

# Protótipo para a análise da Frequência Fundamental da Voz com módulo implementado em dispositivo reconfigurável do tipo FPGA

Ana Maria Jerônimo Soares<sup>1</sup>, Rodrigo Leone Alves<sup>2</sup>

<sup>1</sup> Universidade Potiguar, <sup>2</sup> Instituto Federal de Educação Ciência e Tecnologia do Rio Grande do Norte  
anamaria.soa@gmail.com, rodrigo.leone@ifrn.edu.br

**Resumo:** Pessoas com deficiência auditiva apresentam alterações vocais. Destaca-se que um dos parâmetros mais afetados e que necessita estudo e correção é a Frequência Fundamental da Voz (F0). Com os avanços observados em dispositivos programáveis do tipo FPGA (*Field Programmable Gate Array*), tanto na redução dos custos quanto no aumento da capacidade, evidenciou-se que estes constituem uma excelente alternativa no Processamento Digital de Sinais vocais. A presente pesquisa contempla a implementação de um sistema de análise vocal baseado na computação reconfigurável. Os resultados apontam para o potencial dessa proposta, que deixa como contributo um protótipo capaz de extrair a F0 de sinais de voz capturados e apresentar um retorno para o usuário, logo poderá auxiliar na atuação fonoaudiológica.

**Palavras-chave:** *Computação reconfigurável, Frequência Fundamental da Voz, FPGA.*

## 1. Introdução

Para que a criança aprenda a falar, isto é, desenvolva a produção oral, é preciso que possua a habilidade de percepção auditiva, de modo que possa ouvir os sinais sonoros do ambiente e assim reproduzi-los [1]. Assim, a integridade do sistema auditivo é fator fundamental para o desenvolvimento da fala e da linguagem [2], e conseqüentemente, do processo comunicativo, necessário para a interação do ser humano com o ambiente que o rodeia.

É importante que a sociedade tenha a compreensão de que nem toda pessoa surda é igualmente muda [3], ou seja, tal indivíduo apresenta todas as condições físicas e fisiológicas para falar, porém não desenvolve a oralidade espontaneamente devido ao fato de não escutar os sons do ambiente, e assim reproduzi-los [4].

A pessoa com deficiência auditiva possui como aspecto marcante a sua fala, tendo em vista que a voz apresenta quedas bruscas de altura ou, ao contrário, é monótona e muito aguda. As alterações nos aspectos vocais da pessoa com deficiência auditiva estão associadas ao grau ou tipo de perda auditiva e idade em que esta ocorreu [4].

As alterações citadas são verificadas principalmente nos valores da Frequência Fundamental da Voz (F0) [1], ressonância, ritmo, intensidade, inteligibilidade, qualidade vocal, e velocidade [4]. Ademais, dentre as características apresentadas, a Frequência Fundamental merece atenção [5].

A Frequência Fundamental da Voz, ou F0, trata-se do número de ciclos vibratórios das cordas vocais por segundo, tendo como equivalente perceptual o *pitch* [6]. As pessoas surdas apresentam valores de F0 significativamente alterados [1]. Destaca-se que algumas técnicas e intervenções clínicas, com o suporte de tecnologias, possibilitam a correção do referido parâmetro [5], e conseqüentemente, promove melhorias para a produção oral das pessoas com deficiência auditiva [1].

No que tange o surgimento de novos artefatos tecnológicos com potencial para a implementação de sistemas de Processamento Digital de Sinais de voz, destaca-se os dispositivos eletrônicos reconfiguráveis do tipo FPGA. *Field Programmable Gate Array* - FPGA consiste em um conjunto de circuitos integrados conectados por uma matriz de células lógicas

programáveis [7], e geralmente vem acoplado em kits de desenvolvimento, podendo implementar sistemas complexos em um único componente, dispensando a utilização de um grande número de Circuitos Integrados e tendo como vantagem a reconfigurabilidade, ou seja, qualquer mudança no sistema pode ser obtida alterando-se apenas a configuração/programação do *software*, sem a necessidade de mudar o *hardware*.

Com base nas abordagens analisadas na literatura, que demonstram a relevância do uso de sistemas embarcados e reconfiguráveis para o processamento digital de sinais [1][8], surge a presente proposta, que tem como objetivo prioritário a criação de um protótipo, utilizando FPGA, para analisar as medidas de F0 e demonstrar as alterações nos valores, para posteriormente prover um mecanismo de correção através da substituição sensorial, explorando-se portanto, o sentido visual, de modo que o indivíduo com perdas auditivas possa ter um recurso a mais para análise, monitoramento e controle da sua F0, por meio de uma aplicação alternativa, que utiliza tecnologias com significativo potencial de processamento de sinais, sendo de baixo custo.

Destaca-se que a análise da F0 de surdos em trabalhos acadêmicos utilizando dispositivos lógicos programáveis é um campo amplo, porém ainda pouco explorado [8]. Tal abordagem deixa como contribuição um estudo que apresenta informações relevantes para a comunidade acadêmica e interessados na área.

Em síntese, pode-se destacar que o impacto de se trabalhar a análise de parâmetros vocais em tempo real em dispositivo do tipo FPGA reside no fato de possibilitar a implementação de técnicas de processamento digital de sinais de voz em um componente eletrônico com potencial para melhorar o desempenho do processo.

## 2. Metodologia

O projeto foi configurado no kit de desenvolvimento Altera DE2. O principal *software* EDA utilizado foi o *Quartus II Web Edition* versão 7.2. Ademais, para a codificação do sistema, utilizou-se a linguagem *Very High Speed Integrated Circuit Hardware Description Language* – VHDL, que consiste em uma linguagem de alto nível utilizada para fazer o modelamento estrutural

e comportamental de um sistema a ser implementado em dispositivo reconfigurável [9].

O algoritmo utilizado para a extração da F0 (*Pitch*) foi baseado em outros estudos presentes na literatura [8] [12] e que tiveram resultados satisfatórios, embora não implementados em FPGA. Seu fluxo está esquematizado na Figura 1.

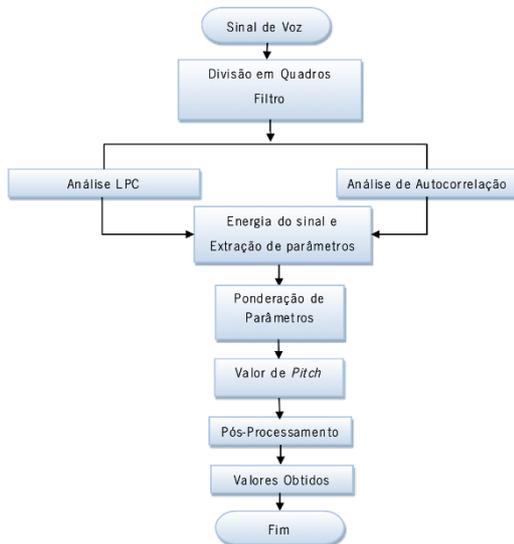


Figura 1 Fluxograma para extração da F0. FONTE: os autores.

Assim, implementou-se as técnicas de pré-ênfase e segmentação do sinal utilizando-se o janelamento *Hamming* com segmentos de 20 ms. Utilizou-se um filtro passa-baixa, de quarta ordem e com frequência de corte de 2000 Hz.

Para a estimação do *Pitch* considerou-se a posição do pico do sinal, isto é, em função da concentração de energia, sendo fundamentada na função de autocorrelação dos resíduos da análise LPC (*Linear Predictive Coding*). Posteriormente, os valores foram classificados e comparados com padrões de referência. Por conseguinte, realizou-se o cálculo da mediana dos resultados dos segmentos vozeados para evitar erros que comprometam os valores finais [8] [12].

Tendo em vista que as extensões das frequências de sons audíveis estão entre 20 Hz e 20.000 Hz, uma frequência de amostragem de 44.100 Hz é capaz de evitar saturação, estando dentro dos padrões de *Nyquist* que determina que a taxa mínima de amostragem deve ser igual ao dobro da frequência máxima contida no sinal [8].

No que tange a conversão do sinal, o conversor A/D está presente no kit didático. Assim, implementou-se um bloco para controlar o fluxo de dados recebidos e enviá-los para o bloco de controle de memória SDRAM, e um bloco para enviar os dados armazenados na memória para serem apresentados no Display de LCD.

Em síntese, o sistema proposto foi projetado de forma que seus parâmetros de entrada foram obtidos a partir da captura de um sinal de voz obtido por meio de um microfone acoplado ao dispositivo eletrônico. O sinal foi posteriormente digitalizado e processado no kit de desenvolvimento, onde extraiu-se a F0 e retornou-se

através de um display de LCD o valor mensurado. Ademais, os usuários poderiam ter além do valor mostrado no display, uma outra resposta visual por meio da indicação de acerto, aproximação ou distanciamento da frequência extraída em relação ao valor esperado, através de variações das cores em um LED RGB.

A Figura 2 apresenta o kit didático com o sistema implementado, mostrando a extração de uma F0 para uma pessoa do sexo feminino.

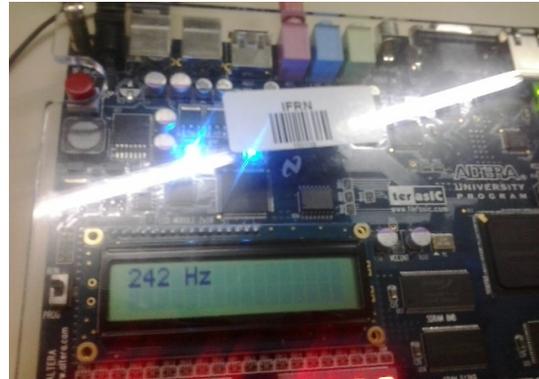


Figura 2. Extração da F0. FONTE: os autores.

Este estudo contou com a participação de 34 voluntários. Ademais, a amostra foi composta por um total de 20 adultos ouvintes, 8 crianças com audição normal, 3 crianças com deficiência auditiva e 3 adultos com deficiência auditiva. A coleta de sinais para o teste do sistema consistiu na captura da emissão da vogal /a/ e de uma sentença/frase.

### 3. Resultados e Discussões

Baseando-se nas observações de estudiosos que analisaram o comportamento da F0 destaca-se que em sujeitos do sexo masculino sem perdas auditivas e fala normal, a frequência vai de 80 Hz até em torno de 114 Hz e de 150 a 250 Hz no sexo feminino [10]. Em crianças os valores da frequência que indicam normalidade estão acima de 250 Hz, embora as formulações em outro estudo tenham demonstrado uma média de 236 Hz para a F0 das crianças [11].

Já outro estudo identificou-se que os homens com audição normal apresentam frequência fundamental entre 100 e 150 Hz, as mulheres possuem os valores do referido parâmetro entre 200 e 250 Hz e as crianças acima de 300 Hz [12]. Por conseguinte, observa-se que nas pessoas com surdez os valores da frequência fundamental são bastantes elevados em comparação com os indivíduos ouvintes.

A Tabela 1 apresenta os valores obtidos com o sistema implementado para os indivíduos adultos com audição normal.

Cabe destacar que buscou-se envolver indivíduos de variadas faixas etárias, todavia, mediante à indisponibilidade de alguns sujeitos convidados em participar do estudo, a amostra apresentou-se um pouco desproporcional, sobretudo no grupo dos adultos em que apenas um sujeito apresenta 51 anos, estando os demais abaixo dos 27 anos, porém tal fato não afeta as análises.

Tabela 1. F0 extraída nos ouvintes adultos. FONTE: os autores.

Nº	Sexo	Idade	F0 /a/	F0 frase
1	Feminino	19 Anos	234 Hz	240 Hz
2	Feminino	18 Anos	229 Hz	235 Hz
3	Feminino	22 Anos	239 Hz	243 Hz
4	Masculino	21 Anos	105 Hz	107 Hz
5	Feminino	21 Anos	237 Hz	239 Hz
6	Feminino	21 Anos	209 Hz	226 Hz
7	Feminino	21 Anos	224 Hz	228 Hz
8	Feminino	24 Anos	239 Hz	241 Hz
9	Masculino	19 Anos	110 Hz	114 Hz
10	Masculino	19 Anos	109 Hz	111 Hz
11	Feminino	22 anos	242 Hz	248 Hz
12	Feminino	25 anos	233 Hz	239 Hz
13	Masculino	17 anos	155 Hz	157 Hz
14	Masculino	51 anos	105 Hz	112 Hz
15	Masculino	23 Anos	98 Hz	102 Hz
16	Feminino	17 Anos	232 Hz	236 Hz
17	Feminino	17 Anos	105 Hz	107 Hz
18	Feminino	27 Anos	235 Hz	238 Hz
19	Masculino	23 Anos	100 Hz	105 Hz
20	Masculino	21 Anos	159 Hz	161 Hz

Para a vogal /a/, os achados vão desde 98 Hz a 159 Hz nos indivíduos do sexo masculino e de 209 Hz a 242 Hz nas mulheres. Com relação a sentença adotada os valores tiveram um pequeno aumento, de 102 Hz a 161 Hz nos homens e nas mulheres de 226 Hz a 248 Hz. Esses corroboram com a literatura [11] que demonstra valores em patamares aproximados.

Os resultados observados nos participantes de números 13 e 20 são considerados elevados segundo correlatos teóricos [10] [11], pois os mesmos são do sexo masculino e ouvintes e mesmo assim a F0 ultrapassa a faixa de 150 Hz, entretanto esse aumento é justificado pelo fato desses indivíduos possuírem voz excessivamente aguda. Os valores extraídos nas crianças com perda auditiva são apresentados na Tabela 2.

Tabela 2. F0 de crianças com perdas auditivas. FONTE: os autores.

Nº	Sexo	Idade	F0 /a/	F0 frase
1	Masculino	10 Anos	398 Hz	409 Hz
2	Masculino	11 Anos	435 Hz	438 Hz
3	Feminino	7 Anos	293 Hz	298 Hz

Nas crianças de número 1 e 2, a elevada F0, deve-se a perda auditiva severa, por este motivo a fala dos mesmos torna-se lenta e pouco inteligível. Já na criança nº 3 o valor está mais aproximado da normalidade, pois esta criança tem o auxílio do Implante Coclear, tendo um ganho na audição [1]. Os valores extraídos nos adultos com perda auditiva são apresentados na Tabela 3.

Tabela 3. F0 de adultos com perdas auditivas. FONTE: os autores.

Nº	Sexo	Idade	F0 /a/	F0 frase
1	Masculino	17 Anos	231 Hz	235 Hz
2	Feminino	25 Anos	305 Hz	307 Hz
3	Feminino	15 Anos	288 Hz	291 Hz

Em ambos os sexos os valores extraídos foram elevados se comparados ao padrão esperado, tendo em vista que em surdos a frequência fundamental da voz é elevada, tanto no sexo masculino quanto no feminino [12], pois eles não possuem o *feedback* auditivo para monitorar a produção vocal que acaba sendo realizada com muito esforço [13]. Por conseguinte, ao analisar-se a aplicabilidade do sistema para a extração da F0 em crianças sem perdas auditivas, evidenciou-se correlação. Os padrões vocais esperados na infância estão acima de 250 Hz, com *pitch* agudo [10], logo, os valores obtidos no presente estudo estão de acordo com a literatura.

Destaca-se que dois participantes (números 3 e 4) estão na fase da puberdade, onde há um decréscimo gradual desse parâmetro [4] [10] [11], conforme apresenta a Tabela 4.

Para todas as categorias analisadas os valores obtidos com o sistema implementado em FPGA estão de acordo com outros estudos que abordam o padrão vocal da frequência fundamental da voz tanto de indivíduos ouvintes quanto de indivíduos que apresentam algum tipo de perda auditiva e complicações no desenvolvimento da oralidade[1] [4] [5], e tais pesquisadores utilizaram outras técnicas vigentes com apoio de *softwares* para extração os resultados são semelhantes [8] [10] [11].

Mesmo em função do pequeno número de casos analisados e alcance do estudo, nota-se que em todos os experimentos realizados a utilização do sistema configurado no FPGA mostrou-se como uma alternativa viável, sendo ainda de baixo custo para a

implementação de sistemas de Processamento Digital de Sinais de Voz.

Tabela 4. F0 de crianças sem perdas auditivas. FONTE: os autores.

Nº	Sexo	Idade	F0 /a/	F0 frase
1	Feminino	7 Anos	264 Hz	271 Hz
2	Feminino	10 Anos	253 Hz	255 Hz
3	Masculino	13 Anos	175 Hz	180 Hz
4	Feminino	12 Anos	228 Hz	235 Hz
5	Masculino	10 Anos	252 Hz	255 Hz
6	Masculino	8 Anos	275 Hz	283 Hz
7	Masculino	6 Anos	287 Hz	288 Hz
8	Feminino	8 Anos	262 Hz	265 Hz

#### 4. Conclusões

Os resultados encontrados com a extração da frequência, seu processamento e armazenamento no kit didático utilizado foram considerados satisfatórios. Os dados obtidos, quando comparados com os valores de referência destacados na literatura, possibilitam destacar que o sistema implementado em FPGA para realizar a captura e cálculo da frequência fundamental da voz atende as necessidades para o qual foi projetado.

Com a aplicação proposta, que merece ser ampliada, espera-se contribuir em estudos posteriores para o aperfeiçoamento das técnicas de avaliação, terapia e correção de características da fala. Os resultados encontrados, mesmo que em curto período e com um pequeno número de testes, mostraram que os valores da F0 estimados estão dentro dos padrões esperados para os grupos e o quanto o melhoramento deste estudo pode contribuir significativamente no processo de estimular em trabalhos acadêmicos, pesquisas que promovam alternativas viáveis no desenvolvimento de novos dispositivos de análise e correção de características vocais, sobretudo em pessoas com deficiência auditiva.

#### Bibliografia

- [1] Alves, R. L.; Soares, A. M. J.; Freire, R. C. S.; Lima, C. M. G. S. (2016) Synthesizable and prototypic visual-tactile system-in FPGA: an alternative to analysis and improvement of the voice quality for the hearing impaired people. *Revista Holos*, n. 32, v. 02. DOI: <http://dx.doi.org/10.15628/holos.2016.4086>.
- [2] Costa, N. T. (2016) *Audibilidade para fala e reconhecimento de fala em crianças com deficiência auditiva*. Tese (Doutorado). Pontifícia Universidade Católica de São Paulo. São Paulo.
- [3] Reis, D. S. (2013) *Formação docente e educação de surdos: um encontro com a diferença, cultura e identidade surda*. Dissertação de Mestrado. Universidade Federal de Rondônia. Porto Velho.
- [4] Bevilacqua, M. C.; Martinez, M. A. M.; Balen, S. A.; Pupo, A. C.; Reis, A. C. M.; Frota, S. (2011) (Orgs.). *Tratado de Audiologia*. 880p. São Paulo: Editora Santos.
- [5] Kremer, R. L.; Gomes, M. L. (2014) A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental. *ReVEL* v. 12, n. 23. <http://www.revel.inf.br>. Acesso em 22/10/2017.
- [6] Christmann, M. K.; Brancalioni, A. R.; Freitas, C. R.; Vargas, D. Z.; Keske-Soares, M.; Mezzomo, C. L.; Mota, H. B. (2015) Uso do programa MDVP em diferentes contextos: revisão de literatura. *Rev. CEFAC*, 17(4): 1341-1349.
- [7] Silva, L. M. D. *Proposta de arquitetura em hardware para FPGA da técnica Q-learning de aprendizagem por reforço* (2016). Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte, Natal- RN.
- [8] Alves, R. L. *Correção da Frequência Fundamental da Voz por meio da Realimentação Tátil e Visual*. Tese (Doutorado). Universidade do Minho, Braga: 2016.
- [9] Espindola, B. M. (2011). *Desenvolvimento e uso de módulos para processamento de sinais em FPGA*. Trabalho de Conclusão de Curso. Instituto Federal de Santa Catarina.
- [10] Behlau, M.; Madazio, G.; Feijó, D.; Pontes, P. (2001). Avaliação de voz. In: Behlau, M., Organizadora. O livro do especialista. Rio de Janeiro: Revinter, p.85-245.
- [11] Behlau, M. S.; Tosi, O. Pontes, P.L. (1985). Determinação da frequência fundamental e suas variações em altura ("Jitter") e intensidade ("Shimmer"), para falantes do português brasileiro. *Acta AWHO*; 4:5-9.
- [12] Barros, A. T. (2004). *Estimulação Tátil aplicada ao ensino da fala*. Tese de doutorado, Universidade Federal de Campina Grande.
- [13] Lejska, M. (2004). Voice field measurements: a new method of examination: the influence of hearing on the human voice. *J Voice*, 18(2):209-15.