

Uma Abordagem Baseada em Agentes Para Um Sistema de Classificação de Timbres

Eduardo P. Teixeira¹, Eder M. N. Gonçalves¹, Diana F. Adamatti¹

¹Programa de Pós-Graduação em Computação (PPGComp)
Centro de Ciências Computacionais (C3)
Universidade Federal do Rio Grande (FURG)
Rio Grande – RS – Brasil

{eduardoteixeira, edergoncalves, diananaadamatti}@furg.br

Abstract. *This paper proposes a agent-based approach to timbre recognition, focusing on the parallelization of the classification model. For this, we assign a method of recognition of timbres to different agents, where each agent is a specialized entity in a particular timbre, characteristic of a specific instrument, seeking a distributed solution for solving the timbre recognition problem.*

Resumo. *Este trabalho propõe uma abordagem baseada em agentes para o reconhecimento de timbres, com enfoque na autonomia dos agentes ao modelo de classificação de timbres. Para isto, atribui-se um método de reconhecimento de timbres a diferentes agentes, onde cada agente é uma entidade especialista em um determinado timbre, característico de um instrumento específico, visando uma solução ao problema de reconhecimento de timbres de forma distribuída.*

1. Introdução

Existem quatro principais dimensões em sons: altura, intensidade, duração e timbre. A quarta dimensão, o timbre, é a mais vaga e complexa das dimensões [Eronen et al. 2001, Casey et al. 2008]. Até mesmo para a percepção humana, o reconhecimento de timbres é uma tarefa difícil, bem como a definição desta característica. O *American National Standards Institute* define timbre de uma maneira puramente excludente:

“...atributo de sensação em que um ouvinte pode julgar que dois sons com a mesma intensidade e altura são diferentes” [Klingbeil 2009].

São muitos os trabalhos cujo tema é o reconhecimento e identificação de timbres, dentre os quais podem ser citados [Helmholtz and Ellis 2009, Strong 1963, Luce and Clark Jr 1967, Benade 2012, Nordqvist 2004, Klapuri 2004, Kitahara 2007]. Esta característica abstrata é de grande interesse no campo de MIR (*Music Information Retrieval*) [Casey et al. 2008].

A tecnologia de Sistemas Multiagentes é uma nova forma promissora para a performance musical interativa, em outros termos, execução colaborativa, onde agentes instrumentistas se unem na produção de uma performance musical, como visto em [Sampaio et al. 2005, Sampaio et al. 2008].

Em trabalhos recentes, essa tecnologia foi adaptada para resolver problemas específicos em um escopo musical limitado, como detecção de pulso, simulação de instrumentos ou acompanhamento automático. Sendo uma área bastante consolidada, adequada

para solucionar problemas que exijam distribuição, seja de natureza lógica ou geográfica, e em que a complexidade do problema seja minimizada por esta abordagem. Sendo assim, Sistemas Multiagentes são úteis em várias subáreas da Computação Musical.

Neste contexto, esse artigo apresenta uma abordagem baseada em agentes para a solução do problema de reconhecimento de timbres, visando um sistema escalável e paralelizável, onde cada agente atua como um especialista em um determinado instrumento, sendo o responsável por sua classificação.

2. Fundamentos Musicais

O conceito mais básico por trás de qualquer área de estudo musical é a definição de som. Ele é produzido quando um objeto (a fonte sonora) vibra e faz o ar ao seu redor se mover [Rumsey and McCormick 2012]. Este efeito pode ser representado como uma esfera que pulsa regularmente, com centro na fonte sonora, e seu tamanho oscila levemente entre maior e menor que o normal. Assim que pulsa, o som irá comprimir e rarefazer o ar ao seu redor, resultando em uma série de compressões e rarefações viajando para longe da esfera, similar a uma versão tridimensional de uma pedra que cai sobre um lago. Se a pressão varia de acordo com um padrão repetitivo, diz-se que o som tem uma forma de onda periódica. Se não houver nenhum padrão discernível, é chamado de ruído. Entre esses dois extremos existe um vasto domínio de sons quase periódicos e quase ruidosos [Roads 1996].

As ondas sonoras são ondas de compressão causadas por vibrações, mas a música de uma sinfonia varia consideravelmente do choro de um bebê ou do sussurro de um confidente.

Todas as ondas sonoras podem ser caracterizadas por sua *altura*, pela sua *intensidade*, pela sua *duração*, e pela sua qualidade sonora ou *timbre* [Lapp 2003]:

- A altura é uma característica do som que faz referência a nossa percepção de agudos e graves. Fisicamente, sons agudos possuem maiores frequências, e sons graves menores. O ser humano é capaz de ouvir em um intervalo entre 20Hz e 20.000Hz. Baleias e golfinhos escutam frequências ainda maiores [Lapp 2003].
- A intensidade do som está relacionada com a amplitude da onda de som. A maioria das pessoas tem algum reconhecimento da escala de decibéis (dB). Eles podem ser capazes de dizer que 0dB é o limiar de audição e de que o som na pista ao lado de um jato acelerando é de cerca de 140dB.
- A duração diz respeito ao tempo da onda sonora, o seu período, e o tempo que o som leva até cessar. É uma importante característica ao se estudar aspectos rítmicos dos sons.
- O timbre é a mais vaga e complexa das quatro dimensões dos sons [Eronen et al. 2001]. Considerável energia e esforços foram aplicados para promover o entendimento do timbre, uma das características mais abstratas da música.

2.1. Representações dos Sons

Além da frequência fundamental de um som, que representa o seu “tom”, ou a sua altura, podem haver muitas frequências presentes em uma forma de onda. Uma representação

no domínio da frequência, ou espectrograma, mostra as frequências principais contidas em um som. Os componentes individuais de frequência do espectro podem ser referidos como harmônicas e parciais. Frequências harmônicas são múltiplos inteiros da frequência fundamental [Roads 1996], e podem ser facilmente identificadas em um analisador de espectro como visto na Figura 1.

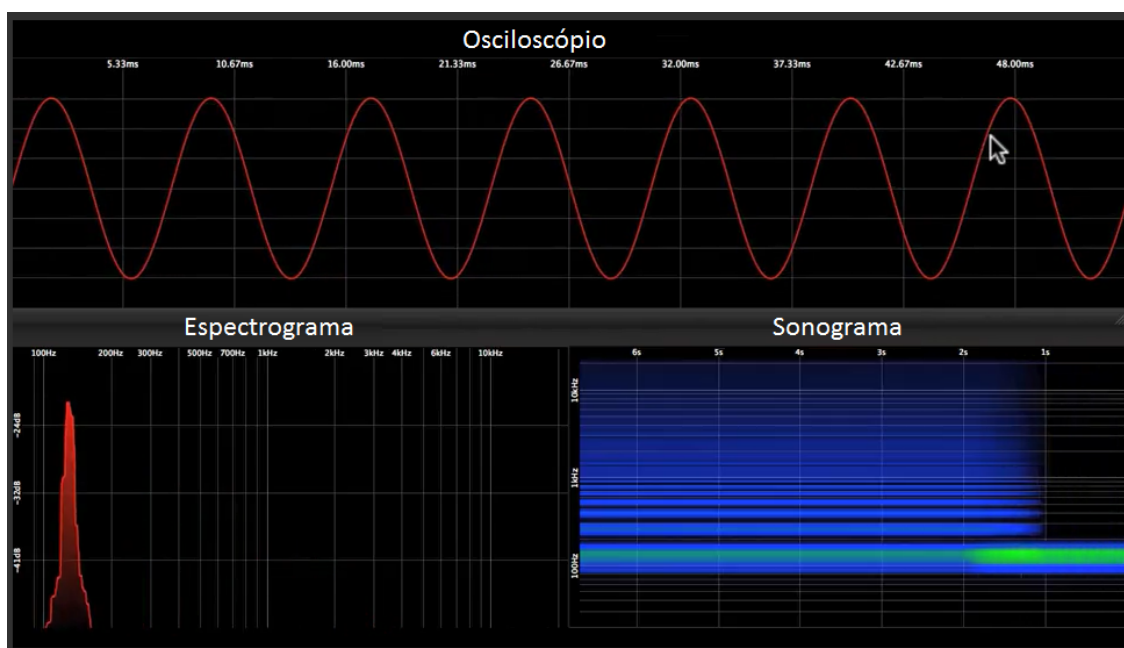


Figure 1. Diferentes representações de uma entrada sonora. Acima, um sinal criado por um osciloscópio. Abaixo, a esquerda, seu respectivo espectro no domínio da frequência. Abaixo, a direita, a visualização de um sonograma.

2.1.1. Osciloscópio

Um osciloscópio é usado para indicar a forma de onda de um som. Ele aceita sinais sonoros em forma elétrica e exibe suas análises em tela. O osciloscópio exibe um ponto em movimento que varre horizontalmente um número de velocidades fixas da esquerda para a direita e cuja deflexão vertical é controlada pela tensão do sinal de som (positivo para cima, negativo para baixo). Deste modo, representa graficamente a forma de onda do som, uma vez que varia com o tempo. Muitos osciloscópios tem duas entradas e podem traçar duas formas de onda ao mesmo tempo. Isto pode ser particularmente útil para a comparação das fases relativas de dois sinais [Rumsey and McCormick 2012].

2.1.2. Analisador de Espectro

O analisador de espectro funciona de diferentes formas, dependendo do método de análise de espectro. Um analisador em tempo real mostra um espectro de linha em constante atualização, e mostra os componentes do sinal de entrada na escala horizontal, juntamente com as suas amplitudes na escala vertical de frequência [Rumsey and McCormick 2012].

Em um analisador de espectro é possível observar as harmônicas (ou parciais), que são características da fonte sonora utilizadas no reconhecimento de timbre.

No domínio da frequência é possível recuperar informação musical com os mais diversos objetivos.

2.1.3. Sonograma

Em um sonograma é representado um espectrograma que varia com o tempo. Este tipo de representação é útil no reconhecimento de fala entre outras aplicações. Auxilia a visualizar as alterações nas frequências harmônicas em sons sendo executados ao longo do tempo. Ele representa três dimensões: a frequência (em Hertz, na vertical), o tempo (em segundos, na horizontal) e a intensidade (em decibéis, representado com diferentes colorações).

2.2. O Reconhecimento de Timbres

Como exemplo de trabalho estado da arte em MIR, que trabalha com o timbre como característica principal, pode-se citar [Devi et al. 2012], que trata o timbre como um conjunto de características, como a envoltória sonora (a forma como o som se inicia, se mantém e termina ao longo do tempo) para a amplitude e a frequência; tempo de ataque (início de cada nota musical); decaimento (em alguns instrumentos o som sofre um decaimento após o ataque até se estabilizar); sustentação (corresponde ao tempo de duração da nota musical) e intensidade [Devi et al. 2012]. Essas características podem ser observadas na Figura 2.

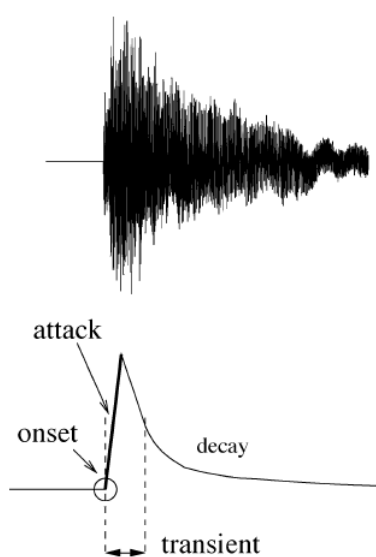


Figure 2. Formato do espectro de uma única nota sendo executada [Bello et al. 2005].

Historicamente, os primeiros estudos datam dos anos cinquenta, onde pode-se citar o trabalho de [Helmholtz and Ellis 1954], que evidenciou que as amplitudes relativas das parciais harmônicas de um som, muito mais que suas fases relativas, são determinantes primários do timbre.

Já [Strong 1963] interpretou o espectro de vários instrumentos de orquestra e demonstrou que o oboé, clarinete e fagote são identificados primeiramente na base no seu espectro de magnitude.

Com estes trabalhos, começam a surgir um maior número de estudos que buscam determinadas características no formato espectral dos sons para o reconhecimento de timbre, como o trabalho proposto em [Luce and Clark Jr 1967], que demonstrou que a família de metais (instrumentos de sopro como trombone, saxofone, trompete, entre outros), é caracterizada por um único *cutoff* na frequência, e que esta característica se correlaciona fortemente com a identificação desta família de instrumentos. Posteriormente, [Benade 2012] foi além, e mostrou que o corte na frequência é um dos principais determinantes do timbre nos instrumentos de sopro de um modo geral.

Outra característica que vem sendo fortemente correlacionada com o timbre é o centróide espectral, que em termos gerais pode ser definido como o “ponto de balanço” do espectro, se mostra diretamente ligado ao “brilho” do instrumento, uma dimensão primária e subjetiva do timbre, como verificado em [Grey 1977, Lichte 1941, von Bismarck 1974]. Nos trabalhos [Beauchamp 1982a, Beauchamp 1982b], o autor demonstra que o centróide varia em muitos instrumentos com a intensidade do som.

No estudo realizado em [Strong 1963], foi demonstrado que, no reconhecimento de muitos instrumentos, a identificação ocorre principalmente pela envoltória temporal, em parte porque suas envoltórias espectrais não são únicas, lembrando que a definição de timbre é excludente, sendo este responsável por diferenciar sons que possuem mesma altura e intensidade [Klingbeil 2009].

Neste contexto, é possível observar que o reconhecimento do timbre está fortemente ligado à forma do espectro sonoro, que usualmente é dividido em *onset*, *attack*, *transient* e *decay*. Este formato é associado ao caso ideal de uma única nota sendo executada, como ilustrado na Figura 2 por [Bello et al. 2005].

Outra característica que passa a ser analisada em relação à influência de alteração do timbre são os aspectos temporais do som, cuja importância foi reconhecida primeiramente por [Risset and Wessel 1982]. Em sequência a esta descoberta, pode-se citar o trabalho de [Handel 1995], que apresenta a unificação da utilização das características temporais e espectrais no reconhecimento de timbre.

Já [Schoenberg 1978] considerou o timbre como uma segunda dimensão do tom, e hoje sabe-se que o timbre pode ser considerado de alguma forma em uma característica multidimensional, como descrito primeiramente em [Grey 1977], e inclusive pode ser representado visualmente, como foi apresentado no trabalho proposto por [Soraghan 2014].

Um dos primeiros trabalhos completos com enfoque de reconhecimento de timbre, com a utilização de várias das características apresentadas, foi o proposto por [Martin and Kim 1998], que considera que existem duas categorias de características para reconhecimento de timbre: temporal e espectral, ambas com grande importância no reconhecimento de timbre. Neste trabalho algumas características são utilizadas, como:

- Altura: sinais produzem uma estrutura identificável em relação à altura no correlograma, num frame bidimensional, com defasagem no eixo horizontal e a frequência em relação à vertical, sulcos verticais indicam o período do sinal, e

por inversão, a altura.

- **Envoltória espectral:** uma vez que a altura tenha sido detectada, a altura de crista vertical do correlograma pode ser mensurada como uma função da frequência, para se obter uma estimativa da forma da envoltória espectral. O centroide espectral é simplesmente o centroide da envoltória espectral.
- **Intensidade:** a soma da energia na envoltória espectral aproxima a intensidade sonora instantânea do sinal. Acompanhar esta ao longo do tempo leva a medidas simples de modulação de amplitude, o que pode revelar o *Tremolo* e, por correlação com modulações de frequência, ressonâncias. Como sugerido em [Beauchamp 1982a, Beauchamp 1982b], a relação entre a intensidade e o centroide espectral pode ser uma importante correlação perceptual do timbre.
- **Ataque assíncrono:** ao rastrear a envoltória espectral ao longo do tempo, com estimativas de altura concorrentes, é possível medir as características de ataque de um tom harmônico musical de uma forma psicofisicamente adequada.
- **Enarmonia:** os desvios harmônicos no sinal vão ser refletidos como desvios da estrutura vertical no correlograma instantâneo.

Ao todo, o trabalho de [Martin and Kim 1998] utilizou trinta e uma características diferentes, a maioria sendo variações das citadas acima. Dentre as estratégias de reconhecimento adotadas, utilizou-se o método k-NN, hierárquico e não-hierárquico para identificação de timbres.

Como descrito por [Eronen et al. 2001], *Formants* são protuberâncias criadas por uma ou mais ressonâncias na fonte sonora. Elas representam a informação essencial para o reconhecimento de voz e fala, e também para reconhecimento de instrumentos musicais. Um recurso robusto para medir a informação de *Formants*, ou a envoltória espectral suavizada, são os coeficientes cepstrais. Dentre as técnicas utilizadas no trabalho de [Eronen et al. 2001], está *Mel-frequency Cepstral Coefficients* (MFCC) [Davis and Mermelstein 1980].

O método de MFCC se tornou uma das técnicas mais populares na extração de características em sistemas automáticos de reconhecimento de fala [Eronen et al. 2001], pois oferece uma descrição da forma espectral do som bastante eficiente, posteriormente sendo amplamente utilizada no reconhecimento de timbres. Um *cepstro*¹ é o resultado de aplicar a transformada inversa de Fourier (IFFT) ao logaritmo do espectro estimado de um sinal. O MFCC se baseia no procedimento de cepstro. Nele, as bandas de frequência estão posicionadas de forma logarítmica utilizando a escala de Mel². A transformada de Fourier é substituída por uma transformada de cosseno discreta (DCT)³. Ela tem uma propriedade de “compactação de energia” eficiente: a maior parte da informação do sinal tende a se concentrar em alguns componentes do DCT de baixa frequência. É por isso que, por padrão, apenas os 13 primeiros componentes são devolvidos. Por convenção, o coeficiente zero indica simplesmente a energia média do sinal [Lartillot et al. 2014].

¹A denominação “cepstro” vem da palavra “spectro” com as primeiras letras em ordem inversa, fazendo referência ao uso da transformada inversa de Fourier.

²A escala de Mel, nomeada por [Stevens et al. 1937], é uma escala de percepção de alturas julgada por ouvintes a ter espaçamentos iguais entre os sons. Esta escala de frequência se aproxima mais a resposta do sistema auditivo humano do que as faixas de frequências linearmente espaçadas.

³Uma transformada de cosseno discreta (DCT) é uma transformada de Fourier semelhante à transformada de Fourier discreta (DFT), mas utilizando apenas números reais.

3. Abordagem Baseada em Agentes para a Classificação de Timbres

Como cada instrumento possui determinadas características de timbre que são identificadas por diferentes descritores, é possível imaginar que uma solução distribuída e paralela seja adequada para reconhecimento polifônico, aumentando a eficiência na solução de problemas.

Segundo [Thomaz 2009], a tecnologia de agentes se torna particularmente adequada para aplicações musicais, devido à possibilidade de associar um agente computacional com o papel de um cantor ou instrumentista. Ele destaca algumas vantagens dessas associações, como mapear características como desempenho, percepção, adaptação e improvisação de um lado, e processos artificiais no outro. Além disso, é possível definir formas de inter-relação social entre os agentes, que traz esta tecnologia ainda mais perto de performance musical colaborativa.

Para a resolução do problema de reconhecimento de timbres, valendo-se de uma abordagem baseada em agentes, este trabalho utiliza um conjunto de agentes especialistas em determinados timbres, onde cada um é responsável pela classificação de um instrumento, como exemplifica a Figura 3.

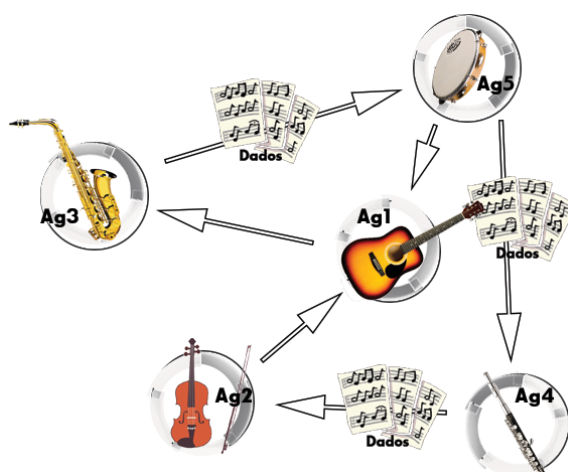


Figure 3. Representação simplificada do sistema proposto. São apresentados cinco agentes especialistas, cada um representado pelo instrumento que ele é encarregado de reconhecer. Dentro do ambiente, os agentes trocam informações entre eles, bem como dados musicais (conjunto de características extraídas) que precisam ser classificados.

Tratam-se de agentes cognitivos, que podem receber um conjunto de características do ambiente ou de outros agentes. Quando um novo sinal de áudio é previamente processado, e suas características são extraídas, esta nova entrada é enviada para o ambiente, onde um agente se encarrega de começar o processo de classificação. Caso ele reconheça as características como sendo ele o reconhecedor correto, com uma tolerância percentual empírica, este retorna ao ambiente notificando que o som foi devidamente classificado. Caso o agente especialista não atinja o percentual esperado, ele encaminha o conjunto de características para um novo agente. Este processo está representado na arquitetura apresentada na Figura 4.

Dentre as vantagens esperadas em adotar um sistema baseado em agentes para a resolução do problema de classificação de timbres pode-se citar:

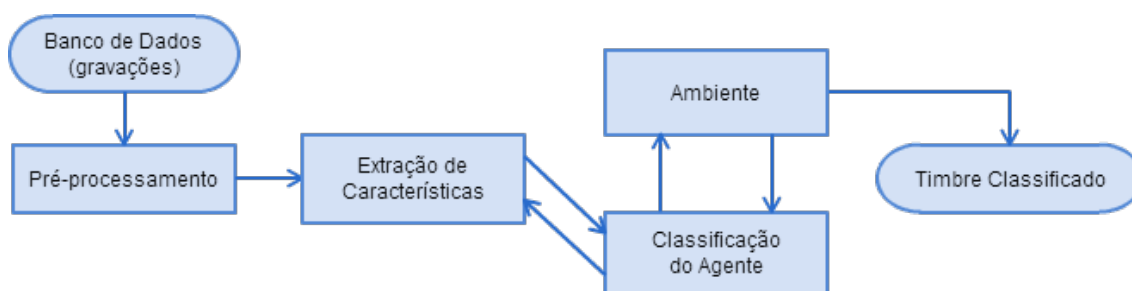


Figure 4. Arquitetura do sistema desenvolvido.

- Maior escalabilidade do sistema, pois é possível adicionar um novo agente, responsável por classificar um novo instrumento, sem realizar um novo treinamento em todo o sistema.
- Paralelização da classificação, pois quando existir uma tarefa de classificação de muitas entradas, a classificação delas poderá ser distribuída entre vários agentes.
- Aperfeiçoamento dos resultados, já que cada agente será específico para cada timbre é possível que eles sejam melhores especialistas do que uma única entidade classificadora, ou que, dependendo da implementação, troquem informação entre si para obter melhores resultados.

4. Desenvolvimento

O sistema baseado em agentes foi desenvolvido no NetLogo [Wilensky 1999], as operações de treino e classificação dos agentes foram desenvolvidas no MatLab [Guide 1998] com a utilização da MIRtoolbox [Lartillot and Toivainen 2007] para a leitura de áudio e a extração de características musicais. A integração entre estas duas ferramentas foi realizada através de uma extensão não oficial do NetLogo, denominada MatNet⁴.

O banco de dados utilizado consiste em uma coleção de arquivos com gravações de notas únicas de diversos instrumentos. Os arquivos de áudio utilizados foram obtidos através da OLPC (*One laptop per child - free sound samples*⁵).

A Figura 5 apresenta a interface gráfica do sistema desenvolvido, onde são representados os agentes e a interação entre eles.

Para o treinamento do agente, através do seu ID, são selecionados arquivos de áudio do banco de dados e aplicado o método de MFCC (*Mel-frequency cepstral coefficients*) apresentado em [Davis and Mermelstein 1980], que consiste em um conjunto de 13 coeficientes para cada arquivo. Desta forma é criada uma matriz de coeficientes para cada agente especialista de tamanho $N \times 13$, onde N representa o número de arquivos utilizados no treinamento. O método de classificação utilizado consiste em aplicar o MFCC para a entrada e comparar com a média da matriz $N \times 13$ do agente, onde as diferenças entre cada um dos 13 coeficientes são somadas e normalizadas. O funcionamento dos agentes, após serem treinados, consiste em aguardar uma atribuição do ambiente ou de outro agente. Quando ele é atribuído de uma classificação, ele realiza o método de

⁴<http://github.com/mbi2gs/netlogo-matlab-extension/wiki>

⁵http://wiki.laptop.org/go/Free_sound_samples

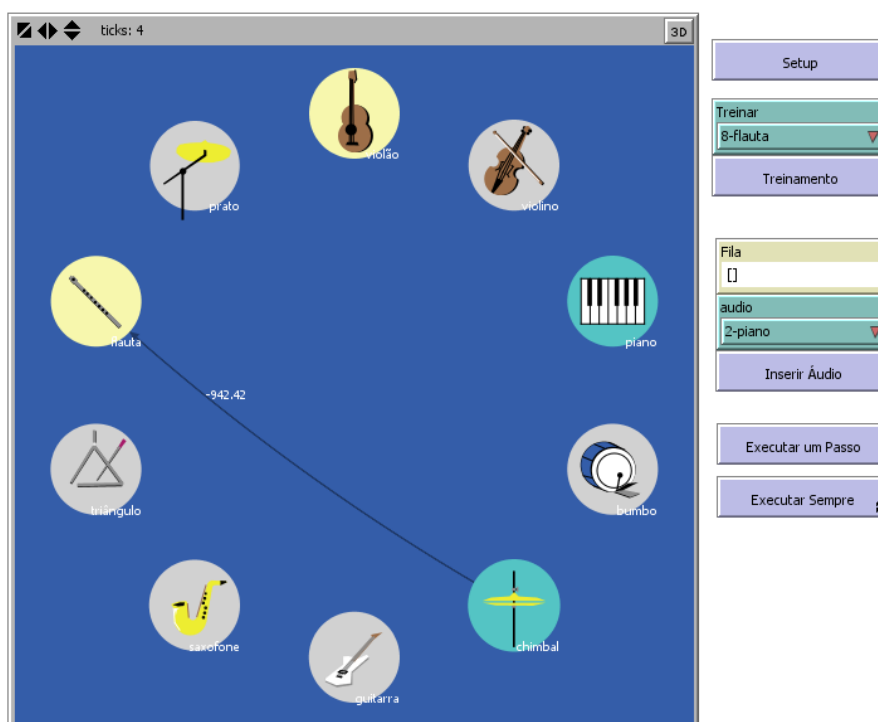


Figure 5. Dentro da visualização de mundo estão representados 10 agentes especialistas, onde o agente chimbau após não conseguir classificar sua entrada a encaminha para o agente flauta.

classificação da entrada: caso a resposta for maior que um limiar empírico, ele reporta ao ambiente que ele é o correto classificador; caso contrário encaminha a entrada para outro agente aleatoriamente.

5. Conclusão

Neste trabalho foi apresentado um sistema baseado em agentes para a solução do problema de classificação de timbres musicais, em que para cada instrumento seria associado um agente especialista capaz de reconhecer apenas um único instrumento. Em um ambiente multiagente, onde vários agentes cognitivos podem realizar trocas de informação, a classificação de timbres seria realizada de forma distribuída, paralela e escalável.

O método de classificação implementado se provou bastante rápido, conforme a sua simplicidade, mas foram observados erros de classificação em instrumentos da mesma família, como o violão e a guitarra, possivelmente pela grande semelhança dos coeficientes de MFCC destes arquivos. Para a solução deste problema, uma provável solução seria utilizar um conjunto de características maior, que explorassem outros comportamentos no domínio do tempo e da frequência. Outra limitação apresentada pelo sistema diz respeito ao método de encaminhamento dos agentes. Em alguns testes de execução, a mesma entrada chegou a passar quatro vezes pelo mesmo agente antes de alcançar o agente correto. Para solucionar este problema, poderia ser implementado um vetor associado a cada entrada, do tamanho do número de agentes no ambiente, servindo como um histórico por onde a entrada circulou, que deverá ser verificado antes de realizar uma atribuição.

References

- Beauchamp, J. W. (1982a). Data reduction and resynthesis of connected solo passages using frequency, amplitude, and “brightness” detection and the nonlinear synthesis technique. *The Journal of the Acoustical Society of America*, 71(S1):S101–S101.
- Beauchamp, J. W. (1982b). Synthesis by spectral amplitude and” brightness” matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society*, 30(6):396–406.
- Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., and Sandler, M. B. (2005). A tutorial on onset detection in music signals. *Speech and Audio Processing, IEEE Transactions on*, 13(5):1035–1047.
- Benade, A. H. (2012). *Fundamentals of Musical Acoustics: Second*. Courier Corporation.
- Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M., et al. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–696.
- Davis, S. B. and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(4):357–366.
- Devi, J. S., Srinivas, Y., and Krishna, N. M. (2012). A study: Analysis of music features for musical instrument recognition and music similarity search. *IJCSI*.
- Eronen, A. et al. (2001). Automatic musical instrument recognition. *Mémoire de DEA, Tampere University of Technology*, page 178.
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277.
- Guide, M. U. (1998). *The mathworks. Inc., Natick, MA*, 5:333.
- Handel, S. (1995). Timbre perception and auditory object identification. *Hearing*, pages 425–461.
- Helmholtz, H. L. and Ellis, A. J. (1954). *On the sensations of tone as a physiological basis for the theory of music (AJ Ellis, Trans.)*. New York: Dover.(Original work published in 1885).
- Helmholtz, H. L. and Ellis, A. J. (2009). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Cambridge University Press.
- Kitahara, T. (2007). Computational musical instrument recognition and its application to content-based music information retrieval. *Unpublished PhD Thesis, Kyoto University, Kyoto, Japan. Retrieved*, 10(31):07.
- Klapuri, A. (2004). *Signal processing methods for the automatic transcription of music*. Tampere University of Technology Finland.
- Klingbeil, M. K. (2009). *Spectral Analysis, Editing, and Resynthesis: Methods and Applications*. PhD thesis, Columbia University.
- Lapp, D. R. (2003). *The physics of music and musical instruments*. Wright Center for Innovative Science Education, Tufts University.

- Lartillot, O. and Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects*, pages 237–244.
- Lartillot, O., Toiviainen, P., and Eerola, T. (2014). *MIRtoolbox 1.6 User's Manual*.
- Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology*, 28(6):455.
- Luce, D. and Clark Jr, M. (1967). Physical correlates of brass-instrument tones. *The Journal of the Acoustical Society of America*, 42(6):1232–1243.
- Martin, K. D. and Kim, Y. E. (1998). 2pmu9. musical instrument identification: A pattern-recognition approach. In *Presented at the 136th meeting of the Acoustical Society of America*. Citeseer.
- Nordqvist, P. (2004). *Sound classification in hearing instruments*. PhD thesis, KTH-S3.
- Risset, J.-C. and Wessel, D. L. (1982). Exploration of timbre by analysis and synthesis. *The psychology of music*, 28.
- Roads, C. (1996). *The computer music tutorial*. MIT press.
- Rumsey, F. and McCormick, T. (2012). *Sound and recording: an introduction*. CRC Press.
- Sampaio, P., Tedesco, P., and Ramalho, G. (2005). Cinbalada: um laboratório multiagente de geração de ritmos de percussao. In *Proceedings of the X Brazilian Symposium on Computer Music*.
- Sampaio, P. A., Ramalho, G., and Tedesco, P. A. (2008). Cinbalada: Multiagent rhythm factory. *J. Braz. Comp. Soc*, 14(3):31–49.
- Schoenberg, A. (1978). *Theory of harmony*. Univ of California Press.
- Soraghan, S. (2014). Animating timbre-a user study. *The 2014 IEEE International Conference on Systems, Man, and Cybernetics*.
- Stevens, S. S., Volkman, J., and Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3):185–190.
- Strong, W. J. (1963). *Synthesis and recognition characteristics of wind instrument tones*. Massachusetts Institute of Technology.
- Thomaz, L. F. (2009). A framework for implementing musical multiagent systems. *6th Sound and Music Computing Conference*, pages 119–124.
- von Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acta Acustica united with Acustica*, 30(3):146–159.
- Wilensky, U. (1999). Netlogo.