Trabalho apresentado no XXXIX CNMAC, Uberlândia - MG, 2019.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Criação de um Modelo Estocástico para a Síntese de Vogais Considerando o Pulso Glotal de Liljencrants-Fant com Parâmetros Unificados

Diego Santana Marques Bahiano [1]
Departamento de Engenharia de Telecomunicações, UFF, Niterói, RJ
Edson Cataldo [2]
Departamento de Engenharia de Telecomunicações, Programa de Pós-graduação em Engenharia Elétrica e de Telecomunicações, UFF, Niterói, RJ

1 Introdução

O ser humano é o único ser capaz de expressar seus sentimentos, pensamentos e vontades através da voz, sendo esta um meio de comunicação de grande importância para a vida em sociedade. O aparelho fonador é constituído pelo aparelho respiratório, pela fonte de vibração (localizada na laringe) e o trato vocal (composto por faringe, boca e nariz). O fluxo aéreo respiratório, ao passar pelas cordas vocais, constituirá uma vibração que irá ressonar pelo trato vocal, baseado na teoria linear fonte-filtro de Gunnar Fant (1970) [2]. Com o avanço tecnológico e a necessidade do estudo das teorias e fenômenos associados à fala, torna-se de suma importância a análise das inúmeras variáveis relacionadas com a produção da voz, tais como a frequência fundamental do som, as medidas de perturbação (jitter e shimmer) e os ruídos. Dessa forma, a construção de modelos probabilísticos que expressem da melhor forma o sistema acústico fonador é um grande desafio para os estudiosos da área. Este artigo tem por objetivo modificar o modelo determinístico do Pulso Glotal de Liljencrants-Fant com parâmetros unificados, considerando alguns destes como processos estocásticos e criando, assim, um modelo mais próximo da realidade.

2 Metodologia

A teoria acústica apresentada por Gunnar Fant [2], em 1960, consiste em modelar o mecanismo de produção da fala como a convolução entre uma fonte de excitação (o pulso glotal) e um sistema de filtros digitais lineares, representando o trato vocal e a irradiação pelos lábios/narinas, conectados em série. A função de transferência do trato vocal se caracteriza por ter apenas pólos, que correspondem às frequências de ressonância (formantes) da fala vozeada, e larguras de banda, as quais consideram os efeitos de perdas

[1] diegobahiano@live.com
[2] ecataldo@im.uff.br

por paredes moles, fricção, condução térmica e pela irradiação nos lábios. O código implementado em Python utiliza o pulso glotal de Liljencrants-Fant com parâmetros unificados, representado na Fig. 1, para a geração de vogais.

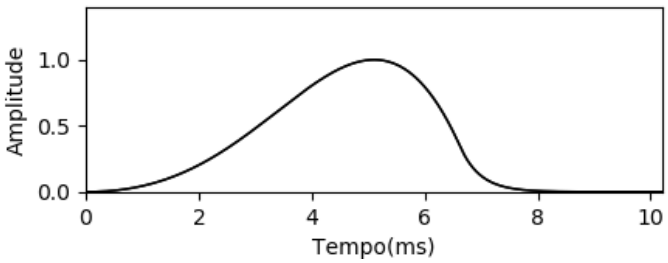

Figura 1: Pulso glotal de Liljencrants-Fant para $T_o = \frac{1}{98}s$.

A equação que representa este modelo depende de diversas variáveis, inclusive do período fundamental (T_o). Neste artigo, esta variável deixa de ser determinística e passa a ser um processo estocástico $T_o(t) = \underline{T_o}(\underline{x} + X(t))^2$, no qual $\underline{T_o} = E\{T_o(t)\}$ é o valor médio de $T_o(t)$. O processo estocástico $\{X(t), t \in \mathbb{R}\}$ e a constante real $\underline{x}$ são escolhidos de forma que, para todo t em $\mathbb{R}$, $E\{(\underline{x} + X(t))^2\} = 1$ e $E\{(\underline{x} + X(t))^4\} < \infty$. O processo X é construído como um processo estocástico gaussiano de segunda ordem, com valores em $\mathbb{R}$, estacionário e ergódico, fisicamente realizável, variando durante o trem de pulsos e criando pulsos de diferentes intervalos de tempo, gerando o fenômeno conhecido como jitter. As vogais com diferentes níveis de jitter podem ser ouvidas em:
https://www.dropbox.com/sh/lio9jg6p41l6e0n/AAB6ohnVZNrBucvTkGfFYM1Wa?dl=0.

3 Resultados e Conclusões

Com o código implementado foram obtidos sons satisfatórios, sendo que o próximo passo é estabelecer novos modelos de processos estocásticos para as variáveis do pulso glotal, a fim de estabelecer uma função densidade espectral de potência para $X(t)$ e valores de jitter que garantam um som mais natural. Possíveis ajustes no algoritmo utilizado também estão sendo analisados, uma vez que, para a geração de vogais mais próximas das que são emitidas pelo ser humano, é indispensável que o trem de pulsos seja o mais próximo possível do que é produzido na glote, além da dependência direta das formantes e larguras de banda definidas no código.

Referências

[1] E. Cataldo and C. Soize. Voice signals produced with jitter through a stochastic one-mass mechanical model. *Journal of voice*, 31 (1), 111.e9-111.e18, 2017.

[2] G. Fant. *Acoustic theory of speech production, 2nd edition.* Mouton, The Hague, 1970.