Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, v. 6, n. 1, 2018.

Trabalho apresentado no XXXVII CNMAC, S.J. dos Campos - SP, 2017.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Identificação de parâmetros em um modelo matemático para geração da voz humana

Edson Cataldo[1]
Departamento de Matemática Aplicada, Programa de Pós-graduação em Engenharia Elétrica e de Telecomunicações, UFF
Eduardo Barrientos[2]
Programa de Pós-graduação em Engenharia Elétrica e de Telecomunicações, UFF

Resumo. A variação dos ciclos glotais causada pelo movimento das cordas vocais gera o fenômeno conhecido como jitter. Sua observação sugere que o fênomeo é aleatório e é caracterizado pelos desvios dos ciclos glotais em relação a um valor médio. Seu estudo é importante em várias aplicações tais como a identificação de alguns tipos de patologias relacionadas à produção da voz. O objetivo deste artigo é usar a construção de um modelo estocástico proposto para o jitter usando um sistema massa-mola-amortecedor para a dinâmica das cordas vocais e então identificar parâmetros desse modelos levando em conta um sinal de voz obtido experimentalmente. As funções densidade de probabilidade da variável aleatória relacionada à frequência fundamental são construídas, para o caso simulado e para o caso experimental, e a distância entre elas minimizada.

Palavras-chave. Modelagem Matemática, Produção da voz humana, Modelos estocásticos.

1 Introdução

Na produção de sons chamados vozados, o fluxo de ar proveniente dos pulmões é modificado em um sinal quase periódico formado por pulsos de ar, chamado de sinal glotal, que é filtrado, amplificado pelo trato vocal, e finalmente irradiado pela boca originando a voz. Como as oscilações das cordas vocais não são exatamente periódicas, os intervalos de tempo correspondentes aos pulsos de ar, que compõem o sinal glotal, não são exatamente os mesmos e essas pequenas flutuações aleatórias são chamadas de jitter. Para discutir e desenvolver modelos de jitter, alguns argumentos envolvem o entendimento dos mecanismos que podem causar o movimento das cordas vocais ser aperiódico, e pode ser modelado usando teoria de probabilidades. Porém, em geral, autores que propõem modelos de jitter não usam modelos matemáticos e entre os que usam apenas poucos consideram modelos estocásticos [1,3,4]. Esse artigo usa um modelo estocástico para o Jitter, baseado no modelo determinístico de produção da voz introduzido por [2]. Dois parâmetros do modelo

[1]ecataldo@im.uff.br
[2]eduardo87e@gmail.com

DOI: 10.5540/03.2018.006.01.0376

estocástico são identificados resolvendo um problema inverso estocástico a partir de um sinal de voz experimental e, a partir dessa identificação, as funções densidade de probabilidade da variável aleatória correspondente à frequência fundamental do sinal simulado e do sinal real são comparadas.

2 Modelo determinístico usado

O modelo completo determinístico correspondente usado é composto de dois subsistemas acoplados pelo fluxo glotal: o subsistema das cordas vocais, chamado *fonte*, e o subsistema do trato vocal, chamado *filtro*. Durante a fonação, o filro é excitado pela sequência de pulsos do sinal glotal. Cada corda vocal é representada por um sistema massa-mola-amortecedor e um sistema simétrico é composto por duas cordas vocais. O trato vocal é representando por uma configuração padrão de tubos concatenados. A figura 1 ilustra um esquema para o modelo.

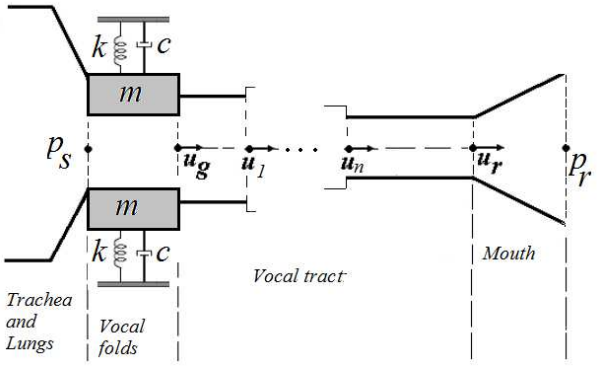

Figura 1: Esquema do modelo de Flanagan e Landgraf (1968).

3 Modelo estocástico do jitter

Consideremos $\{K(t), t \in \mathbb{R}\}$ um processo estocástico indexado por $\mathbb{R}$, com valores em $\mathbb{R}^+$, que modela a rigidez k. A equação da dinâmica das cordas vocais em $x(t)$, a posição das cordas vocais, depende de $u_g(t)$, o fluxo glotal, torna-se uma equação diferencial estocástica não-linear para o processo estocástico $X(t)$ acoplado com o processo estocástico U_g dada pela equação 1:

$$m\,\frac{d^2X(t)}{dt^2} + \{c + c^*(X(t))\}\,\frac{dX(t)}{dt} + K(t)\,X(t) + a_1\,p_B(X(t), U_g(t)) = a_2\,p_s(t)\,, \qquad (1)$$

onde m é a massa correspondente das cordas vocais, c e c^* são valores relacionados com o amortecimento das cordas vocais, a_1 e a_2 são constantes, p_B é uma função relacionada

DOI: 10.5540/03.2018.006.01.0376

à pressão do ar, e $p_s(t)$ é a pressão subglotal. Sob as hipóteses que devem ser satisfeitas para $K(t)$, o modelo escolhido será dado pela equação 2:

$$K(t) = k_0 + (\underline{k} - k_0)(\underline{z} + Z(t))^2\,, \tag{2}$$

onde k_0 e $\underline{k}$ são duas constantes. O processo estocástico Z e a constante real $\underline{z}$ devem ser construídos tais que, para todo t em $\mathbb{R}$, $E\{(\underline{z} + Z(t))^2\} = 1$ e $E\{(\underline{z} + Z(t))^4\} < +\infty$. Consequentemente, $\underline{k} = E\{K(t)\}$ é o valor médio de $K(t)$. O processo estocástico $\{Z(t), t \in \mathbb{R}\}$ é construído como um processo estocástico gaussiano de segunda ordem, indexado por $\mathbb{R}$, com valores em $\mathbb{R}$,centrado, contínuo em média-quadrática, estacionário e ergódico, fisicamente realizável, com função densidade espectral dada pela equação 3:

$$S_Z(\omega) = \frac{1}{2\pi}\frac{a^2}{\omega^2 + b^2} \quad , \quad a > 0 \quad , \quad b > 0\,, \tag{3}$$

onde a e b devem satisfazer $E\left\{(\underline{z} + Z(t))^2\right\} = 1$ que pode se escrever como a equação 4:

$$\underline{z}^2 + \int_{-\infty}^{+\infty} \frac{a^2}{2\pi(\omega^2 + b^2)} d\omega = 1 \quad \Longrightarrow \quad \underline{z}^2 = 1 - \frac{a^2}{2b}\,, \tag{4}$$

com as seguintes desigualdades para a e b,

$$0 < a < \sqrt{2b} \quad , \quad b > 0. \tag{5}$$

Consequentemente, o processo estocásticos Gaussiano Z pode ser visto como o filtro linear $Z = h * N_\infty$ do ruído branco centrado gaussiano N_∞ (processo estocástico generalizado) cuja função densidade espectral é $S_{N_\infty}(\omega) = 1/(2\pi)$, pelo filtro linear causal e estável cuja função resposta em frequência é $\widehat{h}(\omega) = \int_0^{+\infty} e^{-i\omega t} h(t)\, dt = a/(i\omega + b)$ (porque $S_Z(\omega) = |\widehat{h}(\omega)|^2\, S_{N_\infty}(\omega)$). Introduzindo a equação diferencial estocástica de Itô

$$dY(t) = -b\, Y(t)\, dt + a\, dW(t) \quad t > 0\,, \tag{6}$$

com a condição inicial $Y(0) = 0$ *a.s.*, onde W processo de Wiener indexado por $[0, +\infty[$, pode ser provado (Soize, 1994) que a Eq. (6) tem uma única solução $\{Y(t), t \geq 0\}$ tal que para $t_0 \to +\infty$, o processo estocástico $\{Y(t), t \geq t_0\}$ é estocasticamente equivalente ao processo estocástico estacionário Z. Na prática, isso significa que, se t_0 é escolhido suficientemente grande, Y e Z são o mesmo processo estocástico gaussiano centrado estacionário e ergódico de segunda ordem para o qual a função densidade espectral de potência é dada pela Eq. (3). Consequentemente, Eq. (6) pode ser usada para gerar trajetórias do processo Z. Alguns resultados obtidos com a síntese de vogais, no caso determinístico, e com dois diferentes níveis de jitter ($a = 40$ and $a = 160$) podem ser ouvidos em https://www.dropbox.com/s/zuvb1jbge8n9a72/sintvogaisjitter.zip?dl=0 . A ideia é então resolver um processo estocástico inverso para identificar os parâmetros a e b associados com vozes reais obtidas experimentalmente.

DOI: 10.5540/03.2018.006.01.0376

4 Problema inverso estocástico correspondente

Consideremos a duração entre dois sucessivos instantes, o primeiro correspondente ao instante em que a glote abre e o segundo instante quando ela fecha completamente. Essa duração, denotada por T_{fund} é uma variável aleatória e seu inverso é a variável aleatória $F_{fund} = 1/T_{fund}$. O objetivo é então identificar parâmetros a e b tais que a função densidade de probabilidade $f \mapsto f_S(f; a, b)$ definida em $[0\,, +\infty[$, da variável aleatória $F_{fund}(a, b)$ é próxima da função densidade de probabilidade $f \mapsto f_R(f)$ em $[0\,, +\infty[$, associada com a voz real. A distância entre as duas funções densidade de probabilidade será dada pela equação 7:

$$J(a, b) = \int_0^{+\infty} \mid f_S(f; a, b) - f_R(f) \mid \, df\,. \tag{7}$$

Os valores ótimos a_{opt} e b_{opt} são calculados resolvendo o problema de otimização:

$$(a_{opt}, b_{opt}) = \min_{(a,b)\in\mathcal{C}} J(a, b)\,, \tag{8}$$

no conjunto admissível $\mathcal{C}$, usando a equação (5)

$$\mathcal{C} = \{(a, b) \in \mathbb{R}^2 \quad \text{such that} \quad 0 < a < \sqrt{2b} \quad \text{and} \quad b > 0\}\,. \tag{9}$$

As funções densidade de probabilidade são estimadas usando o método de estimação do kernel gaussiano da estatística não-paramétrica. Os valores dos parâmetros considerados para o modelo determinístico correspondente são:
$A_{g0} = 0.05 \times 10^{-2}\, m^2$, $\rho = 0.12\, kg/m^3$, $c_a = 346.3\, m/s$, $\mu = 1.86 \times 10^{-4}\, kg/(m^2 s)$, $m = 0.24 \times 10^{-2}\, kg$, $\ell = 1.4 \times 10^{-2} m$, d $= 0.3 \times 10^{-2}\, m$, $k_0 = 40\, N/m$, $a_1 = 1.87\dfrac{\ell d}{2}$ e $a_2 = \dfrac{\ell d}{2}$. Para o coeficiente de amortecimento foram considerados $c = 0$ e $\alpha = 1$, i.e, somente durante a colisão o amortecimento foi considerado. O método de Monte Carlo foi usado.

DOI: 10.5540/03.2018.006.01.0376

Um sinal experimental foi considerado, denotado por *exper*, que é o sinal de voz de uma mulher produzindo um /a/. A pressão de saída é mostrada na figura 2. Após resolver

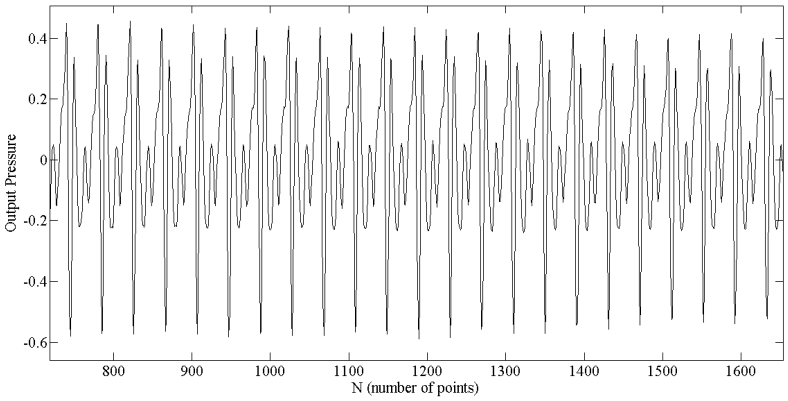

Figura 2: Uma janela da pressão de saída do sinal correspondente à produção da vogal /a/.

o problema estocástico inverso, os valores ótimos obtidos foram $a_{opt} = 3.2834$ e $b_{opt} = -1000$. A figura 3 mostra as duas funções densidade de probabilidade $f_S(.\,; a_{opt}, b_{opt}))$ e f_R (simulada e experimental). O valor correspondente para $J(a_{opt}, b_{opt})$ é 0.098. O

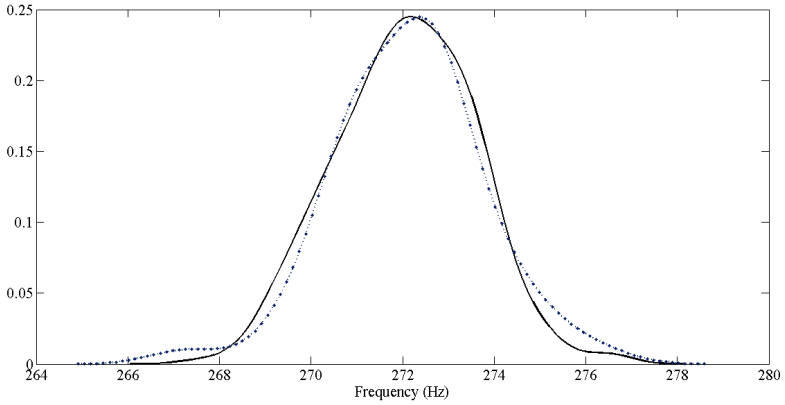

Figura 3: pdf da frequência fundamental aleatória correspondente à voz real (linha cheia) e correspondente à simulada com os valores ótimos dos parâmetros (linha pontilhada).

sinal experimental, *exper*, e o simulado, $Simulated_exper$, com os parâmetros identificados podem ser ouvidos seguindo o link

https://www.dropbox.com/s/52sgf949ybmblb0/sinaisCNMAC2017.zip?dl=0 .

5 Conclusões

Uma abordagem foi proposta para identificar um modelo estocástico que permite gerar os efeitos do jitter em um sinal de voz simulada usando um sinal experimental e um modelo matemático/mecânico que produz voz simulado com efeito de jitter. Com os valores ótimos dos parâmetros do modelo identificado, a pdf construída a partir das vozes simuladas está muito próxima da pdf construída para vozes reais, mostrando que o modelo estocástico considerado é um bom ponto de início para gerar o jitter. O próximo passo é o de usar parâmetros adicionais do modelo de processo estocástico para melhorar a aproximação entre os processos estocásticos correspondentes às vozes consideradas.

Agradecimentos

Os autores agradecem ao CNPq.

Referências

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

[2] J. Flanagan and L. Landgraf. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics AU-16*, 57–64, 1968.

[3] J. Schoengten and R. De Guchteneere. Predictable and random components of jitter. *Speech Communication*, 21, 255–272, 1997.

[4] J. Schoengten, S. Fraj and J. C. Lucero. Testing the reliability of grade, roughness and breathiness scores by means of synthetic speech stimuli. *Logopedics Phoniatrics Vocology*, 40(1) 5–13, 2015.

[5] C. Soize. The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions. *World Scientific*, Singapore, 1994.

