ADCAIJ: Advances in Distributed Computing and Articial Intelligence Journal
Regular Issue, Vol. 7 N. 1 (2018), 29-40
eISSN: 2255-2863
DOI: http://dx.doi.org/10.14201/ADCAIJ2018712940

Ulises: An Agent-Based System For Timbre Classification

Eduardo P. Teixeira[a], Eder M. N. Goncalves[a], and Diana F. Adamatti[a]

[a]Graduation Program in Computer Science (PPGComp) – Computer Science Center (C3) – Federal University of Rio Grande (FURG) Rio Grande – RS – Brasil eduardoteixeira@furg.br, edergoncalves@furg.br, diananaadamatti@furg.br

KEYWORD

The Sound and Music Computing; Timbre Classification; Agent-based System

ABSTRACT

The Sound and Music Computing (SMC) field has grown over the years and every time there are more conferences and specialized researchers in this area. The sub-field of Music Information Retrieval (MIR), one of the main research fields on SMC has focused on getting information from sound data. The most critical issue with regard to the human perception of sound is: what are the qualities of musical instrument sounds to perform recognition of its sound sources. There are four main sound dimensions: pitch, loudness, duration and timbre. The fourth dimension, timbre, is the most vague and complex dimension, a complex and high-level multidimensional property. Recognition of timbres is an area of high interest within MIR, being present in several papers state of the art on SMC. About Multi-Agent Systems (MAS), the term autonomous refers to the fact that the agents have their own existence, regardless of the existence of other agents, and are able to take own decisions without outside interference. Agents technology is particularly suitable for musical applications because of the possibility of associating a computational agent with the role of a singer or instrumentalist as can be seen in works state of art in SMC area. In this context, 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.

1. Introduction

There are four main dimensions in sounds: pitch, loudness, duration and timbre. The fourth dimension, timbre, is the most vague and complex of dimensions (Eronen et al., 2001; Casey et al., 2008). Even for human perception, the recognition of timbres is a difficult task, as well as the definition of this characteristic. The *American National Standards Institute* defines timbre in a purely exclusionary way:

> "...an attribute of auditory sensation by which a listener can judge that two sounds having the same loudness and pitch are different" (Klingbeil, 2009).

There are many works whose theme is the recognition and identification of timbres, among which can be cited (Helmholtz and Ellis, 2009; Strong, 1963; Luce and Clark Jr, 1967; Benade, 2012; Nordqvist, 2004; Klapuri, 2004; Kitahara, 2007). This abstract feature is of great interest in the field of MIR (*Music Information Retrieval*) (Casey et al., 2008).

The technology of Multiagent Systems is a promising new form for interactive musical performance, in other words, collaborative execution, where instrumental agents works togheter in the production of a musical performance (Sampaio et al., 2005; Sampaio et al., 2008).

In recent work, this technology has been adapted to solve specific problems in a limited musical scope, such as pulse detection, instrument simulation or automatic musical accompaniment. Being a well-established area, adequate to solve problems that require distribution, whether of a logical or geographical nature, and where the complexity of the problem is minimized by this approach. Thus, Multiagent Systems are useful in several sub-areas of Sound and Music Computing.

In this context, this work presents an agent-based approach to the problem of timbre classification, aiming at a scalable and parallelizable system, where each agent acts as a specialist in a particular instrument and is responsible for its classification.

2. Musical Fundamentals

The most basic concept behind any area of music study is the definition of sound. It is produced when an object (the sound source) vibrates and the air around it moves (Rumsey and McCormick, 2012). This effect can be represented as a sphere that pulsates regularly, with center in the sound source, and its size oscillates slightly between greater and smaller than normal. As it pulses, the sound will compress and thin the air around it, resulting in a series of compressions and rarefaction traveling away from the sphere, similar to a three-dimensional version of a stone falling on a lake. If the pressure varies according to a repetitive pattern, the sound is said to have a periodic waveform. If there is no discernible pattern, it is called noise. Between these two extremes there is a vast domain of almost periodic and almost noisy sounds (Roads, 1996).

Sound waves are compression waves caused by vibrations, but the music of a symphony varies considerably from the crying of a baby or the whisper of a confidant.

All sound waves can be characterized by their *pitch*, by their *loudness*, by their *duration*, and by their sound quality or *timbre* (Lapp, 2003):

- Pitch is a characteristic of sound that refers to our perception of treble and bass. Physically, high sounds have higher frequencies, and lower bass sounds. The human being is able to listen in a range between 20Hz and 20,000Hz. Whales and dolphins listen to even higher frequencies (Lapp, 2003).
- The intensity of the sound is related to the amplitude of the sound wave. Most people have some recognition of the decibel scale (dB). They may be able to say that 0dB is the threshold of hearing and that the sound on the lane next to a jet accelerating is about 140dB.
- Duration refers to the time of the sound wave, its period, and the time the sound takes until it ceases. It is an important feature when studying rhythmic aspects of sounds.
- The timbre is the most vague and complex of the four dimensions of sounds (Eronen et al., 2001). Considerable energy and efforts were applied to promote the understanding of timbre, one of the most abstract features of music.

2.1. Representations of Sounds

In addition to the fundamental frequency of a sound, which represents its pitch, there may be many frequencies present in a waveform. A frequency domain representation, or spectrogram, shows the main frequencies contained in a sound. The individual frequency components of the spectrum can be referred to as harmonic and partial. Harmonic frequencies are integral multiples of the fundamental frequency (Roads, 1996), and can be easily identified in a spectrum analyzer as seen in Figure 1.

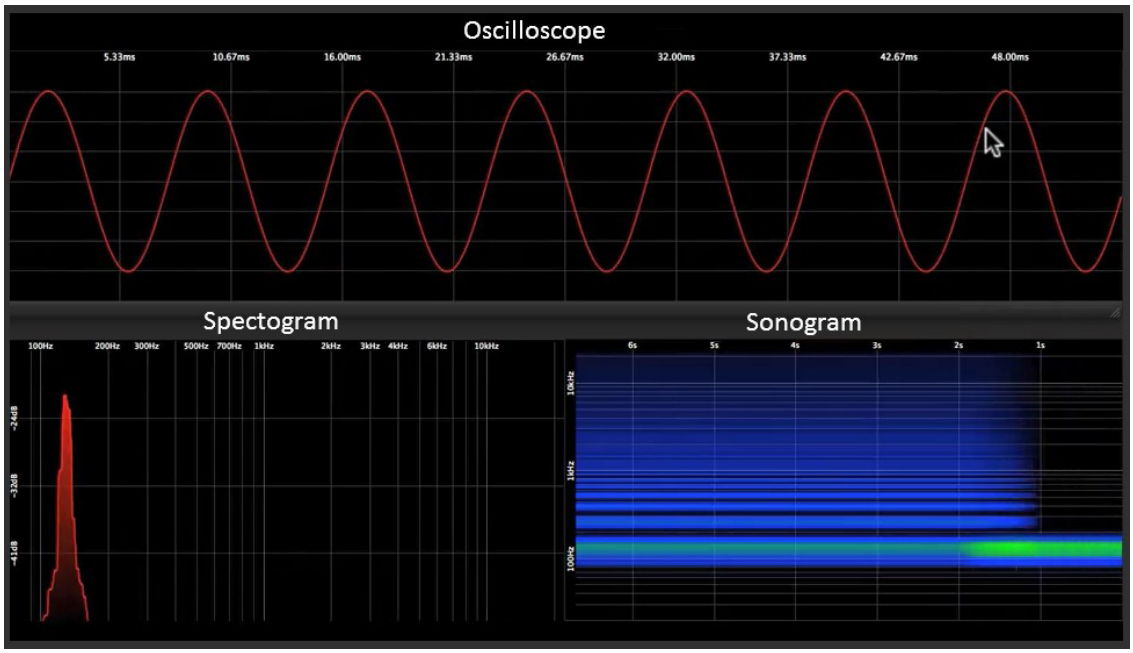

Figure 1: Different representations of a sound input. Above, a signal created by an oscilloscope. Below, the left, its respective spectrum in the frequency domain. Below, on the right, the visualization of a sonogram.

2.1.1. *Oscilloscope*

An oscilloscope is used to indicate the waveform of a sound. It accepts sound signals in electrical form and displays its analyzes on screen. The oscilloscope displays a moving point that sweeps horizontally a number of fixed speeds from left to right and whose vertical deflection is controlled by the sound signal voltage (positive up, negative down). In this way, it graphically represents the waveform of the sound as it varies over time. Many oscilloscopes have two inputs and can trace two waveforms at the same time. This may be particularly useful for comparing the relative phases of two signals (Rumsey and McCormick, 2012).

2.1.2. *Spectrum Analyzer*

The spectrum analyzer works in different ways, depending on the method of spectrum analysis. A real-time analyzer shows a constantly updated line spectrum, and shows the components of the input signal on the horizontal scale, along with their amplitudes in the vertical frequency scale (Rumsey and McCormick, 2012). In a spectrum analyzer it is possible to observe the harmonics (or partial), which are characteristics of the sound source used in the recognition of timbre.

In the field of frequency it is possible to retrieve musical information with the most diverse objectives.

2.1.3. *Sonogram*

In a sonogram is shown a spectrogram that varies with time. This type of representation is useful in speech recognition among other applications. Helps to visualize changes in harmonic frequencies in sounds being played over time. It represents three dimensions: frequency (in Hertz, vertical), time (in seconds, horizontal) and loudness (in decibels, represented with different colorations).

2.2. Timbre Recognition

As an example of state-of-the-art work in MIR, which works with timbre as its main feature, one can (Devi et al., 2012), which treats timbre as a set of characteristics, such as the sound envelope (and its starts, holds, and ends over time) for amplitude and frequency; attack time (beginning of each musical note); decay (in some instruments the sound decays after the attack until it stabilizes); sustain (corresponds to the duration of the musical note) and loudness (Devi et al., 2012). These characteristics can be observed in Figure 2.

Historically, the earliest studies date back to the fifties, where one can cite the work of (Helmholtz and Ellis, 1954), which showed that the relative amplitudes of the harmonic partial of a sound, much more than its relative phases, are primary determinants of timbre.

(Strong, 1963) interpreted the spectrum of various orchestra instruments and demonstrated that the oboe, clarinet and bassoon are first identified at the base in their spectrum of magnitude.

With these works, a larger number of studies are beginning to look for certain characteristics in the spectral format of the sounds for timbre recognition, such as the work proposed in (Luce and Clark Jr, 1967), which demonstrated that the metal family (wind instruments as Trombone, saxophone, trumpet, among others), is characterized by a unique *cutoff* in frequency, and that this feature strongly correlates with the identification of this family of instruments. Subsequently, (Benade, 2012) went further, and showed that frequency cutting is one of the main determinants of timbre in wind instruments in general.

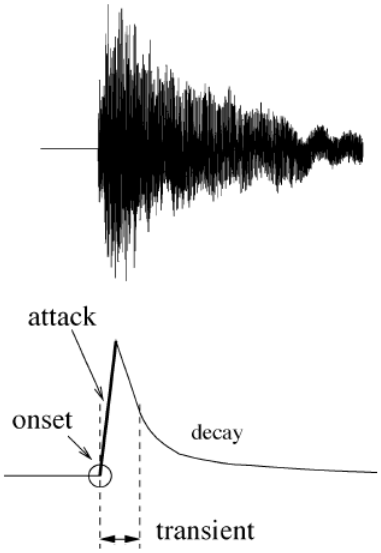

Figure 2: Spectrum format of a single note being executed (Bello et al., 2005).

Another characteristic that has been strongly correlated with timbre is the spectral centroid, which in general terms can be defined as the "swing point" of the spectrum, is directly related to the brightness of the instrument, a primary and subjective dimension of timbre, as verified in (Grey, 1977; Lichte, 1941; von Bismarck, 1974). In the works of (Beauchamp, 1982a; Beauchamp, 1982b), the author demonstrates that the centroid varies in many instruments with the loudness of the sound.

In the study (Strong, 1963), it was shown that, in the recognition of many instruments, the identification occurs mainly by the temporal envelope, partly because its spectral envelope is not unique, remembering that the definition of timbre is exclusive, being responsible for differentiating sounds that have the same height and intensity (Klingbeil, 2009).

In this context, it is possible to observe that the recognition of the timbre is strongly linked to the form of the sound spectrum, which is usually divided into *onset*, *attack*, *transient* and *decay*. This format is associated with the ideal case of a single note being executed, as shown in Figure 2 (Bello et al., 2005).

Another characteristic that starts to be analyzed in relation to the influence of timbre alteration are the temporal aspects of the sound, whose importance was first recognized by (Risset and Wessel, 1982). Following this finding, one can cite the work of (Handel, 1995), which presents the unification of the use of temporal and spectral characteristics in timbre classification.

Since (Schoenberg, 1978) considered the timbre as a second dimension of the tone, and today it is known that timbre can be considered in some way in a multidimensional characteristic, as described first in (Grey, 1977), and can even be represented Visually, as was presented in the paper proposed by (Soraghan, 2014).

One of the first complete works with a focus of recognition of timbre, with the use of several of the presented characteristics, was proposed by (Martin and Kim, 1998), that considers that there are two categories of characteristics for timbre recognition: temporal and spectral, both with Recognition of timbre. In this work some characteristics are used, such as:

- Pitch: Signals produce an identifiable structure in relation to height in the correlogram, in a two-dimensional frame, with horizontal axis lag and frequency relative to vertical, vertical grooves indicate the signal period, and by inversion, pitch.
- Spectral envelope: once the pitch has been detected, the vertical crest height of the correlogram can be measured as a function of the frequency, to obtain an estimate of the shape of the spectral envelope. The spectral centroid is simply the centroid of the spectral envelope.
- Loudness: the sum of the energy in the spectral envelope approximates the instantaneous sound intensity of the signal. Tracking this over time leads to simple measures of amplitude modulation, which may reveal the *Tremolo* and, by correlation with frequency modulations, resonances. As suggested in (Beauchamp, 1982a; Beauchamp, 1982b), the relationship between intensity and spectral centroid may be an important perceptual correlation of timbre.
- Asynchronous Attack: By tracking the spectral envelope over time, with competing pitch estimates, it is possible to measure the attack characteristics of a musical harmonic tone in a psychophysically appropriate way.
- Enarmony: the harmonic deviations in the signal will be reflected as deviations of the vertical structure in the instant correlogram.

In all, the work of (Martin and Kim, 1998) used thirty-one different characteristics, most being variations from those cited above. Among the recognition strategies adopted, the k-NN method was used, hierarchical and non-hierarchical for the identification of timbres.

As described by (Eronen et al., 2001), *Formants* are protuberances created by one or more resonances at the sound source. They represent the essential information for voice and speech recognition, and also for the recognition of musical instruments. A robust feature to measure the information of *Formants*, or the smoothed spectral envelope, is the ceptral coefficients. Among the techniques used in the work of (Eronen et al., 2001), is *Mel-frequency Cepstral Coefficients* (MFCC) (Davis and Mermelstein, 1980).

The MFCC method has become one of the most popular techniques for extracting features in automatic speech recognition systems, (Eronen et al., 2001), as it provides a fairly efficient description of the spectral form of the sound, and is then widely used in timbre recognition. A *cepstrum*[1] is the result of applying the inverse Fourier transform (IFFT) to the logarithm of the estimated spectrum of a signal. The MFCC is based on the cepstrum procedure. In it, the frequency bands are positioned logarithmically using the Mel scale[2]. A Fourier transform is replaced by a discrete cosine transform (DCT)[3]. It has an efficient "energy-compacting" property: most signal information tends to focus on some low-frequency DCT components. This is why, by default, only

1. The name "cepstrum" comes from the word "spectrum" with the first letters in reverse order, referring to the use of the inverse Fourier transform.
2. The Mel scale, named by (Stevens et al., 1937), is a scale of perceived pitches judged by listeners to have equal spacings between sounds.

   This frequency range is closer to the response of the human auditory system than the linearly spaced frequency bands.
3. A discrete cosine transform (DCT) is similar to the discrete Fourier transform (DFT), but it uses only real numbers.

the first 13 components are returned. By convention, the zero coefficient simply indicates the mean energy of the signal (Lartillot et al., 2014).

3. A Agent-Based System For Timbre Classification

Since each instrument has certain timbre characteristics that are identified by different descriptors, it is possible to imagine that a distributed and parallel solution is suitable for polyphonic recognition, increasing the efficiency in problem solving.

According to (Thomaz, 2009), agent technology becomes particularly suitable for musical applications, due to the possibility of associating a computational agent with the role of a singer or instrumentalist. He highlights some advantages of such associations, like to map features such as performance, perception, adaptation and improvisation on one side, and artificial processes on the other. In addition, it is possible to define forms of social interrelation between agents, which brings this technology even closer to collaborative musical performance.

In order to solve the problem of recognition of timbres, using an agent-based approach, this work uses a set of agents specialized in certain timbres, where each one is responsible for the classification of an instrument, as exemplified by Figure 3.

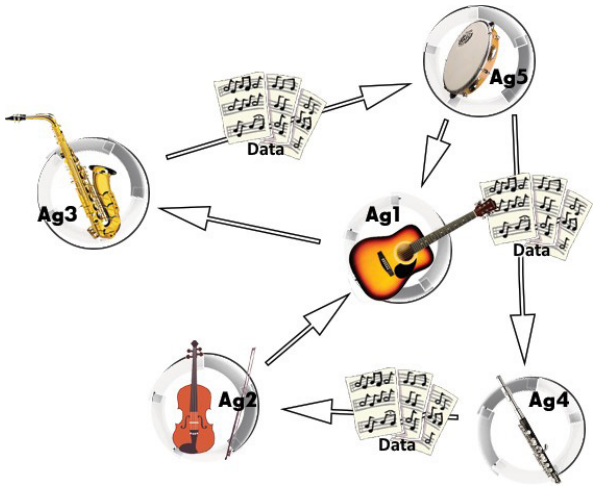

Figure 3: Simplified representation of the proposed system. Five specialist agents are presented, each represented by the instrument that he is responsible for recognizing. Within the environment, agents exchange information between them, as well as musical data (set of extracted features) that need to be classified.

These are cognitive agents, which can receive a set of characteristics from the environment or from other agents. When a new audio signal is preprocessed, and its characteristics are extracted, this new input is sent to the environment, where an agent is in charge of starting the classification process. If it recognizes the characteristics as being the correct recognizer, with an empirical percentage tolerance, it returns to the environment notifying that the sound has been properly classified. If the specialist agent does not reach the expected percentage, he forwards the feature set to a new agent. This process is represented in the architecture shown in Figure 4.

Among the advantages expected in adopting an agent-based system for solving the problem of classification of timbres, we can mention:

- Greater scalability of the system, because it is possible to add a new agent, responsible for classifying a new instrument, without conducting a new system-wide training.
- Parallelization of the classification, because when there is a classification task of many entries, the classification of them can be distributed among several agents.

- Improving the results, since each agent will be specific for each timbre, it is possible that they are better experts than a single classification entity, or that, depending on the implementation, they exchange information among themselves for better results.

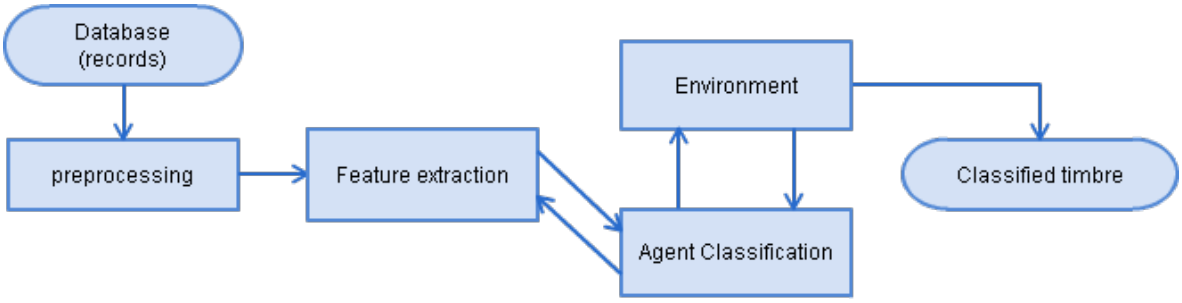

Figure 4: Architecture of the developed system.

4. Development

The agent-based system was developed in NetLogo (Wilensky, 1999). The agent training and classification operations were developed in MatLab (Guide, 1998) using MIRtoolbox (Lartillot and Toiviainen, 2007) for audio reading and extraction of musical features. The integration between these two tools was done through an unofficial extension of NetLogo, called MatNet[4].

The database consists of a collection of files with single note recordings of various instruments. The audio files used were obtained through OLPC (*One laptop per child - free sound samples*[5]).

Figure 5 presents the graphical interface of the developed system, where are represented the agents and the interaction between them.

4.1. The Training Method

The training method receives three NetLogo parameters, the agent ID to be trained, the agent name, and the number of files it will use in training. Basically, this function reads the audio files from the database, executes the MFCC method using MIRtoolbox, and stores 13 coefficients for each audio file in a global variable "base", of dimensions Nx13xM, where N is the number of files used in training and M refers to a third dimension used for each specialist agent. In other words, if 4 agents are trained using 5 files, the size of the array will be 5x13x4. This function can be called in NetLogo individually for each agent or perform training for all agents in the environment.

To exemplify the training procedure, we will consider the call of the training function for the guitar agent (associated to index 0) with three files for training.

$$training(0, \text{“}guitar\text{”}, 3) \quad (1)$$

4. http://github.com/mbi2gs/netlogo-matlab-extension/wiki
5. http://wiki.laptop.org/go/Free_sound_samples

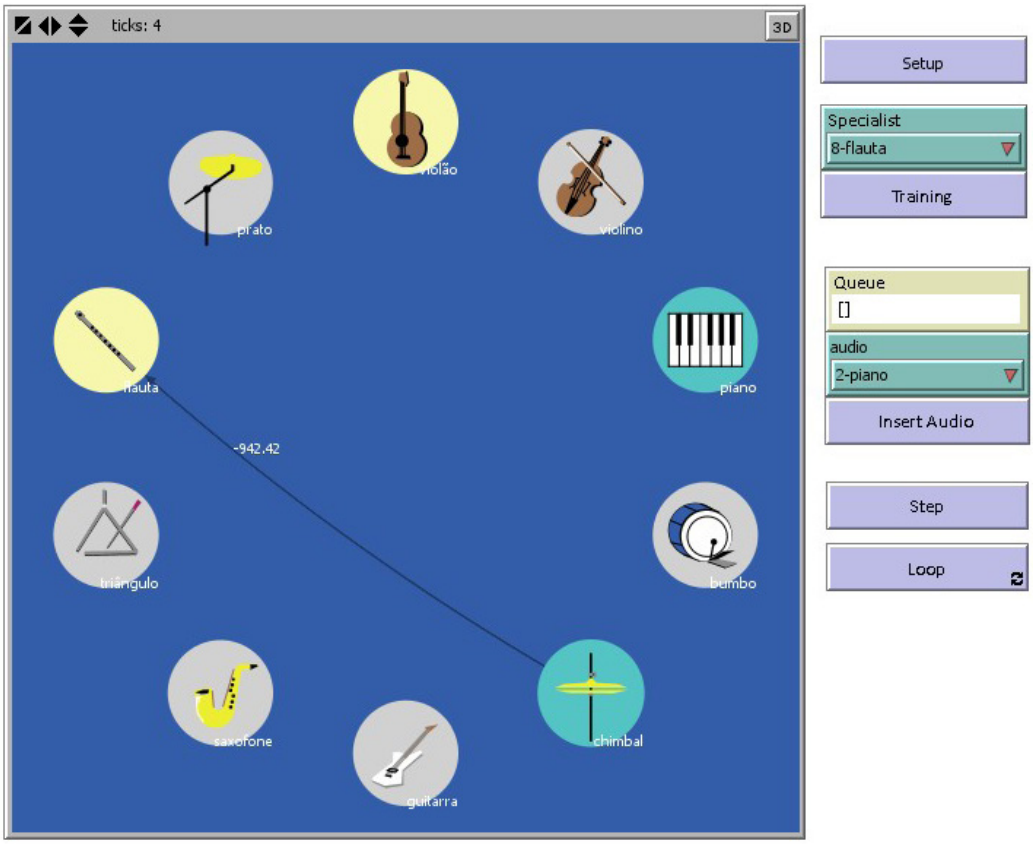

Figure 5: Within the world view are represented 10 agents, where the agent chimbal after not being able to classify its entrance, sends it to the flute agent.

By default, the training function considers the file indexes from 1 to N, in this case, where N = 3, the files "0-1.wav", "0-2.wav" and "0-3.wav". For each of these inputs the training method solve the MFCC, resulting in 13 coefficients for each of the files, as can be seen in Figure 6.

As a result of this procedure, the global variable "base", matrix of dimensions Nx13xM, takes in this case the dimensions 3x13x1. The value of M increases each time a new instrument is trained.

1.2. The Classification Method

The classification method receives three parameters: the identifier of the agent that will classify, the name of the agent and a handle of the corresponding audio file. The MFCC method is applied to the audio to be analyzed in the same way as it is performed in the training process, also obtaining 13 coefficients. Subsequently, the average between the MFCC result of all the files read in the training is used to be used as a comparison source with the file that will be classified. Next, the error between the file and the result of the average term is calculated and normalized by the difference between the highest and the lowest value of the MFCC. In response, the function returns an accumulated value of the normalized error calculation, which within NetLogo will be tested: if it is greater than 0.6, consider the file classified, otherwise it will send to another available agent.

To exemplify the classification procedure, we will consider the call of the classification function for the guitar agent (associated to index 0) with a hi-hat input (associated to index 4) to be tested, as seen in the function call in equation 2.

$$\textit{classification}\ (0, \text{“}\textit{guitar}\text{”}, 4) \tag{2}$$

The first step of the classification procedure is to extract the MFCC from the file to be tested. By default, the input to be compared is not part of the training group, which is assigned the file index of zero. The spectrogram and MFCC result applied to the "4-0.wav" file, which contains the recording to be classified from the hi-hat instrument, are shown in Figure 7.

The next step is to calculate the module of the difference between the mean values of the coefficients obtained in the training with the MFCC values of the input to be classified. These values can be seen in Figure 8.

Finally, the obtained differences are summed in an accumulating variable, representing the total error obtained, in this case 6,22. This value is normalized in relation to the maximum and minimum values of the calculated average, returning only a single real value.

Finally it is retrieved in NetLogo, if it is greater than 0.6 it is considered the correctly classified entry, otherwise it is forwarded to a new specialist agent, who will repeat this procedure.

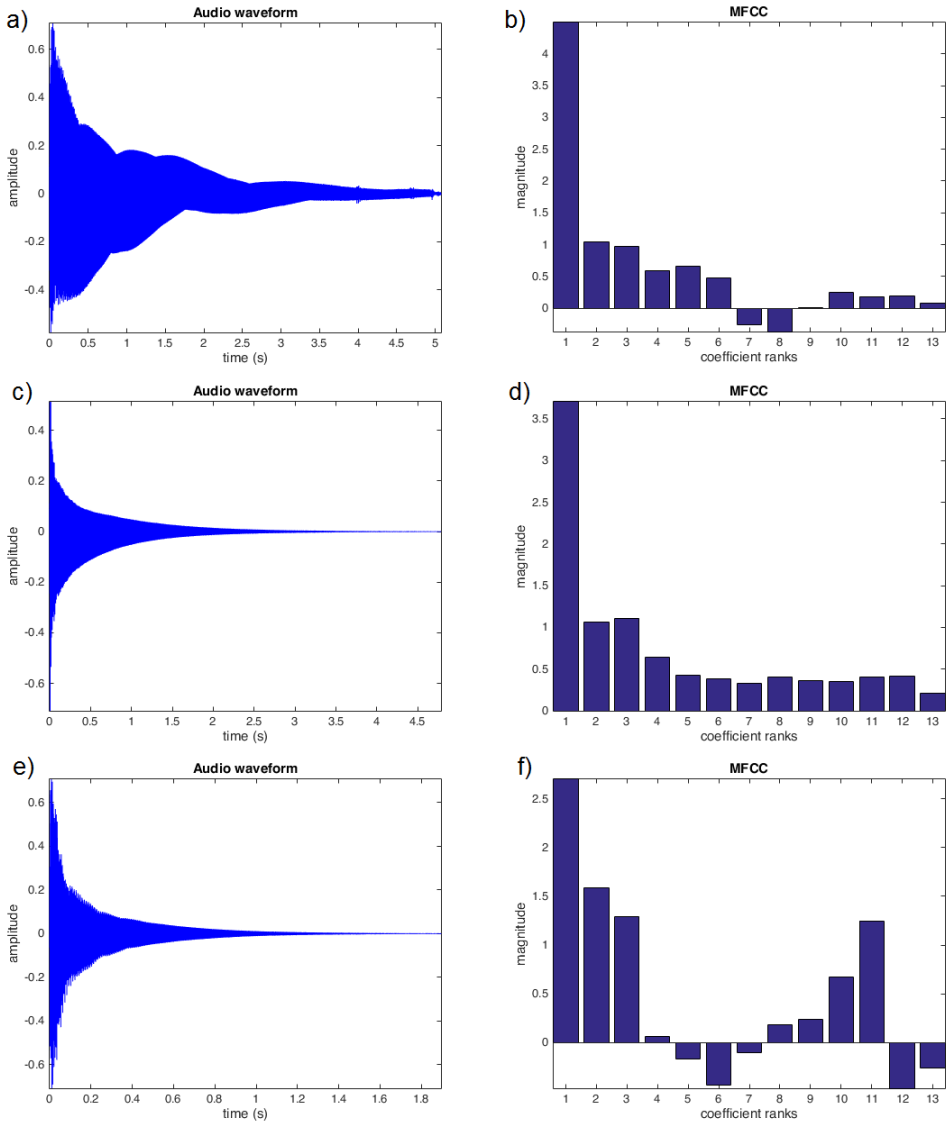

Figure 6: a) spectrogram of file "0-1.wav". b) MFCC of file "0-1.wav". c) spectrogram of file "0-2.wav". d) MFCC of file "0-2.wav". e) spectrogram of file "0-3.wav". f) MFCC of file "0-3.wav".

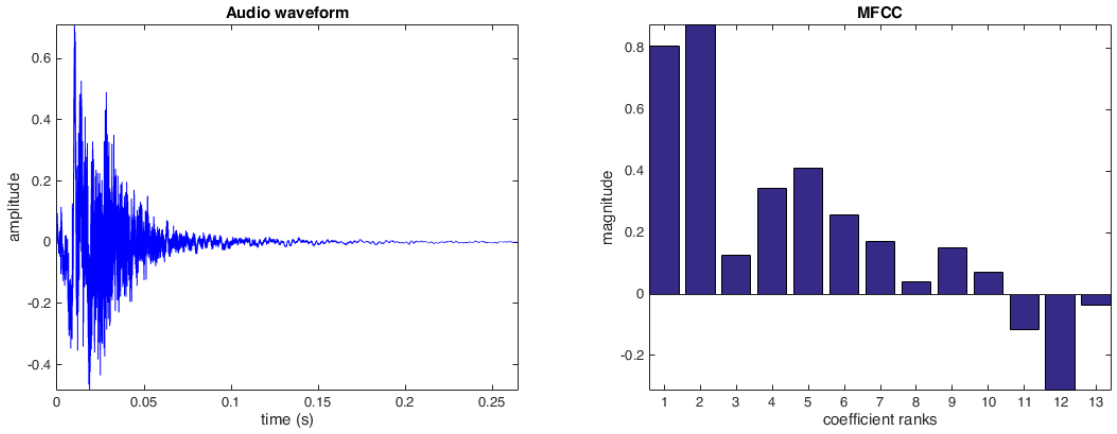

Figure 7: Spectrogram and MFCC of file "4-0.wav".

	1	2	3	4	5	6	7	8	9	10	11	12	13
0-1	4,49	1,04	0,97	0,59	0,65	0,47	-0,25	-0,36	0,00	0,25	0,18	0,19	0,07
0-2	3,70	1,06	1,10	0,63	0,42	0,38	0,33	0,40	0,36	0,34	0,40	0,41	0,21
0-3	2,69	1,59	1,29	0,06	-0,16	-0,43	-0,10	0,18	0,23	0,67	1,24	-0,46	-0,26
Average	3,63	1,23	1,12	0,43	0,30	0,13	-0,00	0,07	0,20	0,42	0,61	0,04	0,00
4-0	0,80	0,87	0,12	0,34	0,41	0,25	0,17	0,04	0,14	0,07	-0,11	-0,31	-0,03
Abs(Dif)	2,83	0,36	1,00	0,09	0,11	0,12	0,17	0,03	0,06	0,35	0,72	0,35	0,03

Figure 8: In the lines 0-1, 0-2 and 0-3 are the results of the MFCC performed in the training of the guitar agent, the fourth line presents the average between the results of the training, the line 4-0 displays the values extracted from the MFCC of the file to be tested, and the last line shows the values of the difference module between the specialist agent average and the file to be tested.

5. Conclusion

In this work an agent - based system was proposed to solve the problem of classification of musical timbres, in which for each instrument a specialist agent capable of recognizing only a single instrument would be associated. In a multi-agent environment, where multiple cognitive agents can perform information exchanges, the classification of timbres would be performed in a distributed, parallel and scalable way.

The classification method implemented proved to be quite fast, according to its simplicity, but classification errors were observed in instruments of the same family, such as classical guitar and guitar. It could be happen possibly because of the great similarity of the MFCC coefficients of these files. To solve this problem, as future works, we will use a larger feature set that would explore other behaviors in the time domain and frequency domain.

Many are the possibilities of future work, considering that this is a pioneering work in timbre classification in a distributed way.

One possible application that has been in great demand by musicians from different areas is the improvement of the musical transcription (transcription is the act of transcribing the music into a documentary form, such as a score). Many software perform musical transcription quite accurately, but still do not show good results in instrument identification. The most commonly used method for transcription is a simple division of heights, based on the frequency of sound. This approach is rather flawed as there are instruments with wide registers, ranging from very low pitches (low frequencies) to very high pitched (high frequencies) sounds, like the piano. A timbre classification system would help to correctly separate the voices of the instruments within a score, taking into

account their timbral characteristics and not just their frequencies. The Ulises system, presented here, has an advantage for an approach like this: because it is a distributed system, to identify new instruments it would not need big software updates, just adding a new specialist agent when necessary.

One more suggestion of future work is the refinement of the classification methods implemented. We used only one method based on the comparison with the average of the MFCC between several entries, but there are different characteristics that help increase the accuracy of the classification. An example is the use of DTW (Dynamic Time Warping), a technique used in speech recognition that has the advantage of comparing temporal sequences independently of their durations. In this sense it is possible to elaborate a much more robust system, in which each specialist agent takes into consideration only the most important characteristics for the recognition of its instrument.

6. 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.

von Bismarck, G., 1974. Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acta Acustica united with Acustica*, 30(3):146-159.

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. *International Journal of Computer Science & Informatics (IJCSI)*, II(1):21-24.

Eronen, A. et al., 2001. Automatic musical instrument recognition. *Memoire de DEA, Tempere 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*. Ph.D. 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*. Ph.D. 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 laboratorio multiagente de geraÃgˇÃcˇo 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., Volkmann, 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.
Wilensky, U., 1999. NetLogo.