processing of the Portuguese Language*, pages 12–23. Springer.
- Kuhn, M. R., Schwanenflugel, P. J., and Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45(2):230–251.
- Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., and Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, 45(3):270–291.
- Wöllmer, M., Schuller, B., Batliner, A., Steidl, S., and Seppi, D. (2011). Tandem decoding of children's speech for keyword detection in a child-robot interaction scenario. *ACM Transactions on Speech and Language Processing (TSLP)*, 7(4):12.