

A Síntese Evolutiva Guiada pela Espacialização Sonora

José Fornari, Adolfo Maia Jr, Jônatas Manzolli
Núcleo Interdisciplinar de Comunicação Sonora (NICS)
e-mail: [fornari, adolfo, jonatas]@nics.unicamp.br
<http://www.nics.unicamp.br>

Sumário:

O método da síntese evolutiva de segmentos sonoros (SESS), inicialmente descrito em [1], usa como função de adequação a medida da distância euclidiana entre segmentos de áudio digital (*waveforms*) para guiar a evolução da população sonora. Este trabalho apresenta uma nova variação do método original onde consideramos como função de adequação a medida do posicionamento espacial do segmento sonoro. Os indivíduos que compõem a população são segmentos de áudio em dois canais (*stereo*) que através de uma função de diferença de tempo inter-aural (ITD) simula as suas localizações espaciais, como fontes sonoras virtuais dispostas num plano horizontal equidistante do usuário ouvinte.

Palavras-Chave: síntese sonora, computação evolutiva, psicoacústica, espacialização sonora.

1. Introdução

Desde o surgimento dos primeiros processos elétricos e eletrônicos com objetivos musicais, vem-se desenvolvendo uma grande quantidade e variedade de métodos de síntese sonora. Estes métodos podem ser organizados em três categorias: métodos lineares, tais como a síntese aditiva [2], métodos não-lineares, como a síntese FM [3] e métodos de edição, como é o caso da síntese *wavetable* [4]. Todas estas categorias de síntese sonora apresentam algo em comum: são métodos invariantes, pois apresentam um único tipo ou padrão fixo de saída (o som sintetizado) para uma condição fixa dos parâmetros de controle do processo de síntese. A síntese evolutiva aqui reportada tem o potencial de produzir variantes sonoras utilizando um mecanismo híbrido de operações determinísticas e não-determinísticas. O som sintetizado evolui ao longo do tempo no sentido de se adaptar a determinadas características ou regras, mesmo que os parâmetros de controle da síntese permaneçam inalterados. A Síntese Evolutiva de Segmentos Sonoros (SESS) é um método baseado na Computação Evolutiva [5], que por sua vez, inspira-se na teoria Darwiniana da evolução das espécies biológicas, valendo-se de analogias computacionais para os processos de reprodução e seleção natural. Existem diversos outros métodos aplicados à música que também são baseados na computação evolutiva. Em *GenJam* [6], um algoritmo genético simula improvisos de Jazz. Um método evolutivo de geração automática de processos de síntese sonora é apresentado em [7]. Um processo de geração evolutiva de padrões rítmicos é dado em [8]. A pesquisa aqui reportada iniciou-se com o desenvolvimento do *VoxPopuli* [9], um software de composição musical interativa que utiliza algoritmos genéticos e funções de adequação para a criação de seqüências musicais. Do ponto de vista de especificidade, verificamos que a SESS é um método pioneiro pela utilização original da computação evolutiva diretamente em síntese sonora i.e. utiliza algoritmos genéticos e funções de adequação não para a manipulação parâmetros de controle de um método determinístico de síntese, mas para a síntese sonora em si, agindo intrinsecamente no segmento sonoro.

2. O método da síntese evolutiva

Na SESS os **indivíduos** são amostras discretas (digitais) de segmentos sonoros com uma dada taxa de amostragem (*amostras/segundos*) e resolução (número de bits). O conjunto de todos os indivíduos/amostras é denominado de **população**. E neste conjunto o processo de evolução será aplicado. O caminho da evolução da população é condicionado através de uma medida de distância dada por uma **função de adequação**, *fitness*, que mede a distância entre parâmetros sonoras dos indivíduos da população com os de outro conjunto de indivíduos, o conjunto alvo. A evolução da população ocorre em estágios, chamados de **geração**. A evolução da população é feita por dois processos: a **reprodução** e a **seleção**. Em cada geração a reprodução gera novos indivíduos e a seleção escolhe o melhor indivíduo da população, ou seja, o mais adaptado aos critérios dados pelo conjunto alvo. No processo de reprodução agem dois operadores genéticos: cruzamento (**crossover**) e **mutação**. O *crossover* troca características sonoras dos indivíduos em reprodução (os progenitores). A mutação insere modificações aleatórias nessas características, aumentando assim a diversidade da população. Chamamos de **genótipo** do indivíduo o conjunto de parâmetros sonoros que o compõem, ou seja, no caso da SESS, são as grandezas acústicas e psicoacústicas. O processo de evolução atua sobre os genótipos dos indivíduos. Na reprodução, o genótipo é modificado pelo *crossover* e pela mutação. Na seleção, o método busca o indivíduo mais adequado aos critérios sonoros do sistema. O grau de adequação de cada indivíduo é medido pela distância entre o seu genótipo e um conjunto de genótipos dos indivíduos do conjunto **alvo**, que condicionam a evolução da síntese evolutiva. A cada geração da população o processo de seleção busca pelo **melhor indivíduo** da população, ou seja, o segmento sonoro com menor distância em relação ao alvo. O resultado sonoro deste método de síntese é um novo segmento sonoro gerado pelo encadeamento temporal dos melhores indivíduos de cada geração da população. Considerando um conjunto alvo inalterado ao longo das gerações, verifica-se que o som sintetizado tende a convergir para sons perceptualmente similares entre si, porém sem nunca se tornarem idênticos. A figura a seguir descreve sucintamente a SESS.

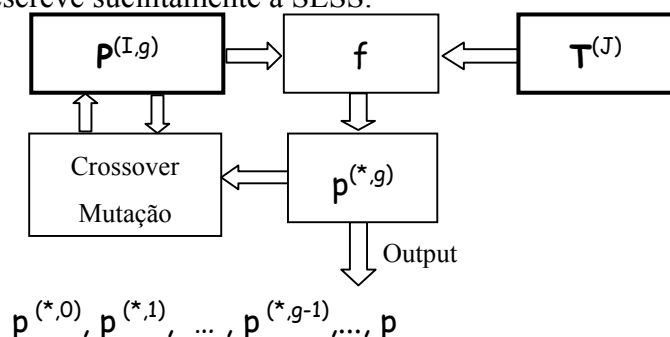


Figura 1: Diagrama básico da SESS

onde:

$P^{(i,g)}$: conjunto população com i indivíduos, na g -ésima geração. $p^{(i,g)}$: i -ésimo indivíduo da população, na g -ésima geração. $p^{(*,g)}$: melhor indivíduo da g -ésima geração da população. $T^{(J)}$: conjunto alvo com J indivíduos. $t^{(j)}$: o j -ésimo indivíduo do conjunto alvo. f : função de adequação.

3. Espacialização sonora aplicada ao método da SESS

Diferente das outras grandezas psicoacústicas tais como *loudness* (percepção da intensidade sonora) e *pitch* (percepção do harmônico fundamental), a percepção da localização espacial de fontes sonora é vinculada à audição binaural (audição simultânea através dos dois ouvidos). Desta forma, adotamos para o nosso modelo a premissa de que o sistema auditivo utiliza a diferença de tempo entre os dois ouvidos para estabelecer a localização de uma fonte sonora. Este modelo

psicoacústico é denominado de ITD (*interaural time differences*) e, conforme será descrito a seguir, pode ser satisfatoriamente simulado em áudio estereofônico (2 canais) através da imposição de um atraso (diferença de fase) entre os dois canais contendo a mesma informação sonora.

3.1 A espacialização dos indivíduos

A diferença de fase entre canais com mesmo som simula a informação binaural dada pela diferença de tempo de chegada do som nos dois ouvidos de um ouvinte. A função ITD simula apenas o posicionamento horizontal e equidistante, também chamado de ângulo de azimute. O posicionamento de elevação e/ou profundidade de objetos sonoros é dado por funções mais sofisticadas tais como as funções HRTF (*head related transfer function*), que serão alvo de estudos futuros.

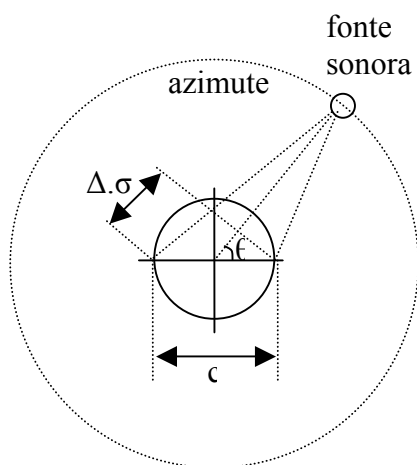


Figura 2. Espacialização Sonora usando a função ITD

3.2 Os processos de reprodução e seleção.

A cada geração, o processo de reprodução cria uma nova população de indivíduos que são descendentes do cruzamento entre cada indivíduo da geração anterior da população com o melhor indivíduo da geração atual. Cada indivíduo da nova geração é criado pelos operadores genéticos: *crossover* e mutação, que compõem o processo de reprodução. O operador *crossover* substitui seções do segmento sonoro de cada indivíduo com a seção similar do melhor indivíduo. A mutação insere mudanças aleatórias no segmento sonoro. A taxa de operação genética é dada por α (taxa de *crossover*) e β (taxa de mutação).

Em cada g-ésima geração, o processo de seleção encontra o melhor indivíduo dentro da população através da medida da distância entre o segmento sonoro de cada indivíduo com o conjunto alvo. A métrica da medida da distância é determinada por uma função de adequação (*fitness*).

4. Resultados experimentais

Neste trabalho usamos áudio digital, amostrado em $f_s = 44100\text{Hz}$. E com 16 bits de resolução. A taxa de crossover manteve-se fixa em $\alpha = 0.5$, e a taxa de mutação esteve fixa em $\beta = 0.1$ e tamanho de indivíduo dado por $N = 2f_s$, implicando assim em segmentos sonoros de 2 segundos de duração. Utilizamos uma população com senoides normalizadas, de mesma intensidade, porém com frequência variando entre $f_{\min}=80\text{Hz}$ a $f_{\max}=4000\text{Hz}$. Os indivíduos do

conjunto alvo são segmentos de ruído-branco normalizado. As variáveis utilizadas são **G**: número de gerações, **M**: número de indivíduos da população, **Q**: número de indivíduos no conjunto alvo. O resultado sonoro desses experimentos pode ser ouvido através do website dado abaixo: <http://www.nics.unicamp.br/~fornari/spatial/>

5. Conclusões e Comentários

Através da audição dos resultados experimentais (melhor percebido utilizando-se fones de ouvido), verifica-se que a função ITD é capaz de simular com sucesso o posicionamento espacial horizontal e equidistante dos indivíduos. A síntese evolutiva orientada pelo posicionamento espacial abre um novo campo de desenvolvimento para a síntese evolutiva. Novos modelos de SESS já estão sendo desenvolvidos utilizando a linguagem de programação para processos em tempo-real Pd (pure data) [14] e [15]. Através destes será possível ampliar o método em uma série de novas inclusões, bem como controlar e escutar os resultados sonoros em tempo-real. Entre outras, o posicionamento espacial pode ser utilizado em processos simulando a reprodução da SESS orientada pela proximidade territorial de indivíduos.

Estamos desenvolvendo um estudo paralelo no sentido de incorporar métodos de espacialização vinculados a SESS como um modelo alternativo para a projeção sonora de obras eletroacústicas. Neste caso, é importante ressaltar que a aplicação do método será no sentido de incorporar no processo de evolução da síntese evolutiva o seu componente de localização espacial. Desta forma, buscamos um novo modelo de projeção onde o espaço passa a ser incorporado na gênese do processo de síntese. O nosso interesse é vincular no próprio material sonoro a sua localização espacial e, por conseguinte, fazer do processo de evolução temporal o desdobramento desta potencialidade implícita.

6. Agradecimentos

Alguns resultados sonoros de simulações deste método podem ser escutados através do link: <http://www.nics.unicamp.br/~fornari>. Este trabalho é parte do projeto de PosDoc da FAPESP, processo número: 04/00499-6. O método da síntese evolutiva é extensivamente explicado em [10], e serviu de inspiração para dois pedidos de patente nacionais [12] e [13].

7. Definições Matemáticas e Modelagem

7.1 Definição de População

Originalmente, cada indivíduo $p^{(i,g)}$ é um segmento de áudio digital de um canal (monofônico) de N pontos, com resolução b e taxa de amostragem f_s .

$$p^{(i,g,[1,N])} = (x_1, x_2, \dots, x_n, \dots, x_N) \quad (1)$$

Para descrever o posicionamento espacial através da função ITD, o indivíduo deve ser descrito em dois canais com a mesma informação sonora, porém deslocados em fase por 2.d pontos, conforme dado abaixo:

$$p^{(i,g,[2,N])} = \begin{cases} \dots, x_{n+dt-1}, x_{n+dt}, x_{n+dt+1}, \dots \\ \dots, x_{n-dt-1}, x_{n-dt}, x_{n-dt+1}, \dots \end{cases} \quad (2)$$

7.2 Ângulo de Azimute

Para calcular o ângulo de azimute usa-se a função abaixo, dada por [11] :

$$\theta = \sin^{-1} \frac{(\Delta \cdot \sigma) \cdot V}{c} \quad (3)$$

onde: θ : ângulo de azimute, $\Delta = \frac{1}{fs}$: período (intervalo de tempo entre amostras). σ : número de pontos de atraso, ou defasagem. V : velocidade do som, dada por 384 m/s. c : distância media entre ouvidos da cabeça humana, adotada em 0,25 m.

De (2) e (3), os indivíduos com localização espacial são descritos por $P^{(i,g,[2,N],\theta)}$, onde:

$$\theta = \sin^{-1} \frac{3072 \cdot dt}{fs} \quad (4)$$

7.3 Função de adequação (Fitness)

Em teoria, existe uma infinita quantidade de métricas que podem ser utilizadas para esta finalidade. Estudamos algumas dessas métricas em [10] e decidimos utilizar a distância Euclidiana por apresentar os resultados mais satisfatórios. Considerando os segmentos sonoros dos indivíduos compostos por N elementos (pontos), temos que a distância Euclidiana entre dois indivíduos v e w é:

$$d_2(v, w) = \left[\sum_{i=1, \dots, N} (w_i - v_i)^2 \right]^{\frac{1}{2}} \quad (5)$$

Considerando $p^{(i,g)}$ o i-ésimo indivíduo da população $\mathbf{P}^{(l,g)}$, em sua g-ésima geração, e $t^{(j)}$, o j-ésimo indivíduo do conjunto alvo $\mathbf{T}^{(j)}$, o melhor indivíduo, ou seja, aquele com menor distância ao alvo, é dado por:

$$p^{(*,g)} = \min \{d_2(p_i, t_j)\} \quad (6)$$

Assim, a função de adequação é definida por:

$$f(P^{(I,g)}, T^{(J)}) = p^{(*,g)} \quad (7)$$

A equação (7) define a função de adequação para medida de distância de segmentos sonoros. Similarmente, a função de adequação para indivíduos com espacialização sonora é definida por:

$$f(P^{(I,g,\theta)}, T^{(J,\theta)}) = p^{(*,g)} \quad (8)$$

Onde a distância Euclidiana entre dois vetores v e w, com o ângulo de ITD como definido em (4), respectivamente para ξ e ψ é:

$$d_2(v^\xi, w^\psi) = \left[(\xi - \psi)^2 \right]^{\frac{1}{2}} \quad (9)$$

8. Referências Bibliográficas

- [1] Fornari, José, Jônatas Manzolli, Adolfo Maia, Furio Damiani. “The Evolutionary Sound Synthesis Method”. *Short-paper do ACM multimedia*, E.U.A. 2001.
- [2] Kleczkowski, P., “Group additive synthesis”. *COMP. MUSIC J.* Vol. 13, no. 1, pp. 12-20. 1989.
- [3] J. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Journal of the Audio Engineering Society*, vol. 21, pp. 526-534, 1973.
- [4] Horner, Andrew; Beauchamp, James; Haken, Lippold. “Methods for multiple wavetable synthesis of musical instrument tones”. *J AUDIO ENG SOC.* Vol. 41, no. 5, pp. 336-356. 1993.
- [5] Fogel, D. B., “Evolutionary Computation: Toward a New Philosophy of Machine Intelligence”, IEEE Press, 46 – 47, 1995.
- [6] Biles, J. A., “Gen Jam: A Genetic Algorithm for Generating Jazz Solos”, *Proceedings of the 1994 International Computer Music Conference*, (ICMC’94), 131—137, 1994.
- [7] R Garcia. "Growing Sound Synthesizers using Evolutionary Methods". *Proceedings of ALMMA 2002 Workshop on Artificial Models*, 2001
- [8] N Tokui, H Iba. "Music composition with interactive evolutionary computation.". *Proceedings of the third International Conference GA2000*, 2000.
- [9] Moroni, A., Manzolli, J., Von Zuben, F., Gudwin, R., “Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition”, *Leonardo Music Journal*, San Francisco, USA, MIT Press, Vol. 10, 2000.
- [10] Fornari, José Eduardo. “Síntese Evolutiva de Segmentos Sonoros”. Dissertação de Doutorado. DSIF/FEEC/UNICAMP. 2003
- [11] Murray, J. C., Erwin, H. R., and S. Wermter, (2004) "Robotic sound source localization using interaural time difference and cross-correlation", presented at KI-2004, September 2004.
- [12] Fornari, José, Jônatas Manzolli, Adolfo Maia. Métodos e Dispositivos Evolutivos para a análise, Processamento e Síntese de sinais digitais unis e multidimensionais, Pedido de Patente. Protocolado no INPI em 23 de Março de 2005, Protocolo: PI0500958-8.
- [13] Fornari, José, Jônatas Manzolli, Método Extrator de Curvas Psicoacústicas de Intensidade Sonora e Frequência Fundamental, Pedido de Patente, Protocolado no INPI em 15 de Dezembro de 2005. Protocolo: 01850064017.
- [14] M Puckette. "Pure Data: another integrated computer music environment". *Proceedings, Second Intercollege Computer Music Concerts*, 1996.
- [15] Site oficial do Pure Data: <http://www.puredata.org> .