

Desenvolvimento de Aplicativo de Aquisição e Processamento de Voz

Development of a mobile Application to Acquisition and Voice Processing

Autor

Pedro Henrique Garcia de Souza
pedroengarcia@gmail.com
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil.

Orientador

Prof. Dr. Paulo Rogério Scalassara
prscalassara@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil.

RESUMO

O foco deste trabalho é desenvolver um aplicativo móvel que opere no sistema operacional Android. Este aplicativo deve coletar, registrar e apresentar dados de um sinal vocal, utilizando somente os recursos de um aparelho celular. O projeto foi realizado em 4 etapas: 1) Escolha do sistema operacional e da interface de desenvolvimento, 2) Desenvolvimento do aplicativo, 3) Aquisição e processamento dos dados, 4) Diagnóstico e apresentação dos dados.

PALAVRAS-CHAVE: Aplicativo Android. Sinal de voz. Detecção do pitch.

ABSTRACT

The focus of this work is to develop a mobile application that runs on Android operating system. This application should collect, register and present the data of a vocal signal, using only the resources of a cellular device. The project will be carried out in 4 steps: 1) Choice of operating system and development interface, 2) Application development, 3) Data acquisition and processing, 4) Diagnosis and data presentation.

KEYWORDS: Android application. Voice signal. Pitch detection.

Recebido: 30 ago. 2018.

Aprovado: 18 set. 2018.

Direito autoral:

Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



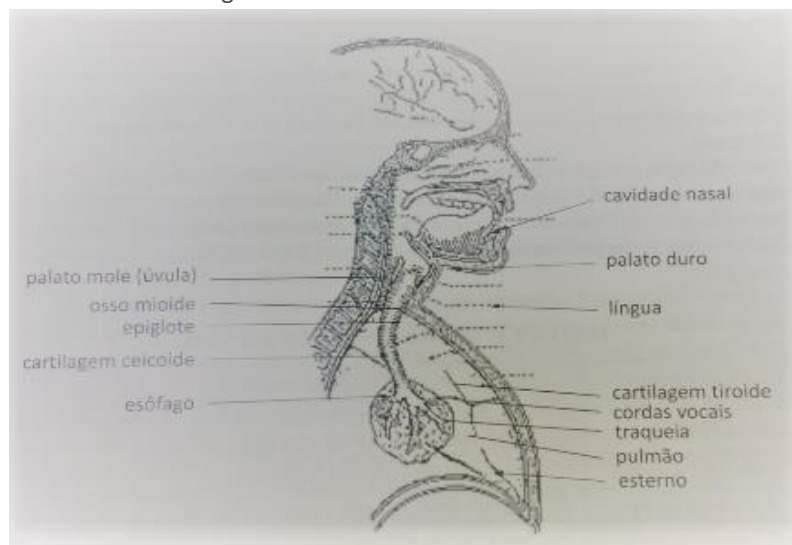
INTRODUÇÃO

A fala representa um dos meios de comunicação mais importantes utilizados pelo homem para expressar seus sentimentos e ideias. Os sons de fala podem ser definidos fisicamente como ondas de pressão, resultantes de vibrações longitudinais das moléculas do ar. O notável é que essas vibrações são provocadas por órgãos do corpo humano destinados às funções vitais de comer, beber e respirar (ALCAIM; ABRAHAM; OLIVEIRA, 2011). No ambiente profissional, a utilização da voz permite a transmissão de confiança, liderança, credibilidade e assertividade (IRIYA, 2014). Os sintomas mais comuns apresentados por profissionais que usam a voz como instrumento de trabalho são: rouquidão, tosse, dor de garganta/queimação, fadiga vocal, garganta seca, perda de voz e variação (ZAMBON, 2012).

Para um diagnóstico preciso de distúrbio vocal, um exame clínico não é eficiente, pois avalia apenas uma situação ideal de uso vocal. É sempre necessário e recomendável realizar uma avaliação do comportamento vocal em uma grande amostra de voz, durante um período prolongado de tempo. O médico otorrinolaringologista e/ou o fonoaudiólogo precisam monitorar a utilização da voz do paciente em longo prazo, a fim de quantificar o tempo, os ciclos, a dosagem e a energia da vibração das pregas vocais e, assim, mensurar a distância percorrida por elas durante um dia de trabalho (HILLMAN et al., 2006). Portanto, o monitoramento do comportamento vocal do profissional é fundamental para realizar diagnósticos mais precisos e, conseqüentemente oferecer ao paciente um tratamento personalizado e efetivo (MORETI et al. (2013).

Nesse sentido, o trabalho originou-se devido a necessidade de um equipamento móvel e de baixo custo, capaz de coletar dados em diferentes períodos do dia, utilizando uma forma não evasiva. Tendo em vista que as informações são provenientes de uma onda de pressão acústica, vindas do aparelho respiratório e mastigatório, apresentado na Figura 1, a utilização de um aparelho celular classifica-se como uma opção viável para esse estudo.

Figura 1 – Mecanismo vocal humano.



Fonte: Fundamento de Processamento de sinais de voz e imagem (ALCAIM; ABRAHAM; OLIVEIRA, 2011).

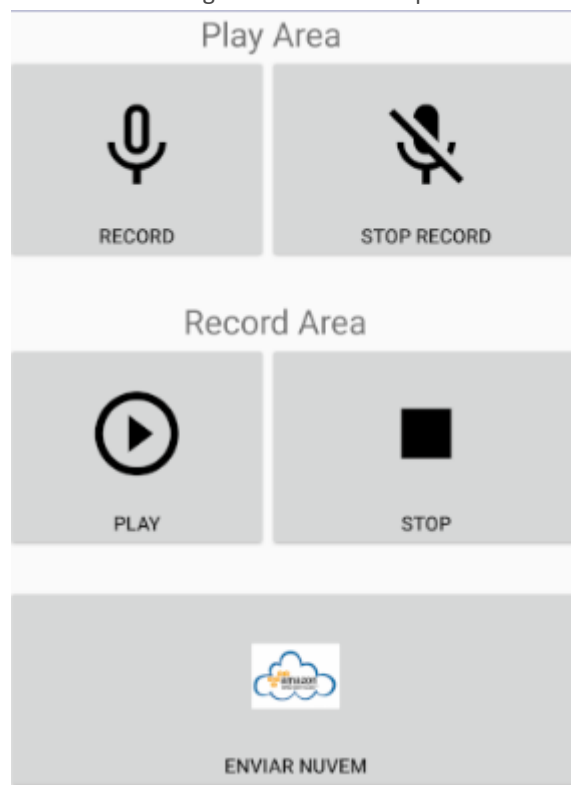
MÉTODOS

O projeto foi realizado em 4 etapas: 1) Escolha do sistema operacional, da interface de desenvolvimento e do armazenamento em nuvem 2) Desenvolvimento do aplicativo 3) Aquisição e processamento dos dados e 4) Diagnóstico e apresentação dos dados.

Na primeira etapa, a escolha do sistema operacional Android levou em conta as vantagens da taxa de licença de publicação de aplicativos, a linguagem de programação utilizada e por possuir uma interface de programação intuitiva, a IDE Android Studio. A escolha da nuvem Amazon S3 se fez pela flexibilidade para controlar quem pode acessar os dados e por disponibilizar um espaço de 5GB no primeiro ano gratuitamente.

O aplicativo desenvolvido consiste em um gravador de voz que possui 5 funções interagindo em tempo real com o usuário, essas funções são: 1) *Record*, onde a gravação do sinal de áudio inicia-se, 2) *Stop Record*, que interrompe o processo de gravação e salva o arquivo de áudio no formato .mp3 no cartão de memória micro SD, 16GB, classe 10, 3) *Play*, onde o arquivo salvo é reproduzido para o usuário, 4) *Stop*, interrompe e pausa a reprodução e 5) *Enviar nuvem*, que quando conectado a internet envia esse arquivo para nuvem. A interface do aplicativo é apresentada na Figura 2.

Figura 2 – Interface Aplicativo.

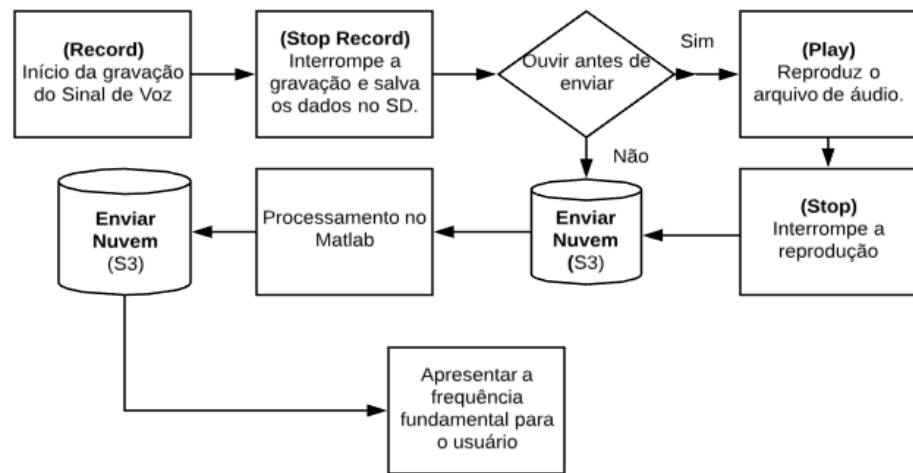


Fonte: Autoria própria.

A terceira etapa do projeto depende diretamente dos resultados obtidos na etapa anterior. O sinal é coletado por um período de tempo e, em seguida, ele é enviado para a nuvem Amazon S3. Uma vez na plataforma online, os dados são coletados pelo software Matlab, onde o processamento é realizado. Nessa etapa,

a frequência fundamental é detectada utilizando a toolbox *Audio System*. A Figura 3 apresenta um esquemático de todo o processo.

Figura 3 – Esquemático do processo.

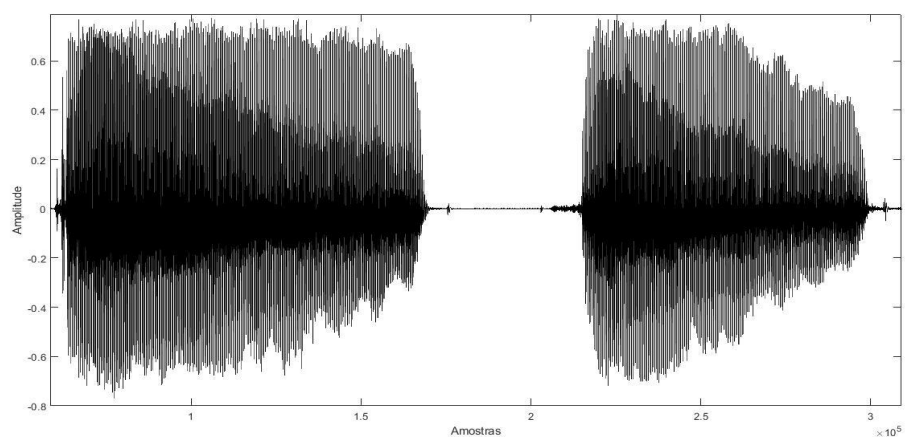


Fonte: Autoria própria.

RESULTADOS E DISCUSSÕES

Na primeira etapa, foram realizados testes a onde foi solicitado que se pronunciasse a vogal /a/ sustentada intercalando momentos de silêncio por um breve período de tempo, conforme sinal ilustrado na Figura 4. Esse sinal foi convertido de .mp3 para .wav, utilizando uma taxa de amostragem de 44.1Khz.

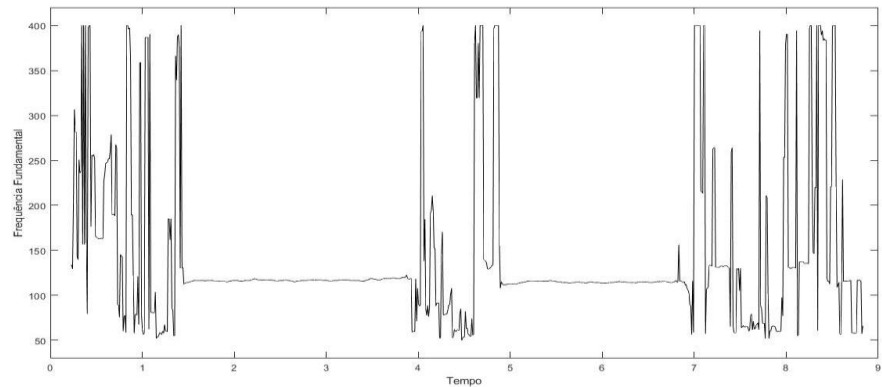
Figura 4 – Teste emitindo a vogal /a/ sustentada.



Fonte: Autoria própria.

Feito o processamento do sinal de áudio no Matlab e foi retornado como resposta um vetor da estimativa da frequência fundamental (pitch) ao longo do tempo apresentado na Figura 5. A valor médio desse vetor foi 116.0526Hz, valor esse que se enquadra dentro da média típica para homens. O algoritmo utilizado foi de cálculo pronto (GONZALEZ; SIRA; BROOKES, 2011).

Figura 4 – Teste emitindo a vogal /a/ sustentada.



Fonte: Autoria própria.

O projeto se encontra ainda na última etapa, ainda não foi possível implementar o retorno do processamento para o aplicativo, sendo assim todos os resultados se encontram na nuvem.

CONSIDERAÇÕES FINAIS

Tendo em vista a necessidade de monitoramento das características vocais ao longo de períodos do dia, surgiu o estudo para o desenvolvimento do aplicativo. Os resultados do processamento têm o potencial para direcionar novas pesquisas de extração de parâmetros em cima do valor de Pitch em cada instante de tempo. O projeto encontra-se na etapa final de apresentação do processamento pela interface no usuário no celular. Espera-se que o aplicativo final possua uma interface simples e útil para a detecção de parâmetros da voz.

AGRADECIMENTOS

O presente trabalho foi realizado com o apoio da Universidade Tecnológica Federal do Paraná - UTFPR - Brasil.

REFERÊNCIAS

- ALCAIM, Abraham; OLIVEIRA, Carlos Alexandre dos Santos (Ed.). **Fundamentos do processamento de sinais de voz e imagem**. Rio de Janeiro: Interciência, 2011. 312 p.
- IRIYA, R. **Análise de sinais de voz para reconhecimento de emoções**. Dissertação (Mestrado em Ciência). Escola Politécnica da Universidade de São Paulo, 2014. 116 p.
- ZAMBON, F. **Epidemiology of Voice Disorders in Teachers and Nonteachers in Brazil: Prevalence and Adverse Effects**. *Journal of Voice*, Vol. 26, No. 5. 2012.

HILLMAN R; HEATON J; MASAKI A., ZEITELS S; CHEYNE H. (2006). **Ambulatory Monitoring of Disordered Voices**. Annals Otology, Rhinology, Laryngology. 115:795-801.

MORETI F; RAIZE T; SAUDA C; LOURENÇO L; OLIVEIRA G; BEHLAU M. (2013). **Grau de quantidade de fala e intensidade vocal de teleoperadores em ambiente laboral e extralaboral**. Revista da Sociedade Brasileira de Fonoaudiologia. 385-390.

GONZALEZ; SIRA; BROOKES, M. **A Pitch Estimation Filter robust to high levels of noise (PEFAC)**. 19th European Signal Processing Conference. Barcelona, 2011, pp. 451-455.