Josue Valentin Uscata Barrientos

**Modelagem Estocástica de Voz e Classificação
dos Sinais Obtidos Usando Redes Neurais
Artificiais**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Mecânica da PUC-Rio

Orientador Prof. Sampaio, Rubens
Co-Orientador Prof. Cataldo, Edson

Rio de Janeiro
novembro de 2010

Josue Valentin Uscata Barrientos

**Modelagem Estocástica de Voz e Classificação
dos Sinais Obtidos Usando Redes Neurais
Artificiais**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Mecânica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Sampaio, Rubens

Orientador

Departamento de Engenharia Mecânica –PUC-Rio

Prof. Cataldo, Edson

Co-Orientador

Departamento de Matemática Aplicada
Programa de Mestrado em Engenharia de
Telecomunicações–UFF

Prof. Hans Ingo Weber

Departamento de Engenharia Mecânica –PUC-Rio

Prof. Fernando Alves Rochinha

Departamento de Engenharia Mecânica – UFRJ

Prof. Thiago Gamboa Ritto

Departamento de Engenharia Mecânica – UFF

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 30 de novembro de 2010

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Josue Valentin Uscata Barrientos

Graduou-se em Engenharia Mecatronica na Universidad Nacional de Ingenieria Lima-Perú em 2006.

Ficha Catalográfica

Uscata Barrientos, Josue Valentin

Modelagem Estocástica de Voz e Classificação dos Sinais Obtidos Usando Redes Neurais Artificiais / Josue Valentin Uscata Barrientos; orientador: Sampaio, Rubens; co-orientador: Cataldo, Edson. — 2010.

100 f: il.(color) ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Mechanical Engineering, 2010.

Inclui bibliografia.

1. Engenharia Mecânica— Teses.
2. Modelos Mecânicos; . 3. Produção da Voz; . 4. Mecânica Estocástica; . 5. Redes Neurais Artificiais.. I. Sampaio, Rubens. II. Cataldo, Edson. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Mechanical Engineering. IV. Título.

CDD: 621

Agradecimentos

A Deus por tudo que me deu. Aos meus pais, Maximo e Maxima, aos meus irmãos, Edgar e Marcos, à minha família em geral, ao meu orientador, Professor Rubens Sampaio, e ao meu co-orientador, Professor Edson Cataldo, pelo apoio, simpatia de sempre, e incentivo para a realização deste trabalho, aos meus amigos e colegas da PUC-Rio, com quem compartilhei muitos momentos agradáveis. Para todos que foram gentis comigo durante minha estadia no Brasil.

Resumo

Barrientos, Josue Valentin Uscata; Sampaio, Rubens; Cataldo, Edson. **Modelagem Estocástica de Voz e Classificação dos Sinais Obtidos Usando Redes Neurais Artificiais**. Rio de Janeiro, 2010. 100p. Dissertação de Mestrado — Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo desta dissertação é classificar sinais de vozes, usando redes neurais, obtidos por meio de um modelo mecânico-estocástico para produção da voz humana, esse modelo foi construído a partir de uma abordagem probabilística não-paramétrica para considerar incertezas do modelo. Primeiro, uma rede neural artificial foi construída para classificar sinais de vozes reais, normais e provenientes de sujeitos com patologias nas cordas vocais. Como entradas da rede neural foram usadas medidas acústicas extraídas dos sinais glotais, obtidos por filtragem inversa dos sinais de vozes reais. Essa rede neural foi usada, posteriormente, para classificar sinais de vozes sintetizadas geradas por um modelo estocástico da produção da voz humana, no caso particular da geração de vogais. O modelo estocástico da produção da voz humana foi construído tomando por base o modelo determinístico criado por Ishizaka e Flanagan. Incertezas do modelo foram consideradas através de uma abordagem probabilística não-paramétrica de modo que matrizes aleatórias foram associadas às matrizes de massa, rigidez e amortecimento do modelo. Funções densidade de probabilidade foram construídas para essas matrizes, usando o Princípio da Máxima Entropia. O método de Monte Carlo foi usado para gerar realizações de sinais de vozes. Os sinais obtidos foram então classificados usando a rede neural construída previamente. Das realizações obtidas, alguns sinais de vozes foram classificados como normais, porém outros foram classificados como provenientes de sujeitos com patologias nas cordas vocais. Os sinais com características de patologia foram classificados em três grupos: nódulo, paralisia unilateral e outras patologias.

Palavras-chave

Modelos Mecânicos; Produção da Voz; Mecânica Estocástica; Redes Neurais Artificiais.

Abstract

Barrientos, Josue Valentin Uscata; Sampaio, Rubens (Advisor) ; Cataldo, Edson (Co-Advisor). **Stochastic Voice Modeling and Classification of the Obtained Signal Using Artificial Neural Networks**. Rio de Janeiro, 2010. 100p. MSc. Dissertation — Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

The aim of this thesis is to classify voice signals, using neural networks, obtained through a mechanical stochastic model for voice production, this model was built from a nonparametric probabilistic approach to take into account modeling uncertainties. At first, an artificial neural network was constructed to classify real voice signals, normal and produced by subjects with pathologies on the vocal folds. As inputs for the neural network were used acoustic measures extracted from the glottal signals, obtained by inverse filtering of the real voice signals. This neural network was used, later, to classify synthesized voice signal generated by a stochastic model of the voice production, in the particular case of vowels generation. The stochastic model was constructed from the corresponding deterministic model created by Ishizaka and Flanagan, in 1972. Modeling uncertainties were taken into account through a nonparametric probabilistic approach such that random matrices were associated to mass, stiffness and damping model matrices. Probability density functions were constructed for these matrices using the Maximum Entropy Principle. The Monte Carlo Method was used to generate realizations of the voice signals. The voice signals obtained were then classified using the neural network previously constructed. From the realizations obtained, some voice signals were classified as normal, but others were classified as produced by subjects with pathologies on the vocal folds. The signal with pathologies characteristics were classified into three groups: nodulus, unilateral paralysis and other pathologies.

Keywords

Mechanical Models; Voice Production; Stochastic Mechanics; Artificial Neural Networks.

Sumário

1	Introdução	12
1.1	Motivação	13
1.2	Metodologia	13
1.3	Objetivos	14
1.4	Organização da Dissertação	14
2	Fund. da Produção de Voz e Patologias das Cordas Vocais	16
2.1	Introdução	16
2.2	Produção de Sinal de Voz	16
2.3	Patologias das Cordas Vocais	17
3	Características do Sinal de Voz e do Sinal Glotal	20
3.1	Sinal Glotal	20
3.2	Características do Sinal Glotal	20
3.3	Sinal de Voz	25
3.4	Características do Sinal de Voz	26
4	Redes Neurais Artificiais	29
4.1	Introdução	29
4.2	Estrutura do neurônio	30
4.3	Tipos de Arquitetura das Redes Neurais - Rede Multilayer Perceptron (MLP)	33
4.4	Algoritmo Retro-propagação	37
5	Modelagem Mecânica Determinística da Produção da Voz	43
5.1	Introdução	43
5.2	Modelo de Ishizaka e Flanagan	43
5.3	Outros Modelos de Produção da Voz	50
6	Mod. Est. de Inc. e sua Apl. no Caso de Prod. de Voz	52
6.1	Modelagem Estocástica de Incertezas	52
6.2	Abordagem Paramétrica Aplicado ao Modelo de Ishizaka e Flanagan	56
6.3	Abordagem Não Paramétrica Aplicada ao Modelo de Ishizaka e Flanagan	59
7	Avaliação de sinais de Voz	67
7.1	Processo de Aprendizado.	68
7.2	Processo de Avaliação.	82
8	Conclusões.	87
A	Sinais de Voz Gerados	94

Lista de figuras

2.1	Esquema do Aparelho Fonador (19)	17
2.2	Vista superior das cordas vocais.	17
2.3	Pequenos nódulos nas cordas vocais.	18
2.4	Paralisia unilateral da corda vocal esquerda.	18
3.1	Esquema de um pulso glotal.	20
3.2	Fase fechada do ciclo glotal.	21
3.3	Fase de abertura do ciclo glotal.	21
3.4	Fase de fechamento do ciclo glotal.	22
3.5	Fase aberta do ciclo glotal.	22
3.6	Fase fechada do ciclo glotal.	23
3.7	Ciclo glotal com fase primaria e secundaria de abertura.	23
3.8	Exemplo de um sinal de voz correspondente a um trecho da vogal /a/.	26
3.9	Exemplo de um sinal de voz real na produção de um /a/.	26
4.1	Esquema básico do neurônio biológico (24).	29
4.2	Modelo básico de um neurônio artificial de tipo Perceptron (24).	30
4.3	Funções de Ativação: (a) Função Degrau. (b) Função linear. (c) Função Logsig. (d) Função Tansig. (24).	32
4.4	Arquitetura de uma rede neural MLP (n-i-j-k). (24).	34
4.5	Configuração da rede MLP (24).	35
4.6	Treinamento supervisionado (24).	37
4.7	(a) Fase de propagação (b) Sinais funcionais. (24).	38
4.8	(a) Fase de Retro-propagação (b) Sinais de erro (24).	39
4.9	Cálculo de erro e_j na camada de saída (24).	41
4.10	Calculo de erro e_j na camada escondida (24).	42
5.1	Esquema do modelo de Ishizaka e Flanagan.	43
5.2	Equivalência eletro-acústica de um guia de ondas planas.	46
5.3	Representação por linha de transmissão do sistema fonador.	46
5.4	Modelo mecânico para as cordas vocais e (b) sistema vocal utilizado. (Flanagan e Landgraf modelo, 1968).	50
5.5	Representação da glote no modelo de Lous et al.	51
6.1	Diferentes tipos de incertezas influenciando o resultado.	53
6.2	Função distribuição de probabilidade experimental (em cima) e função distribuição de probabilidade simulada(abaixo).	59
6.3	Convergência de F_0 com o número de amostras.	65
6.4	Distribuição de probabilidade estimada.	66
6.5	Pressão radiada, similar a pressão gerada pelo modelo médio.	66
6.6	Pressão radiada não periódica, similar à observada com papilomas nas cordas.	66
6.7	Pressão radiada similar à observada com papilomas nas cordas.	66
6.8	Pressão radiada similar à observada com nódulos nas cordas.	66

7.1	Processos envolvidos na identificação de patologias.	68
7.2	Processo que envolve os passos para a aprendizagem das RNA's.	68
7.3	Coeficiente de dispersão em percentagem da característica acústica <i>Mean Pith</i> .	73
7.4	Coeficiente de dispersão em percentagem das características acústicas <i>Jitter(local)</i> e <i>Jitter(rap)</i> .	73
7.5	Coeficiente de dispersão em percentagem das características acústicas <i>Jitter(absolute)</i> e <i>Jitter(ppq5)</i> .	74
7.6	Coeficiente de dispersão em percentagem das características acústicas <i>Jitter(ddp)</i> e <i>Mean Autocorrelation</i> .	74
7.7	Coeficiente de dispersão em percentagem das características acústicas <i>Mean harmonics – to – noise ratio</i> e <i>Mean noise – to – harmonics ratio</i> .	74
7.8	Características acústicas normalizadas para observar suas tendências.	75
7.9	Comparação de perfis a) Entre as cinco características acústicas <i>Jitters</i> , e b) Entre as características acústicas <i>Jitter (ppq5)</i> e <i>Mean Noise – to – harmonics ratio</i> .	76
7.10	Características acústicas escolhidas oferecidas pelo PRAAT.	76
7.11	Melhor visualização das características acústicas: a) HRF e b) dH12.	77
7.12	Ampliação das características acústicas: a) HRF e b) dH12.	77
7.13	Comparação das características acústicas normalizadas CIQ, OQ1 e OQ2.	78
7.14	Coeficiente de dispersão das características acústicas CIQ, OQ1 e OQ2.	78
7.15	Características acústicas normalizadas QOQ, NAQ e OQa.	79
7.16	Coeficiente de dispersão das características acústicas QOQ, NAQ e OQa.	79
7.17	Características acústicas normalizadas SQ1 e SQ2.	79
7.18	Coeficiente de dispersão das características acústicas SQ1 e SQ2.	80
7.19	Características acústicas escolhidas oferecidas pelo APARAT.	80
7.20	Processo que envolve os passos para a avaliação das vozes.	83
7.21	Sinal obtido do modelo mecânico estocástico	83
A.1	Sinal 1 gerado pelo Modelo Mecânico Estocástico.	94
A.2	Sinal 2 gerado pelo Modelo Mecânico Estocástico.	94
A.3	Sinal 3 gerado pelo Modelo Mecânico Estocástico.	95
A.4	Sinal 4 gerado pelo Modelo Mecânico Estocástico.	95
A.5	Sinal 5 gerado pelo Modelo Mecânico Estocástico.	95
A.6	Sinal 6 gerado pelo Modelo Mecânico Estocástico.	96
A.7	Sinal 7 gerado pelo Modelo Mecânico Estocástico.	96
A.8	Sinal 8 gerado pelo Modelo Mecânico Estocástico.	96
A.9	Sinal 9 gerado pelo Modelo Mecânico Estocástico.	97
A.10	Sinal 10 gerado pelo Modelo Mecânico Estocástico.	97
A.11	Sinal 11 gerado pelo Modelo Mecânico Estocástico.	97
A.12	Sinal 12 gerado pelo Modelo Mecânico Estocástico.	98
A.13	Sinal 13 gerado pelo Modelo Mecânico Estocástico.	98
A.14	Sinal 14 gerado pelo Modelo Mecânico Estocástico.	98
A.15	Sinal 15 gerado pelo Modelo Mecânico Estocástico.	99
A.16	Sinal 16 gerado pelo Modelo Mecânico Estocástico.	99

A.17 Sinal 17 gerado pelo Modelo Mecânico Estocástico.	99
A.18 Sinal 18 gerado pelo Modelo Mecânico Estocástico.	100
A.19 Sinal 19 gerado pelo Modelo Mecânico Estocástico.	100
A.20 Sinal 20 gerado pelo Modelo Mecânico Estocástico.	100

Lista de tabelas

7.1	Resumo de sinais de voz no banco de dados.	69
7.2	Características acústicas oferecidas pelo PRAAT.	70
7.3	Características acústicas usadas do PRAAT.	70
7.4	Exemplo das medidas obtidas pelo APARAT.	71
7.5	Médias para cada medida acústica.	72
7.6	Coeficientes de dispersão para cada medida acústica.	72
7.7	Características acústicas selecionadas como entradas da rede neural artificial.	80
7.8	Resumo dos resultados da classificação de vozes.	82
7.9	Resumo dos resultados da classificação de vozes.	82
7.10	Exemplo das características acústicas obtidas pelos programa PRAAT e APARAT.	84
7.11	Resumo dos resultados da classificação de vozes.	85
7.12	Resultado da avaliação na identificação de patologias.	85
7.13	Resultado da avaliação na identificação de patologias.	86