IDENTIFICAÇÃO NÃO-PARAMÉTRICA DE NÚCLEOS MEMS UTILIZANDO REDES NEURAIS RECORRENTES

Marcos B. Ketzer[1]; Manuel M. P. Reimbold[2];
Universidade Regional do Noroeste do Estado do Rio Grande do Sul - UNIJUÍ
Caixa Postal 480 – 98700-000 – Rio Grande – RS
{marcos.ketzer, manolo}@unijui.edu.br

Resumo. *Este trabalho apresenta um método de identificação comportamental não-paramétrica de núcleos de micro-resonadores. Para a identificação do sistema é utilizado uma rede neural recorrente, com capacidade de representação de termos lineares e não-lineares. Uma plataforma de testes é criada no software de elementos finitos ANSYS®. Os dados de entrada e saída obtidos da plataforma são utilizados para o treinamento da rede neural. O algoritmo de aprendizagem utilizado é o back-propagation. Ao final são obtidos dados que validam a rede neural treinada.*

Palavras-chave: MEMS, Identificação de Sistemas, Redes Neurais Artificiais.

1. INTRODUÇÃO

A modelagem e simulação de micro-estruturas é uma tecnologia emergente que vem ganhando interesse nos setores: automotivos, biomédicos, de telecomunicações, industrial, aeroespacial e militar. Dadas as suas dimensões de micrometros e sua alta-fragilidade que dificultam seu estudo, uma solução adotada é a otimização dos macromodelos que emulam o comportamento característico dos dispositivos.

Na aplicação de produção em lote (*batch*), as técnicas para testes e avaliação de microdispositivos, visando à garantia e confiabilidade do produto, se tornam ainda mais relevantes, pois se relacionam ao tempo de processo e ao custo do produto final. Sendo assim é necessário o desenvolvimento de técnicas de avaliação do comportamento destes microssistemas, de forma intuitiva, inteligente e semi-automática.

Neste contexto o uso de técnicas de identificação de sistemas, o qual não se vê restrito só à MEMS (*Microeletromechanical Systems*), se mostra uma alternativa ao problema. Estes

1 – Acadêmido do curso Engenharia Elétrica, bolsista PIBIC-UNJUÍ.

2 – Professor Mestre do curso de Engenharia Elétrica.

algoritmos possuem um legado na literatura técnica, como o que pode ser visto em Aguirre (2004) e Ljung (1999), que facilita a aplicação dos mesmos.

Os métodos de identificação de sistemas, segundo a literatura Aguirre (2004), podem ser divididos em paramétricos e não-paramétricos. Os métodos paramétricos são aqueles que resultam num modelo matemático tal como a função de transferência, cujos parâmetros podem ter uma interpretação física. Os métodos não-paramétricos têm por característica a total ausência a respeito da forma de função, a qual deve ser estimada a partir de ajustes de parâmetros livres. Por fim, esta resulta num modelo com uma quantia elevada de parâmetros que não tem uma interpretação individual.

Com base nestes conceitos, as redes neurais têm características de apresentarem, em certo grau, a modelagem não-paramétrica, pois modelam sistemas sem se preocupar com a ordem ou tipo dos modelos em si, mas buscando uma generalização da sua representação.

Sendo assim, o objetivo deste trabalho é a utilização de redes neurais para criar um modelo com capacidade de "imitar" o comportamento da planta estudada, que neste caso é um MEMS. Desta forma se torna possível realizar testes e avaliações que necessitam de grande quantidade de dados, sem a utilização da plataforma, mas através da generalização a partir da rede neural, tornando o processamento rápido e intuitivo.

2.0 METODOLOGIA

2.1 MEMS

A topologia MEMS em estudo é conhecida na literatura por comb-drive. O desenho do microssistema pode ser visto na figura 1. O comb-drive funciona com base na força eletrostática desenvolvida pelo arranjo capacitivo. A força desenvolvida, perpendicular a estrutura móvel, é responsável pelo movimento da estrutura.

O equacionamento do núcleo resulta num sistema típico massa-mola-amortecedor, que pode ser visto na equação 1. É conhecido que o efeito capacitivo não é linear. Devido a propriedades intrínsecas o sistema real não se comporta de maneira exatamente linear como o modelo propõe.

$$x(t) = \left[-\frac{d^2x(t)}{dt^2} - \frac{dx(t)}{dt} + f_e(t) \right] \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \qquad (1)$$

Onde:

$$M = \frac{\alpha}{\gamma} \qquad D = \frac{\beta}{\gamma} \qquad K = \frac{1}{\gamma}$$

x(t) = deslocamento do núcleo;

$f_e(t)$ = força eletrostática;

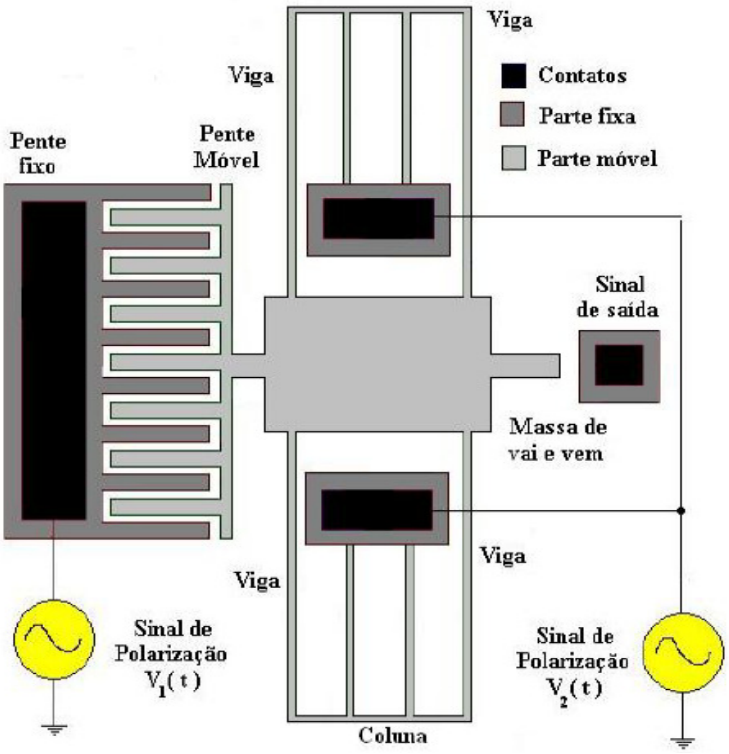
M = Massa equivalente da estrutura ou micro-núcleo;

D = Viscosidade equivalente;

K = Elasticidade equivalente;

Maiores detalhes sobre o funcionamento do micro-ressonador são discutidos em Tang (1990).

Figura 1. Topologia MEMS em estudo.



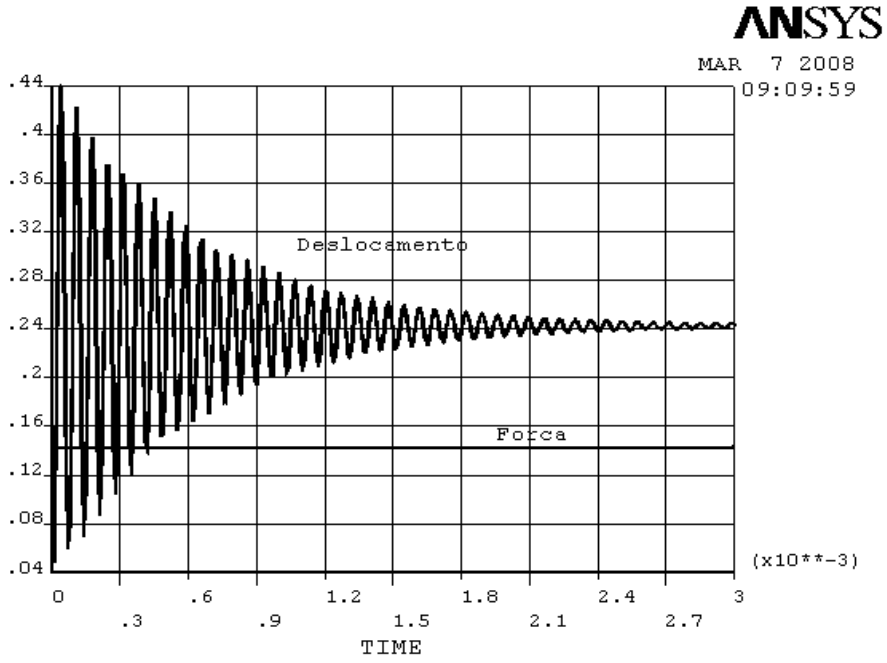
2.2 Plataforma de testes

A coleta de dados é feita a partir da plataforma de testes desenvolvida com o aplicativo ANSYS ®. A técnica dos elementos finitos consiste em dividir o modelo geométrico espacial em pequenos elementos, com as propriedades de interesse modeladas nos nós dos elementos. Posteriormente, o modelo é solucionado através de métodos numéricos. A escolha dos elementos esta diretamente relacionada aos fenômenos físicos de interesse. Assim também, as

técnicas de resolução das equações diferencias parciais e o passo de discretização, tanto com relação ao tempo como em relação ao espaço, determinam a precisão do método. A modelagem por elementos finitos permite que um sistema, previamente não modelado, seja resolvido como um conjunto de sistemas menores cujas propriedades físicas são consagradas. Sendo assim, é possível observar comportamento do sistema, independente de sua complexidade, apesar de ser limitado em precisão.

Os dados de entrada e saída obtidos da plataforma criada são utilizados para o treinamento da rede neural. A resposta do comb-drive ao sinal degrau pode ser visto na figura 2.

Figura 2. Resposta ao degrau obtida da plataforma de teste.



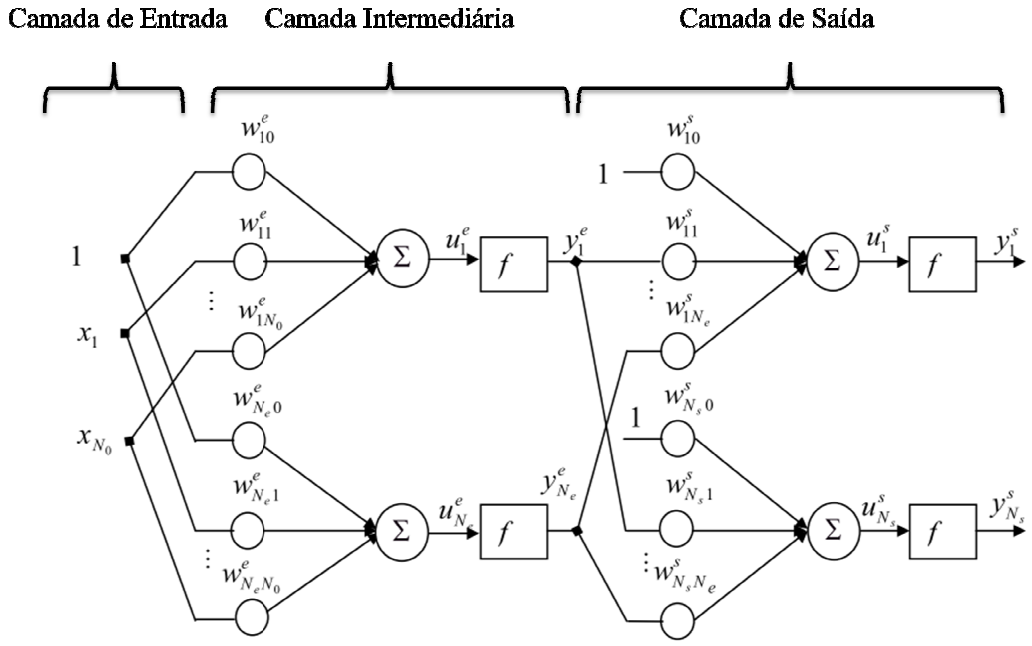
2.3 Redes Neurais

As redes neurais artificiais são sistemas computacionais que imitam funcionalidades do sistema nervoso biológico. O neurônio é sintetizado através de uma função que tem sua saída relacionada às suas entradas que podem ser de origem externa a rede, ou proveniente de outros neurônios. O algoritmo de aprendizado corrige a influencia destas entradas, de forma a "mapear" o problema na rede. Através de exemplos diferentes, a rede é treinada e generaliza as situações a que foi submetida.

Duas classes de redes neurais têm recebido atenção: as redes neurais estáticas e as redes neurais recorrentes.

As redes neurais estáticas assim como recorrentes possuem três camadas: camada de entrada, camada intermediária e camada de saída. A figura 3 mostra o diagrama da organização da rede neural.

Figura 3. Camadas da Rede Neural.



A rede neural da figura 3 apresenta a saída conforme a equação 2.

$$\hat{f}(\mathbf{x}) = F_s\left[\mathbf{W}F_e\left[\mathbf{W}^e\mathbf{x}\right]\right] \tag{2}$$

Onde $\mathbf{W}$ contém os pesos ajustáveis, $\mathbf{x}$ é o vetor de entradas e F_s e F_s são funções vetoriais. Estas camadas são comumente aplicadas em problemas de reconhecimento de padrões, onde os pesos são ajustados no sentido de otimizar a saída da rede. A saída da rede, em forma escalar, para uma rede neural de uma camada, é dada pela equação 3.

$$\begin{aligned} u^s{}_k(l) &= \sum_{j=1}^{Ne} w^s{}_{kj}\, y^e{}_j(l) + w^s{}_{k0} \\ &= \sum_{j=1}^{Ne} w^s{}_{kj}\, f(u^e{}_j(l)) + w^s{}_{k0} \\ &= \sum_{j=1}^{Ne} w^s{}_{kj}\, f\Big(\sum_{j=1}^{No} w^s{}_{kj}\, x_i(l) + w^s{}_{k0}\Big) + w^s{}_{k0} \end{aligned} \tag{3}$$

Onde,

$\boldsymbol{w}$ = peso;

$\boldsymbol{x} = [\boldsymbol{x_1}...\boldsymbol{x_{N0}}]^{\mathbf{T}}$ é o vetor de entrada;

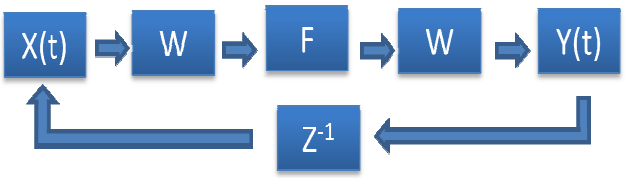
$\boldsymbol{y} = [\boldsymbol{y^s_1}... \boldsymbol{y^s_1}]^{\mathbf{T}}$ é o vetor de saída;

Sendo a camada de saída '*j*' e a camada intermediária '*i*', e a terminação '*0*' uma constate (bias).

As redes neurais recorrentes, mais conhecidas a partir dos trabalhos de Hopfield, apresentam a capacidade de memória endereçada por conteúdo, que permite a solução de problemas mais complexos, entre esses, a identificação de sistemas dinâmicos.

Originalmente as redes neurais recorrentes, como proposta por Hopefield, eram de uma única camada. Porém, algumas topologias com diferentes posições dos perceptrons e dos deslocamentos foram desenvolvidas mais tarde. A rede neural recorrente utilizada neste trabalho é também multicamada. A topologia simplificada da rede neural recorrente utilizada na identificação dos micro-núcleos pode ser visto na figura 4.

Figura 4. Topologia de rede neural desenvolvida.



O algoritmo *back-propagation* tem base no gradiente descente. Os pesos são ajustados conforme a equação 3.

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{a\partial J(\mathbf{w})}{\partial \mathbf{w}} \tag{4}$$

Onde, $\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}$ é função do erro e da derivada da função de ativação. Maiores detalhes sobre o algoritmo de aprendizagem *back-propagation* pode ser visto em Haykin (2001).

Considerando as dimensões do problema deve ser dimensionado o tamanho e estrutura da rede. Em geral, o tamanho da rede proporciona a esta, a capacidade de representar dados mais complexos, porém com um processo mais lento. Além disso, deve se considerar que redes superdimensionadas causam imprecisão numérica. Sendo assim é preciso um ajuste ótimo que estabeleça a melhor precisão que a rede pode oferecer, sem requerer muito processo e memória.

Ainda além do tamanho da rede é necessário escolher as funções de ativação. As funções de ativação determinam as capacidades de representação da rede, estando intimamente relacionadas aos erros de convergência numérica.

Conhecendo a modelo da planta como a equação 1, a ordem é escolhida de forma coerente. Para este trabalho foram utilizados 4 parâmetros de entrada e 4 parâmetros como realimentação da saída, sendo cada um destes deslocados em 1 amostra (z^{-1}), sucessivamente. A camada oculta têm 5 neurônios, a função de ativação da camada de saída é linear, e para a camada intermediária é utilizada a função tangente hiperbólico. Para modelagem de sistemas dinâmicos, cabe dizer que a utilização de funções de ativações lineares possui melhor desempenho. Isto porque, a grande maioria destes sistemas necessita representar os sinais em diversas amplitudes. As funções lineares permitem uma melhor convergência da rede para estes casos, e reduz os erros por truncamento numérico.

3. RESULTADOS

Os resultados obtidos com a rede neural foram comparados com as obtidas através de um modelo ARMAX utilizando o estimador estendido dos mínimos quadrados. As discussões do modelo e do estimador podem ser encontradas em Aguirre (2004) e Ljung (1999).

Para a determinação do modelo a ser estimado, é adotada a identificação caixa-cinza. Esta modalidade utiliza a equação 1, e aplica métodos de discretização de equações diferenciais. Os coeficientes são organizados no vetor de regressores. Os resultados obtidos com o modelo linear estimado e a resposta da plataforma podem ser vistos na figura 5. O erro residual entre os o modelo linear identificado e o ANSYS é visto na figura 7.

Para treinamento da rede, esta é considerada estática (identificação *off-line*). Para isto são utilizados pontos que correspondem aos deslocamentos da rede recorrente. A utilização dos dados é totalmente aleatória de forma a permitir a rede generalizar o comportamento para todo o intervalo igualmente.

Para a rede neural treinada, a verificação é obtida com a excitação direta da rede, utilizando o modelo neural recorrente. Os resultados obtidos pela rede neural e a plataforma são mostrados na figura 6. O erro residual entre a plataforma e a rede neural treinada é visto na figura 8.

Em princípio, considera-se o modelo ARMAX capaz somente de representar termos lineares. É verificado que a rede neural tem a capacidade de modelar características lineares que o ARMAX não foi capaz. Isto demonstra que as técnicas padrão de identificação “caixa cinza”, considerando que o modelo em estudo é, em teoria, linear, são capazes de dar significados aos parâmetros identificados, mas não de modelar, por completo, a planta em

análise. Sendo assim, para análises onde se desejar testes com resultados os mais próximos possíveis da dinâmica do dispositivo, as redes neurais são mais indicadas

Figura 5. Erro entre a plataforma de teste e o modelo ARMAX.

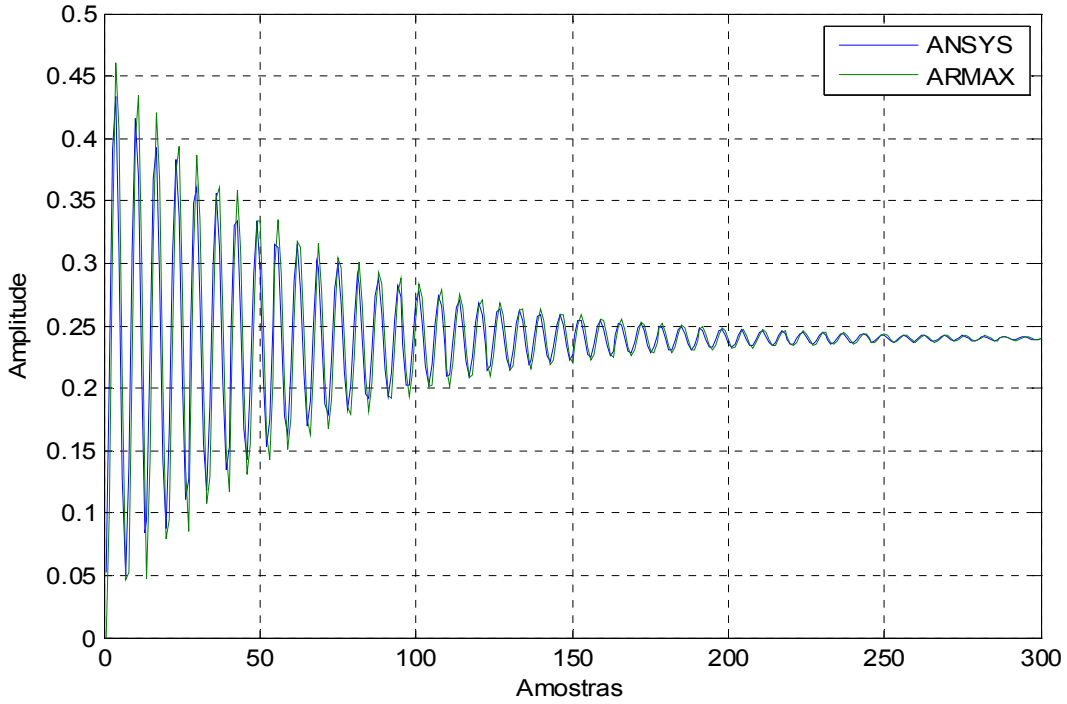


Figura 6. Erro entre a plataforma de teste e o modelo ARMAX.

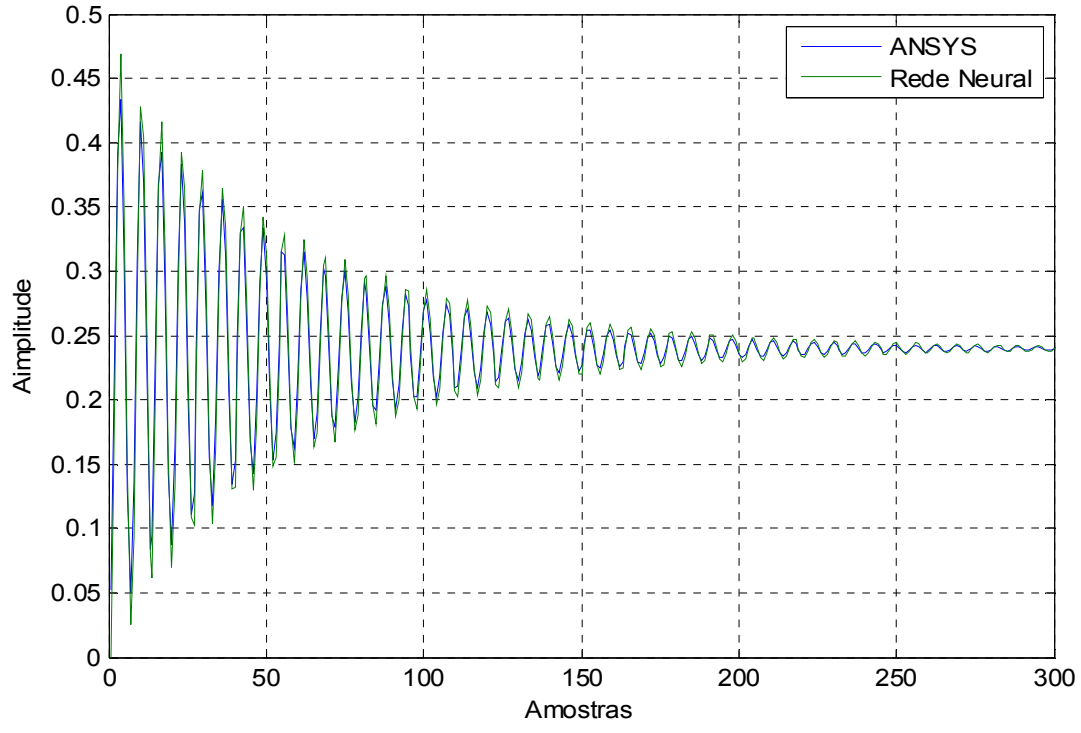


Figura 7. Erro entre a plataforma de teste e o modelo ARMAX.

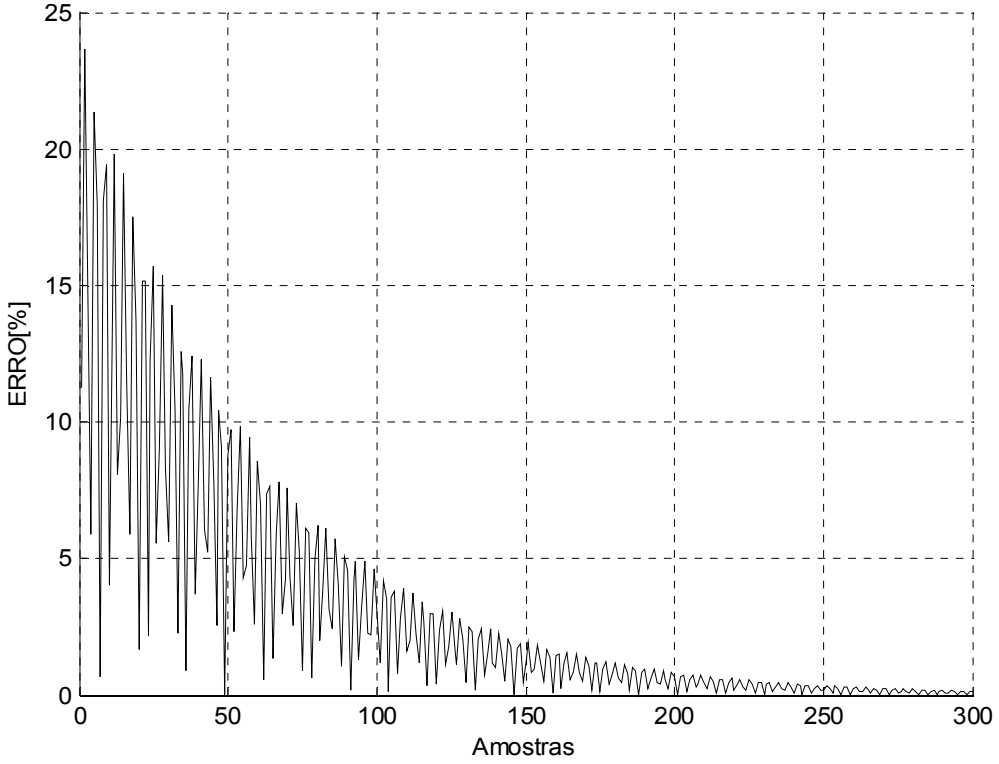
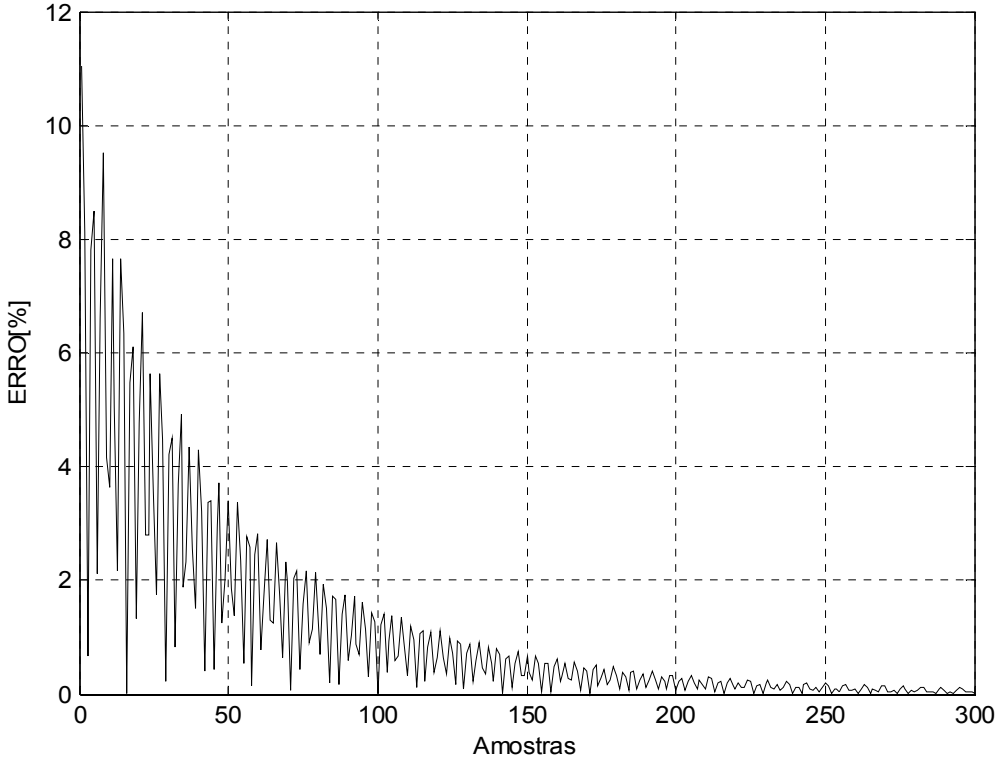


Figura 8. Erro entre a plataforma de teste e o modelo rede neural recorrente.



4. CONSIDERAÇÕES FINAIS

A ferramenta de identificação proposta oferece uma alta-precisão na simulação de dispositivos, sem sua modelagem matemática específica. A partir da planta obtida é possível estudar comportamentos não presentes nos modelos lineares, tornando ainda mais robusta à detecção de falhas.

Como exemplo de aplicação do modelo neural obtido, este pode ser útil na verificação do comportamento em freqüência do MEMS. Neste caso não seria necessário submeter o dispositivo a todas as frequências, mas a rede neural treinada extrairia, ainda, o comportamento não presente nos modelos lineares, aproximando-o às características reais de operação.

Agradecimentos

Os autores agradecem pelos recursos e suporte à PIBIC/UNIJUÍ.

REFERÊNCIAS

[1] Aguirre, L. A. "Introdução a Identificação de Sistemas – Técnicas Lineares e Não-lineares Aplicadas a Sistemas Reais", Ed. UFMG, 2 º Ed., 2004.

[2] Haykin S., "Redes Neurais – princípios e práticas", Boockman, Porto Alegre, 2001

[3] Ljung, L. "Systems Identification. Theory for the user", Prentice Hall, London, 1999.

[4] Tang W. C., "Electrostatic comb-drive of lateral poly-silicon resonators", Sensors and Actuators A, 21-23, pp.328-331, 1990.