

Construindo Trilhas Sonoras Dinâmicas Em Jogos Utilizando Sistemas Fuzzy

Marcelo Cardoso Silva Felipe M. G. França Giordano Ribeiro Eulalio Cabral*

Universidade Federal do Rio de Janeiro, PESC/COPPE, Brasil

*Universidade Federal de Pernambuco, Centro de Informática, Brasil

Resumo

O uso de áudio em jogos digitais (trilhas sonoras e efeitos) tradicionalmente segue um paradigma em que faixas pré-gravadas e estáticas são tocadas a partir de certos eventos que acontecem no jogo. Esse tipo de abordagem pode levar a uma trilha sonora que se torna muito repetitiva e enfadonha, o que pode diminuir o interesse do usuário no jogo e prejudicar sua experiência. Uma das possíveis alternativas para tratar essa questão é gerar o conteúdo musical em tempo real, utilizando padrões de comportamento do jogador como fonte de dados a fim de conseguir uma trilha sonora dinâmica e variada. Desta forma, este artigo propõe aplicar a teoria dos sistemas *fuzzy* na geração de sequências de acordes em tempo real, como uma maneira de trazer dinamismo à trilha sonora. A proposta utiliza variação de tons para adequar a trilha sonora a diferentes estados emocionais do jogador.

Palavras-chave: desenvolvimento de jogos, geração dinâmica de áudio, computação musical, sistemas *fuzzy*

Contato dos autores:

{marcelocardoso, felipe}@cos.ufrj.br
*giordanorec@gmail.com

1. Introdução

Em jogos para computador, tradicionais, é possível perceber um grande enfoque na qualidade da parte gráfica em comparação com o uso de áudio. Em jogos, um paradigma comum corresponde a ter trilhas sonoras estáticas e efeitos sonoros que são reproduzidos em resposta a certos eventos relacionados ao jogador (e.g., disparo de um tiro). O áudio (em particular, a trilha sonora) torna-se um elemento de distração em vez de ter uma função relevante no jogo. Outro problema relaciona-se com a repetição em demasia de trilhas sonoras, que pode tornar a experiência do jogo enfadonha. Existem trabalhos, como Berndt *et al.* [2012], que apresentam uma coleção de técnicas que tentam driblar ou reduzir a repetição musical inerente nesse paradigma, mas esses tipos de prática não resolvem o problema por completo. Embora esse tipo de solução possa vir a oferecer ambientes sonoros mais sofisticados, eles ainda se baseiam no paradigma comum sobre uso de áudio em jogos.

Uma solução para contornar os problemas mencionados é utilizar geração dinâmica de áudio em

jogos, como uma maneira de trazer dinamismo e variabilidade à trilha sonora.

Em termos de geração dinâmica de áudio, Wooller *et al.* [2005] dividem os possíveis tratamentos em algoritmos transformadores (*transformational*) e geradores (*generative*).

Os algoritmos transformadores alteram parâmetros ou propriedades de uma passagem sonora enquanto o conteúdo em si permanece o mesmo. Por exemplo, uma determinada música poderia sofrer alteração de tom ou tempo, sem alteração do conteúdo da música. Brown e Kerr [2009] discutem uma série de técnicas baseadas em algoritmos transformadores que se baseiam nos seguintes elementos: padrões (modelos harmônicos), abstrações (para isolar estrutura musical), recombinações (criar variações ao escolher dentre padrões ou trilhas distintas), transformações (transposição, expansão, contração, inversão, *arpeggiation*, entre outros) e probabilidade (seleção aleatória e aplicação de tabelas de Markov). Brown e Kerr [2009] aplicam algumas das técnicas descritas anteriormente como uma forma de se alterar a expressividade e o “estado de espírito” do ambiente sonoro. Mais especificamente, eles alteram os seguintes elementos músicas para atingir esse objetivo: faixa de notas (transposição, modulação), afinamento da densidade rítmica, articulação ou duração de notas, *loudness*, tempo e timbre.

Neste artigo, é proposta uma alternativa para se gerar trilhas sonoras em jogos aplicando técnicas *fuzzy*. Um dos objetivos da proposta é eliciar diferentes estados emocionais do jogador.

Este artigo está organizado como se segue. A Seção 2 apresenta algumas considerações sobre emoções em música e jogos. A Seção 3 descreve trabalhos relacionados. A Seção 4 apresenta a nossa proposta e a Seção 5 apresenta discussões e conclusões.

2. Emoção em Música e Jogos

Reconhecimento de emoções em música ainda é um tema em aberto. Yang e Chen [2012] conduziram uma extensa pesquisa sobre métodos encontrados na literatura para se reconhecer emoções em música. Uma das questões principais está relacionada com a conceitualização de emoção, que é bastante subjetiva.

De acordo com Paiva [2012], “em termos de tonalidade, os acordes maiores estão associados a

emoções positivas (alegria, contentamento) enquanto que acordes menores têm a conotação contrária [...] Um indivíduo pode estar eufórico e querer continuar nesse estado ou acalmar-se. [...] em teoria é possível escolher música de acordo com o estado de espírito que se deseja induzir.”

Liljedahl *et al.* [2007] cunharam o termo “*scary shadow syndrome*” para se referir a situações de onde a imaginação (de algo real ou não) poderia ser o ingrediente chave para se eliciar emoções nos jogadores (principalmente medo e apreensão). As situações típicas onde isso se aplica envolvem elementos como suspense e surpresa. Um exemplo icônico de filme que aplica esse tipo de técnica é *Psicose* de Alfred Hitchcock. Um exemplo de jogo que se utiliza dessa técnica é *The Audio Flashlight* [Valente *et al.* 2008; Valente *et al.* 2009], que utiliza uma série de trechos de música pré-gravados para criar um “radar sonoro”. O *The Audio Flashlight* é um jogo não-visual para telefones celulares, onde o jogador utiliza o radar sonoro para encontrar um tesouro escondido no cenário (uma sala escura). Quando o jogador se aproxima do objeto, a música do radar torna-se mais “dramática” (aumento do tempo e melodias mais complexas). Quando se afasta, a música se torna mais “calma” (diminuição do tempo e melodias menos complexas).

3. Trabalhos Relacionados

Os algoritmos generativos criam a música dinamicamente a partir de elementos básicos (como fragmentos de melodias), de acordo com um conjunto de regras definidas pelo programador. Collins [2009] apresenta alguns exemplos de jogos que se utilizaram desse tipo de técnica, como *Ballblazer* (1984) e *Spore* (2008). No caso de *Spore*, as amostras de sons foram criadas e inseridas no jogo como parte do seu conteúdo. Quando o jogo está em execução, regras são utilizadas para se definir melodias e andamentos, criando a trilha sonora em tempo real.

Robertson *et al.* [1998] descrevem um gerador dinâmico de músicas voltado para um ambiente virtual educacional utilizado para se treinar habilidades de escrita em crianças. Esse gerador dinâmico utiliza suspense e surpresa como uma maneira de alterar estados emocionais (como medo) dos usuários do ambiente virtual. Segundo Lombard e Ditton [1997], um dos objetivos de se usar essa técnica é reforçar o grau de presença nos ambientes virtuais, como uma forma de aumentar a imersão e tornar a experiência de aprendizado mais motivadora. A imersão é uma propriedade central relacionada com a experiência que um jogo pode oferecer aos usuários.

Livingstone e Brown [2005] propuseram um ambiente musical dinâmico (aplicado a jogos) onde a trilha sonora se ajusta ao “estado de espírito” do momento atual do jogo, em tempo real. Esse ajuste ocorre através da alteração de propriedades da música, utilizando regras que mapeiam emoções para

propriedades musicais. O mapeamento de emoções para propriedades musicais resultou em um mapa bidimensional denominado *2D Dimensional Emotion Space* (2DES). Esse mapa é dividido em 8 octantes e contém emoções relacionadas à raiva, tristeza, depressão, empolgação, felicidade e estados “sonhadores” (*dreamy*). As propriedades musicais principais relacionadas às emoções foram modo, tempo, *loudness*, complexidade harmônica e articulação. Outras propriedades (consideradas como secundárias) foram ritmo, timbre e tonalidade.

Adam *et al.* [2014] propuseram um sistema para gerar música dinamicamente. Esse sistema recebe como entrada cinco tipos de parâmetros que são modelados para se relacionar com propriedades musicais: estresse, energia, semente, “número de padrões” e “repetição de padrões”. Os dois primeiros parâmetros (estresse e energia) são usados para se caracterizar o “estado de espírito” (*mood*). Estresse é mapeado para níveis de harmonia e dissonância musical. Energia é mapeada para ritmo, tempo e andamento. A semente tem a ver com aspectos relacionados à variabilidade (aleatória) da música. O “número de padrões” tem a ver com o número de sequências de 4 tempos que devem ser geradas. O último parâmetro determina quantas vezes uma sequência qualquer (de 4 tempos) deve ser repetida.

Em relação à aplicação de técnicas *fuzzy* para música, Cavalcanti *et al.* [2012] apresentam uma tabela com os expoentes musicais de cada nota (dó-ré-mi-fá-sol-lá-si), gerados a partir da determinação das frequências das notas musicais utilizando a equação de Mersenne. Uma função de pertinência define o grau de intensidade i com que um elemento pertence a um conjunto *fuzzy*. Este grau de intensidade i é um número real contido no intervalo $0 \leq i \leq 1$. As funções de pertinência triangulares dependem de três parâmetros escalares (a , b e c) e são comumente usadas em sistemas *fuzzy* devido a sua simplicidade. O cálculo do expoente musical leva em consideração o valor da divisão da frequência em Hz da nota pelo número 11 (número de ouro da música) e os números primos 2, 3 e 5 que, usados como funções de pertinência triangulares, se tornam capazes de representar as notas musicais. A partir dos expoentes musicais, são propostas análises de acordes duplos utilizando os operadores *fuzzy* OU e E. Os resultados obtidos com o uso desses operadores sugerem outra forma de execução do acorde, isto é, através do uso de apenas uma nota. Por exemplo, o acorde formado pelas notas dó₄ (central) e mi₄, cujas frequências são, respectivamente, $24 \cdot 11$ e $30 \cdot 11$, gera o expoente musical $120 \cdot 11$ que corresponde ao mi₆, pois, em uma operação *fuzzy* OU, se considera o máximo entre os expoentes musicais dó (310) e mi (111), donde se obtém o expoente 311 que é o máx. entre 3 e 1, o máx. entre 1 e 1 e o máx. entre 1 e 0. Calculando a frequência em Hz, temos $23 \cdot 31 \cdot 51 = 120 \cdot 11 = 1320$ Hz (mi₆). Com isso, observa-se que a execução do acorde dó₄mi₄ pode ser gerada pela nota mi₆. Seguindo

o mesmo exemplo, a demonstração do uso do operador E chega ao resultado $6 * 11 = 66$ Hz (dó₂), uma vez que, neste caso, se considera o valor mínimo entre os expoentes musicais dó (310) e mi (111). O mesmo raciocínio aplicado aos acordes duplos pode ser aplicado aos acordes triplos.

4. Proposta

A proposta se baseia na ideia apresentada por Plans e Morelli [2012], que descrevem o uso de normas advindas de experimentos em geração de música procedural com o objetivo de delinear novas maneiras de se produzir música e áudio dinamicamente a partir das métricas de jogo do usuário. Essas métricas são frustração, desafio e diversão com o objetivo de gerar música adaptativa considerando o EDPCG em Yannakakis e Togelius [2011]. O EDPCG (*Experience-Driven Procedural Content Generation*) é um *framework* para geração dinâmica e procedural de conteúdo, baseando-se em modelos computacionais orientados à experiência do usuário.

As adaptações de Plans e Morelli [2012] consistem na avaliação das métricas em tempo real para permitir a implantação de um mecanismo musical capaz de reagir às mudanças de humor do jogador durante o jogo. Para isso, foi criada uma única métrica-alvo para expressar o humor do jogador ($target\ excitement = fun - 0.5 \times challenge - 0.5 \times frustration$) e estimulá-lo se: 1) ele não está tendo um jogo frustrante; 2) ele está se divertindo; e 3) não é desafiado por muito tempo. A partir daí é proposto um gerador de música tonal simples baseado em algoritmos genéticos padrões que produzem as sequências harmônicas e a melodia (período). A sequência harmônica se baseia em acordes criados a partir da nota dó na escala maior, ou seja, dada a nota fundamental (dó), é gerado o acorde dó-lá-sol. Em tempo de execução, cada vez que uma nova sequência de acordes se faz necessária, é feita uma avaliação histórica das sequências que já foram tocadas, levando em consideração um índice de novidade histórica. O mesmo se aplica para construção de períodos. Merece destaque que o construtor de períodos gera um conjunto de frases (período), e, cada frase é uma sequência individual de notas MIDI.

Uma hipótese inspirada em Plans e Morelli [2012] é proposta para se trabalhar com as métricas frustração, desafio e diversão, a fim de ajustar a trilha sonora dinamicamente e, por conseguinte, alterar o estado emocional do jogador. Essa hipótese baseia-se nos seguintes itens:

1. Gerar trilha sonora em diferentes tons. A ideia é criar um mapeamento das notas musicais fundamentais (dó, ré, mi, fá, sol, lá e si) para determinados estilos musicais. Esse mapa será usado para alterar a trilha sonora para versões mais tristes, alegres ou clássicas da música. Uma versão “alegre” poderia ser uma música na escala de dó maior, por exemplo.

2. Fuzzificar as métricas originalmente propostas por Plans e Morelli [2012], sob forma de variáveis linguísticas que representam, cada uma delas, três conjuntos *fuzzy*: Baixo, Médio e Alto. Essa fuzzificação pretende avaliar o grau de entusiasmo do jogador em determinado instante através de regras.

O grau de entusiasmo é utilizado para selecionar (e gerar) o estilo adequado para a trilha sonora. Por exemplo, pode haver uma regra que selecione um estilo “alegre” para a trilha sonora:

```
R60: SE frustração é Baixo E desafio é Baixo E diversão é Alto ENTÃO estilo é Alegre.
```

Para cada regra, estará associado ao seu consequente, um estilo que será intensificado de acordo com a oitava do piano. Por exemplo, se o estilo “alegre”, representado pela regra R60, estiver associado ao dó central (dó₄), considerando que o número que identifica a regra representa o valor MIDI da tecla do piano, as regras com valores maiores representarão as oitavas acima do dó central e as com valores menores, as oitavas abaixo.

Dada a combinação dos resultados obtidos pelas regras disparadas, a defuzzificação da variável linguística estilo gera um valor de frequência que será usado para determinar a nota musical fundamental do acorde que irá compor a próxima sequência de acordes a ser gerada.

Junto com a ideia inicial apresentada anteriormente, formulam-se as seguintes perguntas de pesquisa:

1. Quais os estados emocionais do jogador que poderiam ser influenciados pela manipulação dinâmica do áudio ?
2. Quais estilos musicais são adequados para serem usados em jogos, a fim de eliciar os estados emocionais do jogador ?
3. É possível fuzzificar os estados emocionais do jogador a fim de se obter uma métrica para o grau de entusiasmo baseada em regras *fuzzy* ?
4. Quais as técnicas computacionais que poderiam ser empregadas na geração dinâmica de áudio ?
5. É coerente associar os consequentes das regras *fuzzy* a uma variável linguística que represente as frequências em Hertz das notas musicais ?

5. Discussão e Conclusões

A manipulação dinâmica do áudio pode trazer inovação e benefícios à experiência que um jogo fornece. No entanto, primeiro é necessário entender que estados emocionais poderiam ser elicitados em um jogo, o que levou à pergunta de pesquisa (1). A pergunta de pesquisa (2) é relacionada à (1) porque é

preciso entender que estilos musicais são mais adequados para se elicitar as emoções e como mapeá-los em escalas e tonalidades. Além disso, a resposta a essa pergunta servirá para criar uma base de conhecimento *fuzzy* onde os antecedentes das regras devem gerar consequentes adequados ao contexto do jogo. A pergunta (3) pretende auxiliar a descoberta da viabilidade em se aplicar sistemas *fuzzy* na avaliação do grau de entusiasmo. Em particular, pretende-se comparar os resultados desta pesquisa com aqueles obtidos por Plans e Morelli [2012] para verificar se existe algum tipo de melhoria. Para responder à pergunta (4), é preciso fazer um levantamento das técnicas existentes para geração dinâmica de áudio em jogos e escolher aquelas que sejam mais adequadas para implementação da hipótese proposta neste artigo. Finalmente, a pergunta (5) ajudará a mapear o estilo musical através de notas que servirão de base para geração de acordes.

Benefícios de pesquisa em geração dinâmica de áudio e áudio em geral, para jogos:

- Oportunidade de criar trilhas sonoras variadas e dinâmicas, que possam estimular e alimentar o interesse do jogador;
- Oportunidade de inovar em relação ao *gameplay*, ao tentar explorar o áudio como um veículo para elicitar estados emocionais nos jogadores;
- Oportunidade de melhorar um componente crucial da experiência do jogador, a imersão. O áudio pode ser manipulado dinamicamente para criação de efeitos como os descritos pela “*scary shadow syndrome*”, Liljedahl *et al.* [2007].

Referências

- ADAM, T., HAUNGS, M. E KHOSMOOD, F., 2014. Procedurally Generated, Adaptive Music for Rapid Game Development. Em *Proceedings of the 9th International Conference on the Foundations of Digital Games*. FDG 2014. Ft. Lauderdale. Available at: <http://fdg2014.org/>.
- BERNDT, A., DACHSELT, R. E GROH, R., 2012. A Survey of Variation Techniques for Repetitive Games Music. Em *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound*. AM '12. New York, NY, USA: ACM, pp. 61–67. Available at: <http://doi.acm.org/10.1145/2371456.2371466> [Accessed July 27, 2014].
- BROWN, A.R. E KERR, T., 2009. Adaptive music techniques. Em A. C. Sorensen, ed. *Improvise: The Australasian Computer Music Conference 2009*. Improvise: The Australasian Computer Music Conference. Brisbane: Australasian Computer Music Association, pp. 26–31. Available at: <http://acma.asn.au/> [Accessed July 27, 2014].
- CAVALCANTI, J.H.F., DE MELO, H., DA ROCHA SOUTO, C. E CAVALCANTI, M.T., 2012. *Lógica Fuzzy Aplicada às Engenharias*, João Pessoa, PB.
- COLLINS, K., 2009. An Introduction to Procedural Music in Video Games. *Contemporary Music Review*, 28[1], pp.5–15.
- LILJEDAHL, M., PAPWORTH, N. E LINDBERG, S., 2007. Beowulf: an audio mostly game. *Proceedings of the international conference on Advances in computer entertainment technology*, pp.200–203.
- LIVINGSTONE, S.R. E BROWN, A.R., 2005. Dynamic Response: Real-time Adaptation for Music Emotion. Em *Proceedings of the Second Australasian Conference on Interactive Entertainment*. IE '05. Sydney, Australia, Australia: Creativity & Cognition Studios Press, pp. 105–111. Available at: <http://dl.acm.org/citation.cfm?id=1109180.1109196> [Accessed July 27, 2014].
- LOMBARD, M. E DITTON, T., 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, 3[2], pp.0–0.
- PAIVA, R.P., 2012. A música e as emoções. Disponível em: <http://www.eventpoint.com.pt/articles/a-musica-e-as-emocoes> [Acessado Julho 24, 2014].
- PLANS, D. E MORELLI, D., 2012. Experience-Driven Procedural Music Generation for Games. *IEEE Transactions on Computational Intelligence and AI in Games*, 4[3], pp.192–198.
- ROBERTSON, J., DE QUINCEY, A., STAPLEFORD, T. E WIGGINS, G., 1998. Real-time music generation for a virtual environment. Em *Proceedings of ECAI-98 Workshop on AI/Alife and Entertainment*. Citeseer.
- VALENTE, L., DE SOUZA, C.S. E FEIJÓ, B., 2008. An exploratory study on non-visual mobile phone interfaces for games. Em *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*. IHC '08. Porto Alegre, Brazil, Brazil: Sociedade Brasileira de Computação, pp. 31–39.
- VALENTE, L., SOUZA, C.S. DE AND FEIJÓ, B., 2009. Turn off the graphics: designing non-visual interfaces for mobile phone games. *Journal of the Brazilian Computer Society*, 15[1].
- WOOLLER, R., BROWN, A.R., MIRANDA, E., DIEDERICH, J. E BERRY, R., 2005. A framework for comparison of process in algorithmic music systems. Em B. David & E. Ernest, eds. *Creative Industries Faculty*. Generative Arts Practice. Sydney, Australia: Creativity and Cognition Studios, pp. 109–124. Available at: <http://eprints.qut.edu.au/6544/> [Accessed July 27, 2014].
- YANG, Y.-H. E CHEN, H.H., 2012. Machine Recognition of Music Emotion: A Review. *ACM Trans. Intell. Syst. Technol.*, 3[3], pp.40:1–40:30.
- YANNAKAKIS, G.N. E TOGELIUS, J., 2011. Experience-Driven Procedural Content Generation. *IEEE Transactions on Affective Computing*, 2[3], pp.147–161.