

Síntese Evolutiva de Segmentos Sonoros - SESS

José Eduardo Fornari¹, Jônatas Manzolli^{1,2}, Adolfo Maia Jr.^{1,3}

¹Núcleo Interdisciplinar de Comunicação Sonora

²Departamento de Música – Instituto de Artes

³Departamento de Matemática Aplicada – IMECC

Universidade Estadual de Campinas

(UNICAMP)

Rua da Reitoria, 165 - Cidade Universitária "Zeferino Vaz"

CEP: 13 091 - 970 - Caixa Postal: 6166 – CAMPINAS - SP

{fornari,jonatas,adolfo}@nics.unicamp.br

Abstract. *The Evolutionary Sound Synthesis method (essynth) was initially described in [Manzolli 2001]. It uses a Population set of waveforms that evolve in time by the interaction of two processes: reproduction (done by genetic operators) and selection (done by the Hausdorff distance). The synthesized sound is the evolution path of best individuals in time. This paper presents the essynth's first model where the reproduction and selection processes are done directly on the waveforms . The sound results are available in audio, at: <http://www.nics.unicamp.br/~fornari/essynth1>.*

Resumo. *A síntese evolutiva de segmentos sonoros (essynth) foi inicialmente apresentada em [Manzolli 2001]. Esta utiliza uma população de segmentos sonoros que evolui ao longo do tempo pela interação de dois processos: reprodução (composto por operadores genéticos) e seleção (calculado pela distância de Hausdorff). O som sintetizado é o caminho temporal da evolução do melhor indivíduo da população. Este trabalho apresenta o primeiro modelo proposto da síntese evolutiva, onde os processos de reprodução e seleção são feitos diretamente sobre o segmento sonoro. Resultados sonoros podem ser escutados em: <http://www.nics.unicamp.br/~fornari/essynth1>.*

1. Introdução

A síntese evolutiva de segmentos sonoros (SESS) é um método que se baseia em princípios da Computação Evolutiva para geração de sons, conforme foi inicialmente proposta por [Manzolli, 2001]. Na SESS segmentos sonoros são tratados como indivíduos que pertencem exclusivamente a dois conjuntos. O primeiro é o conjunto chamado População que contêm segmentos sonoros muito simples, tais como sons monofônicos senoidais, que irão evoluir ao longo do tempo, representado por gerações da população. O segundo conjunto é chamado de Alvo e contêm segmentos mais complexos, tais como ruído-branco. A evolução dos indivíduos na população é guiada pelas características dos indivíduos do Alvo, assim este corresponde ao que seria a pressão condicionante do meio-ambiente na evolução de uma população biológica. Neste cenário a evolução da população de sons ocorre no sentido de, a cada geração, ocorrerem sons mais complexos, a partir de sons inicialmente muito simples. A geração sonora na síntese evolutiva é feita por dois processos independentes: Seleção, que

seleciona o melhor indivíduo de cada geração da população e Reprodução, que cria novos indivíduos na população. É importante ressaltar que o caminho traçado em cada processo de síntese evolutiva é único (nunca se repete) ainda que tenda a convergir para uma similaridade perceptual sonora. Dessa forma o método de síntese aqui apresentado tem o diferencial, em relação às outras formas de síntese sonora, de não ser determinística, mas evolutivo.

2. O Método da Síntese Evolutiva de Segmentos Sonoros

O método da SESS pode ser visto na figura abaixo (Figura 1).

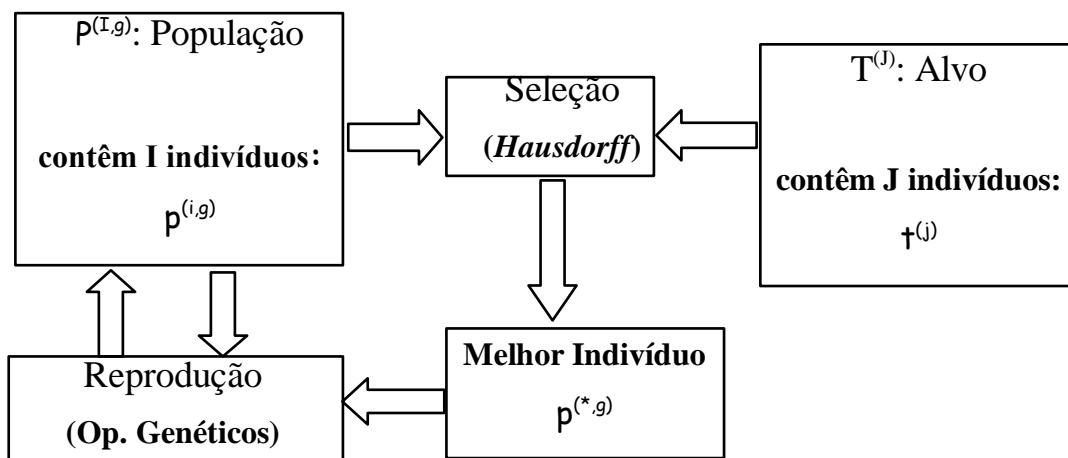


Figura 1. Diagrama da síntese evolutiva de segmentos sonoros.

Onde:

$p^{(i,g)}$: i-ésimo indivíduo da g-ésima geração. Todos os indivíduos são compostos por N pontos de áudio monofônico, amostrados à uma taxa de amostragem F_s Hz e resolução b bits. Valores típicos são: $F_s=44.1\text{KHz}$ e $b=16$ bits.

$p^{(*,g)}$: o melhor indivíduo da g-ésima geração

$P^{(I,g)}$: g-ésima geração do conjunto população de I indivíduos. O conjunto população antes do início do processo evolutivo é denotado por $P^{(I,0)}$.

$T^{(J)}$: o conjunto alvo contendo J indivíduos.

$t^{(j)}$: indivíduos do conjunto alvo.

No processo de **reprodução** agem dois operadores genéticos: **crossover** e **mutação**. O crossover permuta segmentos de cada indivíduo da população e o melhor indivíduo. A mutação insere modificações aleatórias em cada indivíduo promovendo assim a diversidade da população. O grau de ação dos operadores genéticos é dado pelas taxas alfa (taxa de crossover) e beta (taxa de mutação). Valores típicos utilizados são: alfa=0,5 e beta=0,1.

No processo de **seleção**, cada indivíduo da população é comparado com todos os indivíduos do conjunto alvo, através de uma função de adequação (*fitness function*). No método SESS, esta comparação é feita pela distância de *Hausdorff* que mede a distância entre conjuntos através da norma euclidiana (L2) tomada como sendo o Mínimo do conjunto que contém todas as distâncias tomadas entre cada elemento dos dois conjuntos. A aplicação de Hausdorff para síntese de som, no sentido de medir a similaridade entre conjuntos de amostras sonoras, é um desenvolvimento original deste trabalho. Outros pesquisadores como [Polansky, 1996], utilizaram processos de medida similares, chamados de métricas morfológicos. Todavia, as medidas de similaridade foram tomadas entre conjuntos representados por eventos como notas, padrões rítmicos e perfis melódicos.

3. O Experimento

Foram realizadas 8 simulações da síntese evolutiva de segmentos sonoros, onde usamos diferentes tipos de indivíduos tanto na população quanto no alvo. Como: senoides e ondas quadradas, em diferentes alturas (ou *pitch*). O resultado da síntese é a seqüência gerada pelo processo de evolução do melhor indivíduo de cada geração da população de segmentos sonoros. Nas simulações realizadas o som sintetizado é apresentado como uma seqüência resultante do entrelaçamento (*overlap-and-add*) de cada segmento sonoro que corresponde a cada melhor indivíduo.

Table 1. Resultados experimentais.

População	Alvo	N,M, Q, G	Resultados
squares	white-noise	N=16384, M=50, Q=3, G=20	s1
sines	white-noise	N=16384, M=50, Q=3, G=20	s2
sawtooths	white-noise	N=16384, M=50, Q=3, G=20	s3
sines	white-noise	N=4096, M=50, Q=3, G=40	s4
sines	white-noise	N=16384, M=100, Q=3, G=80	s5
sines	sines	N=4096, M=100, Q=10, G=40	s6
sines	squares	N=4096, M=100, Q=10, G=40	s7
squares	sines	N=4096, M=100, Q=10, G=40	s8

Onde:

$$\begin{array}{llll} \text{alfa} & = & 0.5 & (\text{taxa} \quad \quad \quad \text{de} \quad \quad \quad \text{crossover}) \\ \text{beta} & = & 0.1 & (\text{taxa} \quad \quad \quad \text{de} \quad \quad \quad \text{mutação}) \end{array}$$

$$Fs = 11025 \text{ (taxa de amostragem)}$$

$$b = 16 \text{ bits (resolução da amostra)}$$

N: comprimento de todos os indivíduos, onde: tempo = N / fs [s]

M: número de indivíduos na População, onde indivíduos estão espalhados em uma escala

logarítmica entre $f_{\min} = 80\text{Hz}$ e $f_{\max} = 4\text{KHz}$.
Q: número de indivíduos do Alvo.
G: número de gerações da síntese evolutiva.

Todos estes resultados sonoros estão disponíveis para *download* (em formato MP3) no link: <http://nics.unicamp.br/~fornari/essynth1>. A coluna “Resultados” descreve o nome do arquivo sonoro disponível na web.

4. Conclusão e comentários finais

Consideramos que a proposição da síntese evolutiva de segmentos sonoros (SESS) fornece bases para a criação de uma nova metodologia para desenvolver sínteses sonoras que não são determinísticas, mas evolutivas. O termo evolutivo é aqui utilizado no sentido de fazer uma analogia à Evolução Biológica. Neste sentido, evolução sonora descreve um processo no qual uma população inicial de sons se aproxima gradativamente de outra, denominada de Alvo. A síntese evolutiva de fato gera sons com grande riqueza de padrões a partir de sons simples (tais como populações descritas por formas de ondas senoidais). Entendemos que o controle da SESS é intuitivo, pois é dado apenas pela manipulação dos indivíduos do conjunto alvo, que determina o curso da evolução das gerações de melhores indivíduos. Ou seja, o processo de síntese é guiado por amostras sonoras e não por valores numéricos ou parâmetros. Nesta fase inicial o SESS, apesar de apresentar resultados sonoros interessantes, o método, do ponto de vista computacional, é ainda bastante simples. Muito pode ser desenvolvido uma vez que este é um campo praticamente novo e inexplorado de métodos de síntese sonora.

Como sugestões para trabalhos futuros, temos, entre outras, as que se seguem: Inclusão do conceito de genes e cromossomos dados por parâmetros que descrevam o comportamento perceptual do segmento sonoro, como grandezas psicoacústicas. Inclusão do conceito de gênero nos indivíduos, o que leva à criação de um novo processo de reprodução (sexuada). Variabilidade do número de indivíduos na população, que abre possibilidades para conceitos de extinção ou explosão demográfica. Inclusão de um período de maturação do indivíduo (participando do processo de seleção, mas não do de reprodução), ou seja, infância.

6. Agradecimentos

Gostaríamos de expressar nosso apreço à FAPESP, pelo apoio financeiro ao desenvolvimento deste trabalho, através de bolsa de Pos-Doc no Brasil, processo: 04/00499-6, o que culminou em um pedido de patentes protocolado no INPI em 23 de Março de 2005, sob o número: PI0500958-8 (Patente Requerida).

7. Referências

Fornari, J., Manzolli, J., Maia Jr., A., Damiani F., (2001) “Waveform Synthesis Using Evolutionary Computation”. Proceedings of the V Brazilian Symposium on Computer Music, Fotaleza.

Fornari, J., Manzolli, J., Maia Jr., A., Damiani F., (2001) “The Evolutionary Sound Synthesis Method”. SCI conference. Orlando, USA.

- Manzolli, J., Moroni A., Von Zuben F., Gudwin R., (1999) “An Evolutionary Approach Applied to Algorithmic Composition”, Proceedings of the VI Brazilian Symposium on Computer Music, Rio de Janeiro, Brazil, p. 201-210.
- Manzolli, J., Fornari, J., Maia Jr., A., Damiani F., (2001) “The Evolutionary Sound Synthesis Method”. Short-paper. ACM multimedia, ISBN:1-58113-394-4. USA.
- Moroni, A., Manzolli, J., Von Zuben, F., Gudwin, R., (2000) “Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition”, Leonardo Music Journal, San Francisco, USA, MIT Press, Vol. 10.
- Polansky, L., 1996., Morphological Metrics, Journal of New Music Research, 25, pp. 289-368.