Revista Vórtex | Vortex Music Journal | ISSN 2317–9937 | http://vortex.unespar.edu.br/
D.O.I.: https://doi.org/10.33871/23179937.2023.11.1.7068

Music Segmentation and Similarity Estimation Applied to a Gaze-Controlled Musical Interface

Higor A. F. Camporez, Yasmin M. de Freitas, Jair A. L. Silva,

Leandro L. Costalonga, Helder R. de O. Rocha

Federal University of Espirito Santo| Brazil

Abstract: Assistive technology, especially gaze-controlled, can promote accessibility, health care, well-being and inclusion for impaired people, including musical activities that can be supported by interfaces controlled using eye tracking. Also, the Internet growth has allowed access to a huge digital music database, which can contribute to a new form of music creation. In this paper, we propose the application of Music Information Retrieval techniques for music segmentation and similarity identification, aiming at the development of a new form of musical creation using an automatic process and the optimization algorithm Harmony Search to combine segments. These techniques for segmentation and similarity of segments were implemented in an assistive musical interface controlled by eye movement to support musical creation and well-being. The experimental results can be found in [https://bit.ly/2Zl7KSC].

Keywords: Music Information Retrieval, Eye tracking, Optimization, Musical Interface.

Received on: January 05, 2023. Approved on: March 01, 2023. Online available on: April 15, 2023. Presented at the Ubiquitous Music Symposium (2022) at the State University of Paraná (Brazil). Guest editors: Prof. Dr. Damián Keller, Dr. Tereza Raquel Alcântara-Silva, Dr. Bruno Mesz.

Around the world, millions of people present some sort of disability (visual, auditory, motor, or intellectual) (WHO, 2012). However, technological advancements, especially in terms of assistive technology (AT), have been helping these people, since AT develops products to provide a better life condition specifically for them (CHOI; SPRIGLE, 2011).

The AT makes itself present in several areas including musical activities, allowing disabled people to create music not only for pleasure and well-being but also during the rehabilitation process (AGRES, et al., 2021). For example, instrumental executions can help in motor rehabilitation as well as in cognitive enhancement, since music activates several areas of the brain (LARSEN; OVERHOLT; MOESLUND, 2016). Thus, customization of traditional musical instruments can be done to support their use by disabled people, such as a keyboard with large keys that can be pressed by a handle shown in (LOURO; IKUTA; NASCIMENTO, 2005). Another example is the musical game GenVirtual (CORREA et al., 2007) which uses augmented reality technology for cognitive and motor rehabilitation. A review of accessible digital musical instruments that includes many types of control interfaces such as touchless controllers, brain-computer music interfaces, adapted instruments, wearable controllers and others can be found in (FRID, 2019). These types of technologies include gaze-controlled interfaces that use an eye tracker to estimate the user's eye movements and points of gaze, helping to interact with a device without limb movement (MAJARANTA; BULLING, 2014). A prospective for eye-controlled musical performance can be found in (HORNOF, 2014).

The Internet has changed how musical content has been disseminated, facilitating access to a wide database of musical content. The organization and extraction of information from this enormous online database contributed to the creation of a research field known as music information retrieval (MIR) (FUTRELLE; DOWNIE, 2003). In addition, the insertion of technological resources in musical activities also led to the creation of the ubimus (ubiquitous music) research area (KELLER; LAZZARINI; PIMENTA, 2014; KELLER, 2018). Ubimus uses ubiquitous systems of human agents and material resources for musical purposes. As ubimus uses technologies, such as assistive musical interfaces, to support musical activities, it can bring health care and well-being to people, especially for people with disabilities who have limitations to play physical musical instruments. In addition, these types of interfaces can also increase the ubimus premise of availability

since it includes not only average people (novices in music) but also disabled people.

Within these fields of study, Kutiman ThruYOU[1] is an example of a project that uses a musical content database available on the Internet. The author uses YouTube videos to create new songs with no direct relations between them. Kutiman mixed slices of videos, made by amateur musicians, to create new songs using a manual search process in the YouTube database. A system to help Kutiman's searches, based on MIR techniques, was created to provide content filtering (LINDENBAUM et al., 2010). By means of this system, experiments using the same database used by Kutiman matched on average 60% of his previous choices. Likewise, Mix The City[2] project provides an interactive video clip interface that allows users to compose with slices of videos produced by local artists, enabling them to discover cities through the produced songs and videos. Therefore, non-musicians can create songs, since the available slices of videos have some musical similarities.

This paper is an extension of (CAMPOREZ et al., 2020) which describes strategies to combine musical segments extracted from original songs, aiming for its application in an assistive gaze-controlled musical interface. The interface shows the similarity level between segments to assist users, mainly non-musicians, in the creation process. Thus, we describe methods of music segmentation and similarity extraction to support assistive interfaces which can be seen as ubimus material resources for musical activities. Procedures are proposed for automatic creations and also for users' compositions. It is important to emphasize that the combination of MIR techniques and ubiquitous technologies, such as assistive interfaces, can not only provide well-being for people with disabilities but also increase the availability of this ubimus system. The remainder of this paper is organized as follows. Section 1 describes some gaze-controlled interfaces. Section 2 presents musical feature extraction used to find similarities between segments. Section 3 depicts our assistive interface. The relation of musical features is described in Section 4. In Section 5 is illustrated the harmony search algorithm process. Section 6 shows segment concatenation procedures. The experiments and the results are presented in Section 7 and the conclusions are provided in Section 8.

[1] thru-you.com

[2] mixthecity.britishcouncil.org

1. Gaze-Controlled Interfaces

The EyeMusic (HORNOF; SATO, 2004) is a musical interface that uses a commercial eye tracker, where the device's coordinates (x, y) are used to create a granular synthesis of click-sounding samples through the Max/MSP in real-time. The Oculog (KIM; SCHIEMER; NARUSHIMA, 2007) is a system created using Pure Data programming language with an eye tracker, where the pupil coordinates (x, y) are mapped in note number and velocity, respectively, to control a tone generator. The EyeGuitar (VICKERS; ISTANCE; SMALLEY, 2010) is a system based on the Guitar Hero[3] game style which uses eye tracker information as input to play, however, due to the eye tracker limitations and to be easier to play, it is only necessary to select the correct string and not the action to hit it together like in the original game.

The EyeHarp (VAMVAKOUSIS; RAMIREZ, 2012, 2016) is a gaze-controlled musical interface that is divided into three parts: a) the pie menu, b) the step sequencer, and c) the arpeggiator. The interfaces b) and c) are used for rhythmic and harmonic creation in background, whereas a) is used for melodic construction in real-time. Other gaze-controlled interfaces, outside musical context, are pEYEwrite and pEYEtop (HUCKAUF; URBINA, 2008) where pEYEwrite aims textual writing and pEYEtop focuses on computer navigation. The design of the interface follows a circle structure, named the pie menu. The authors reported that users had better performance in pEYEwrite than in the virtual ABCDE keyboard.

Olhar Musical (CAMPOREZ et al., 2018b, a) is a gaze-controlled musical interface that has six cells to control six different audio samples (segments) by sending MIDI messages to Ableton Live[4]. The interface allows the user to choose to play or pause the audio samples. The use by non-musicians is also an objective of the interface, therefore, it shows levels of similarity between the audio segments to help the user to choose the next audio segment. Audiovisual demonstrations can be accessed on YouTube[5].

Davanzo et al. (2018) proposed a digital musical interface controlled by eye movements, considering comfortable and effective interaction of the user. They consider an eye tracker and a

[3] www.guitarhero.com
[4] https://www.ableton.com/en/live
[5] https://youtu.be/YDj8pr3Cnx8

switch that can be of diverse types, depending on the needs of the performer. Thus, the user needs to look at and activate the switch. Dueto (VALENCIA et al., 2019) is a gaze-operated musical interface that supports melody and harmony creation through multiple input modalities: a) gaze-only, that uses a dwell time of 100ms to activate the selection of a graphical element; b) gaze + switch, where a player needs to look at an element and then use a switch to activate it; and c) gaze + partner, the modality that enables two players, one using gaze-only mode and the other using a multi-touch keyboard.

Cyclops is another gaze-controlled digital instrument created for live performance and improvisation, containing a synthesizer and a sequencer (PAYNE; PARADISO; KANE, 2020). Using the sequencer, the user can record a sequence and play it in a loop, in addition the user can also add effects to the sequence. Kandpal, Kantan and Serafin (2022) developed a real-time gaze-controlled digital interface for musical expression and performance using a physical model of a xylophone where twelve musical keys (an octave in an equitempered scale) were set up in a graphical interface. The authors reported that it is better to place graphical elements in the middle of the screen where the eye tracker has high accuracy.

2. Musical Features Extraction

Lindenbaum et al. (2010) created a system to assist in Kutiman's searches which use the following musical features and measures: a) Beats per Minute; b) Chromagram and Chromatic Scale and c) Cyclic Harmonic Cross-correlation. These attributes and measures are extracted from WAV files. The calculations are defined considering that the vectors $\underline{x}_1$ and $\underline{x}_2$ are two musical segments.

2.1. Beats per Minute

The musical feature beats per minute (BPM) describes the rhythm in which a song is played. Thus, generally speaking, BPM can be seen as the number of repetitions found in one minute that denotes the speed to play a song. DJs often use the BPM to choose which songs to mix, selecting those with similar BPM because they are easier to synchronize. In this work, the BPM relationship between

two musical segments can be calculated by Equation (1) (LINDENBAUM et al., 2010), where a low distance value represents a high similarity. The BPM distance equation is defined as:

$$D_{tempo}(\underline{x}_1, \underline{x}_2) = \frac{\left|tempo(\underline{x}_1) - tempo(\underline{x}_2)\right|}{\left[tempo(\underline{x}_1), tempo(\underline{x}_2)\right]}, \quad (1)$$

where $tempo(x)$ represents the BPM extracted from x.

2.2. Chromagram and Chromatic Scale

The feature chromatic scale contains 12 musical notes equally spaced by the interval of a semitone (BENWARD; SAKER, 2008). The chromagram, or harmonic pitch class profile, represents a histogram of musical notes that shows the energy distribution across the interval pitch classes (GÓMEZ, 2006). A chromagram can be mathematically expressed by

$$\underline{C}_x(b) = \sum_{m=0}^{M} |X_{cq}(b + m\beta)| \quad (2)$$

for M the total of octaves, b the chroma bin number, β the number of bins per octave and X_{cq} the frequency domain signal obtained by the Constant Q Transform (CQT) method (BROWN, 1991) using

$$X_{cq}[k] = \sum_{n=0}^{N(k)-1} w[n,k] \cdot x[n] \cdot e^{-j2\pi n f_k}, \quad (3)$$

where $f_k = 2^{\frac{k}{\beta}} \cdot f_{min}$ is the k^{th} frequency bin, for f_{min} the minimum frequency.

In this work, we considered that $\beta = 12$, f_{min} is approximately 32.7 Hz (C1), $b \in [1,12]$, $\underline{C}_x$ is a 12-length vector and the chromagram is normalized to values between 0 and 1.

2.3. Cyclic Harmonic Cross-correlation

For most listeners, two notes with several common harmonics sound pleasant (THOMPSON, 2015). For example, musical notes with intervals between 4 or 7 semitones share several harmonics, which makes them harmonically compatible (PISTON; DEVOTO, 1987). According to Lindenbaum et al. (2010), one way used to quantify harmonic similarity is based on the cross-relation of pieces' chromagram. Thus, the normalized cross-correlation (YOO; HAN, 2009), which returns values between -1 and 1, is defined as:

$$R_{1,2}(p) = \frac{\sum_l \left[\underline{C}_{x_1}(l) - \underline{\underline{C}}_{x_1}\right]}{\sqrt{\sum_l \left[\underline{C}_{x_1}(l) - \underline{\underline{C}}_{x_1}\right]^2}} \times \frac{\sum_l \left[\underline{C}_{x_2}(l - p \bmod 12) - \underline{\underline{C}}_{x_2}\right]}{\sqrt{\sum_l \left[\underline{C}_{x_2}(l - p \bmod 12) - \underline{\underline{C}}_{x_2}\right]^2}}, \quad (4)$$

where p is the cross-correlation, l is the vector length and $\underline{\underline{C}}_{x_i}$ is the $\underline{C}_{x_i}$ mean value. Following the process described by Lindenbaum et al. (2010), the higher cross-correlation value between $R_{1,2}(0)$, $R_{1,2}(4)$ and $R_{1,2}(7)$ indicates that there are many common harmonics, which reflects a great compatibility. Using the maximum achieved value, the chroma distance is calculated as

$$D_c(\underline{x}_1, \underline{x}_2) = \frac{1}{2}\left[1 - R_{max}(\underline{x}_1, \underline{x}_2)\right], \quad (5)$$

for $R_{max} = max\left[R_{1,2}(0), 0.8R_{1,2}(4), 0.8R_{1,2}(7)\right]$.

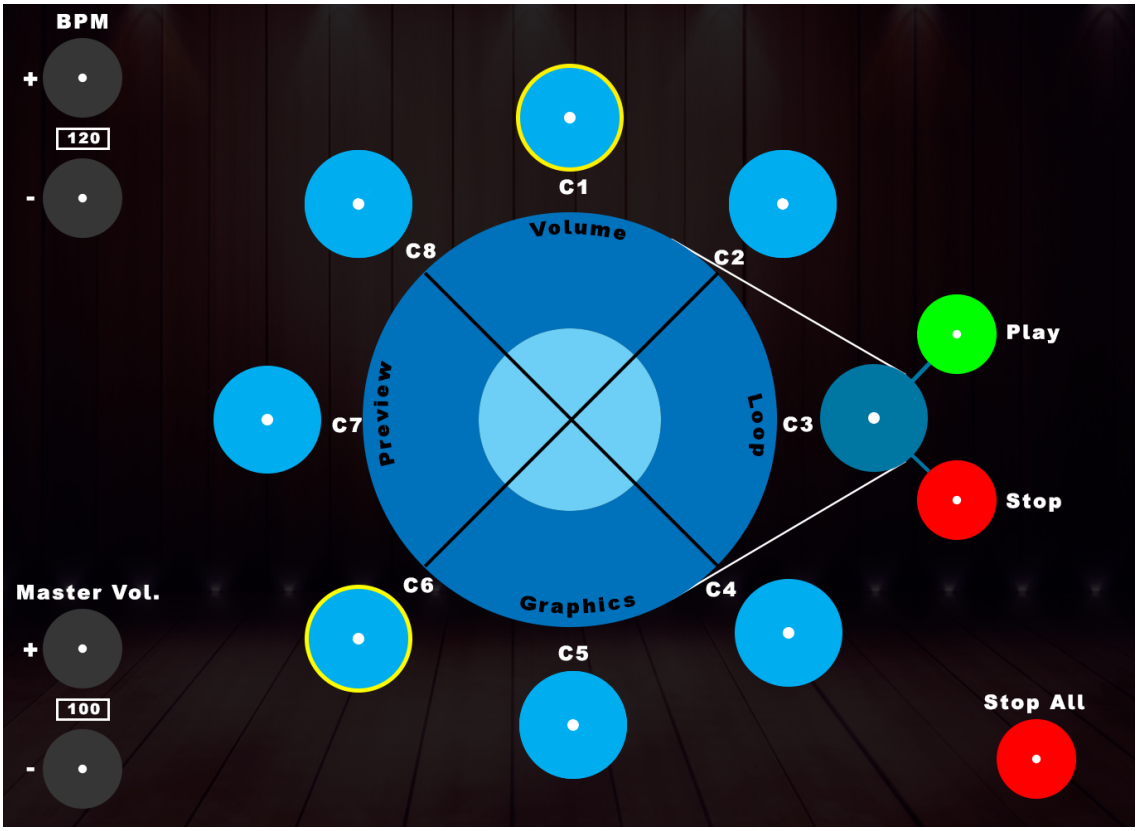
3. Assistive Musical Interface

We designed a new interface for our previous work Olhar Musical (Figure 1) (CAMPOREZ, et al., 2018a). The new design was inspired by (HUCKAUF; URBINA, 2008) that presents good results for circle interfaces structure and in (VAMVAKOUSIS; RAMIREZ, 2012, 2016) which uses

circles with dots in the center to help the user focus on the circle. Figure 1 demonstrates our gaze-controlled musical interface with eight light blue circles (C1, C2, C3, C4, C5, C6, C7, C8) where each one controls a musical sample (a segment). In other to control each light blue circle, there are the buttons play and pause, as well as the center circle that represents some configuration for segments. Thus, users can control volume; loops, number of repetitions; preview, to listen through a specific channel out of the performance; and graphics, to see a visual representation of the segment. Users can also control beats per minute, master volume, and stop all, a button that pauses all segments in execution.

This work is planned to be used by non-musicians as well. Thus, the interface shows similarity levels between the segments through the yellow ring, where the thickness expresses the similarity level. The similarity level between musical segments is defined by using the musical features explained in Section 2. The relationship between the features is shown in Section 4 and the interface usage guide process is shown in Section 6.

FIGURE 1 – Assistive Musical Interface controlled by eye tracking.



4. Musical Features Relationship

Our interface, cited in the previous Section, uses a large number of musical segments that can be cropped from song databases. Then, finding good positions for songs segmentation is extremely important for the quality of the segments and their overall "mixability". Thus, the traditional musical structures, i.e. bar, can provide a good segmentation strategy for our interface.

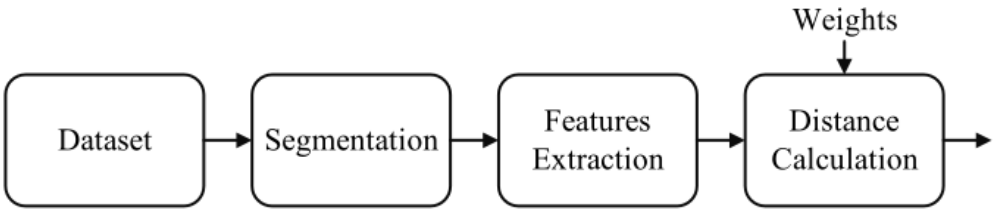
Music bars can be found through the spectral differences in audio samples (DAVIES; PLUMBLEY, 2006). Developed by the Center for Digital Music (C4DM), the bar detector is implemented in the QM Vamp Plugin Library (CANNAM et al., 2014) for Vamp audio analysis plugin system[6]. In addition, the calculations of chroma, BPM, and RMS were done with Librosa library (MCFEE et al., 2015).

According to Lindenbaum et al. (2010), the relation between the musical feature distances can be defined as:

$$D\left(\underline{x}_1, \underline{x}_2\right) = \sum_{i \in A} w_i D_i(\underline{x}_1, \underline{x}_2)\,, \tag{6}$$

where A is a group of features and w_i is the weight set for each feature. A default value for w_i can be obtained by $\frac{1}{|A|}$. Figure 2 shows a block diagram of the process employed in the quantification of segments, as described in Equation (6).

FIGURE 2 – Block diagram of the process used in the quantification of segments.



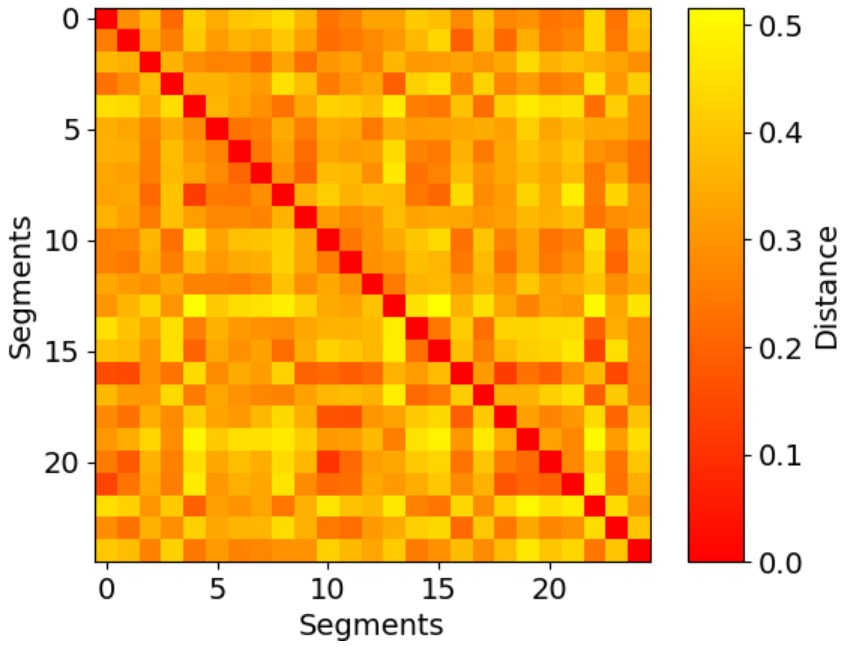
[6] https://www.vamp-plugins.org/

Beginning from the dataset, subsequently, each sound is segmented into bars based on the plugin above-mentioned. Afterward, the features of each segment are extracted, and, finally, a many-to-many relationship is calculated to find similarities between the segments. The relationship can be explained in matrix form, where Figure 3 shows 25 segments relationships. The matrix shows the distances between segments computed by Equation (6), where lower distances (red color) represent a high similarity level between the segments.

5. Harmony Search Algorithm

One way to find the best solution to a problem is knowing all of its solutions. However, this strategy may cost a lot of time to be processed. Thus, optimization algorithms can be used to find satisfactory solutions with reduced processing time (VENTER, 2010). One application example is the optimization of a delivery system where the route can be optimized to minimize some parameter such as traveled distance or time spent (or either) which is named objective function. In this case, the optimizer considers many parameters like traffic information, order address, order priority, weather condition and others.

FIGURE 3 – Similarity matrix for 25 segments.



There are many optimization algorithms inspired by different things such as ant colony (DORIGO; DI CARO, 1999) and grey wolves (MIRJALILI; MIRJALILI; LEWIS, 2014). In addition, there is an algorithm inspired by the improvisations process of Jazz musicians named harmony search (HS) (GEEM, 2010), where a decision variable represents a musician; a musical instrument's pitch range corresponds to the decision variable's range; musical harmony at certain time relates to a solution vector at certain iteration; and audience's acceptance corresponds to the objective function to evaluate the solution. Musical harmony is improved time after time, in the same way, the solution vector is upgraded iteration by iteration.

The HS uses some configuration parameters. Harmony memory size (HMS) is the number of solution vectors handled at the same time; harmony memory considering rate (HMCR) is the rate at which the HS chooses one value randomly from the musician's memory. Thus, (1 - HMCR) is the rate at which the HS accepts one random value from the total value range. The pitch adjusting rate (PAR) is the rate at which the HS modifies the value which was selected from memory, consequently, (1 - PAR) is the rate at which HS keeps the original value. Maximum improvisation (MI) is the number of iterations.

The i-th vector solution, which can be seen as a harmony, is represented by $x^i = \{ x_1^i, x_2^i, x_3^i, \cdots, x_n^i\}$, where x_a^i is a decision variable. Relating to jazz improvisation, a decision variable is a note chosen by the musician that is represented by the variable. Each decision variable contains a certain allowed range (e. g. instrument's pitch range) and the objective function ($f(x^i)$) to be maximized or minimized (e. g. audience's acceptance). The HS initializes randomly the harmony memory (HM) with at least HMS times. The HM can be represented by a matrix:

$$HM = [x_1^1\ x_1^2\ \vdots\ x_1^{HMS}\ \ x_2^1\ x_2^2\ \vdots\ x_2^{HMS}\ \cdots\cdots\ \ddots\ \cdots\ x_n^1\ x_n^2\ \vdots\ x_n^{HMS}\ | f(x^1)\ f(x^2)\ \vdots\ f(x^{HMS})\ ] \qquad (7)$$

where the matrix is sorted by the optimization function $f(x^1) \leq f(x^2) \leq \cdots \leq f(x^{HMS})$.

Generally, in Jazz improvisation, a musician plays a note by selecting it from the allowed range or by adjusting a note from his/her memory. Thus, the HS follows the same idea picking values from the available range, from the HM directly or from the HM with some modification. Therefore, the

HS selects a new value for a decision variable following these rules:

- **Random Selection**: the new value x_i^{new} for a new harmony is randomly selected from an allowed range with the probability of $(1 - HMCR)$.
- **Memory Consideration**: x_i^{new} receives the value x_i^j selected from $HM = \{x_i^1, \cdots, x_i^{HMS}\}$ with probability of $HMCR$. The index j can be calculated by uniform distribution: $j \leftarrow int(U(0,1) \cdot HMS) + 1$.
- **Pitch Adjustment**: if the x_i^{new} is chosen from the HM, it can be adjusted with the probability of PAR. For discrete variables, if $x_i(k) = x_i^{new}$, the pitch-adjusted value becomes $x_i(k+m)$, where $m \in \{-1, 1\}$.

After generating the x^{new}, the HS calculates its objective function value. If the evaluation of x^{new} is better than the worst harmony in HM, the worst harmony is replaced by x^{new}. The HS, for discrete variables, can be represented by the following pseudo-code:

a. **Step 1**: Configure the HS parameters;
b. **Step 2**: Generate HM matrix;
c. **Step 3**: set $i = 1$;
d. **Step 4**: if $rand < HMCR$ go to step 5, otherwise set x_i^{new} with a random value from the available range and go to step 7;
e. **Step 5**: select a harmony randomly from the HM (for example x^k) and set $x_i^{new} = x_i^k$;
f. **Step 6**: if $rand < PAR$ do the pitch adjustment and go to step 7, otherwise just go to step 7.
g. **Step 7**: if $i <= n$ do $i = i + 1$ and go back to step 4, otherwise go to step 8
h. **Step 8**: If the new harmony (x^{new}) is better than the worst harmony from HM, replace the worst harmony by the new one and go to step 9, otherwise go to step 9
i. **Step 9***:* If it is the last iteration go to step 10, otherwise go to step 3.
j. **Step 10**: Return the best harmony

6. Segments Concatenation Procedures

6.1. Automatic Concatenation

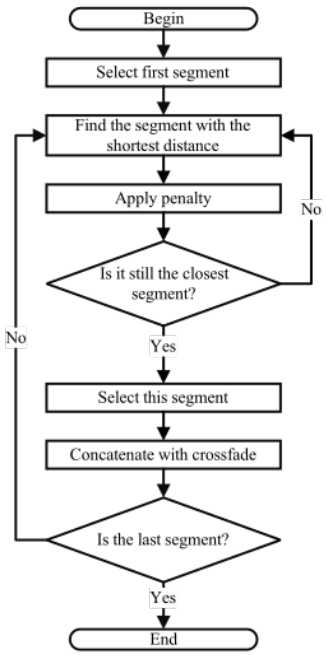
Figure 4 shows a flowchart of the proposed segment concatenation process. After the choice of a propellant segment (segment a), the segment closest to it (segment b) is searched. Next, the segment closest to segment b (segment c) is selected, and this process is repeated until a number of segments is reached.

In the concatenation process, there is a possibility of loops between segments. For example, the segment closest to a is b and the segment closest to b is a. Thus, the segment choice procedure can get inside a loop between a and b. Therefore, to avoid a local minimum, we defined a penalty for the distance, where the expression is defined as:

$$D_{pen} = D \cdot e^{\alpha n}, \tag{9}$$

where n is a number of repetitions and α is a penalty factor.

FIGURE 4 – Segments concatenation processes.



In the concatenation process between two segments, an abrupt transition may occur because of the audio segmentation process. Thus, to avoid these kinds of transitions, we employed a crossfade technique with fade in and fade out according to the following formulas

$$\frac{\sqrt{(1+t)}}{2} \; and \; \frac{\sqrt{(1-t)}}{2}, \tag{8}$$

respectively, where the intersection area is mapped for $-1 \leq t \leq 1$.

To explain the concatenation process using the crossfade technique, Figure 5 shows the concatenation of two sine wave that represents two segments. Figure 5 (a) demonstrates the two segments and the fade curves described in Equation (8). Figure 5 (b) depicts, in the highlighted area, the fade effects applied to the segments where segment 1 was multiplied by the fade out curve and segment 2 was multiplied by the fade in. It is possible to notice a decrease in segment 1 amplitude (fade out) and an increase in segment 2 (fade in). Figure 5 (c) shows the final result which is segment 1 added to segment 2 after the fade process.

FIGURE 5 – Segment concatenation using crossfade.

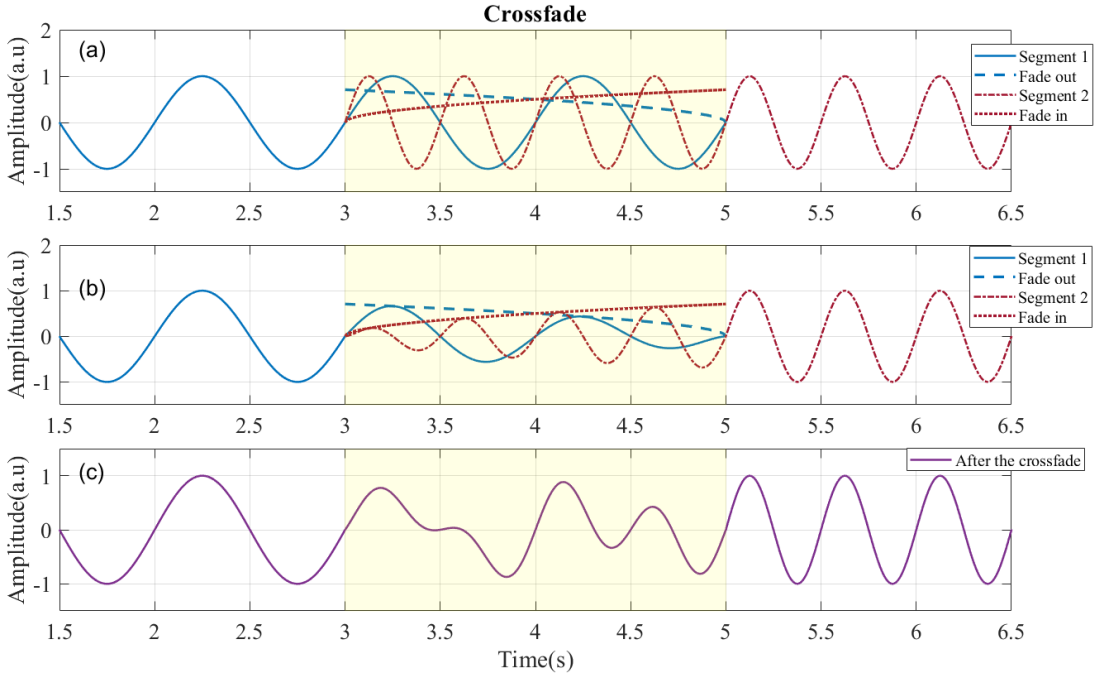


Figure 6 also shows an example of concatenation between two segments with the values of the employed features. The highlighted area shows the intersection where the crossfade was applied.

6.2. HS Concatenation

The segments can also be selected by the HS algorithm. Thus, we used the HS to select n segments represented by $x = \{x_1, x_{2,} \cdots, x_n\}$ to be concatenated. The HS chooses segments to minimize the objective function (cost function) which is defined as:

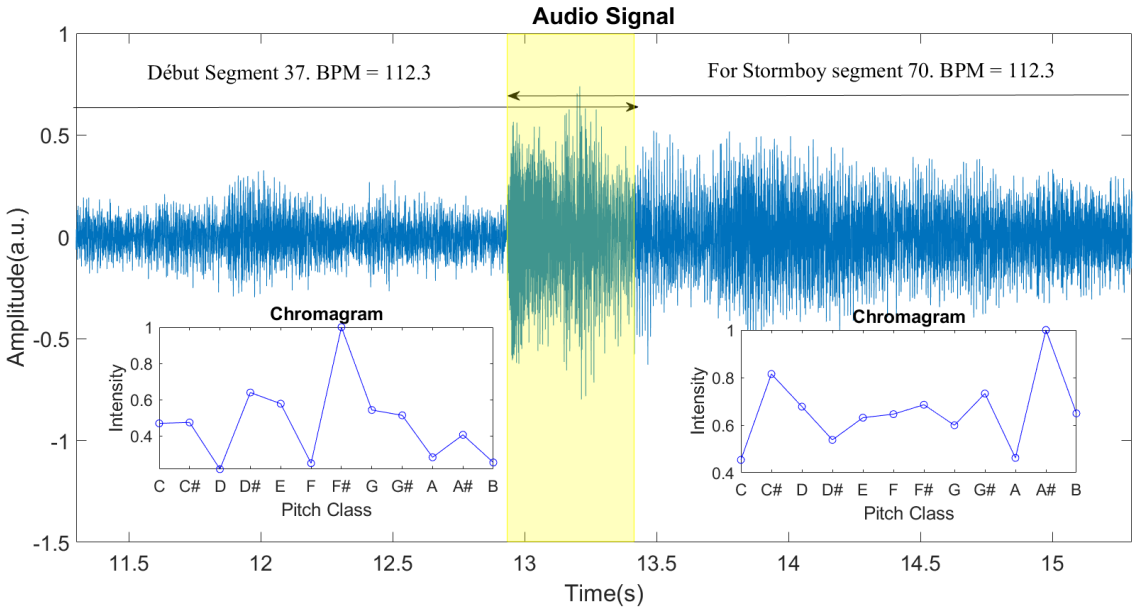
$$f(x) = \left[\sum_{i=1}^{n-1} D(x_i, x_{i+1})\right] \cdot e^{\alpha pen(x)} \quad (10)$$

where D is defined in Equation (6), α is a penalty factor and $pen(x)$ is the sum of the repetitions considering the last 5 segments of each segment. The $pen(x)$ equation is defined as:

$$pen(x) = \sum_{i=5}^{n-1} count\ x_i\ in\ \{x_{i-5}, x_{i-4}, x_{i-3}, x_{i-2}, x_{i-1}\} \quad (11)$$

After the HS finds a good solution, the segments are concatenated following the process described in Subsection 6.1. We configured the parameters of HS as following $HMS = 20$, $HMCR = 0.95$, $PAR = 0.2$, $MI = 25000$, and $n = 20$.

FIGURE 6 – Features of two concatenated segments.



6.3. Semi-automatic Concatenation

One of the goals of our assistive musical interface is its use by non-musicians. Thus, we use the musical features explained in Section 4 to facilitate users' choices. In the interface, when the user focuses on a cell, the gray window (Figure 7) presents the musical features as well as the relationship (distance) between the last choice and the focused cell.

FIGURE 7 – The assistive musical interface showing musical features information.

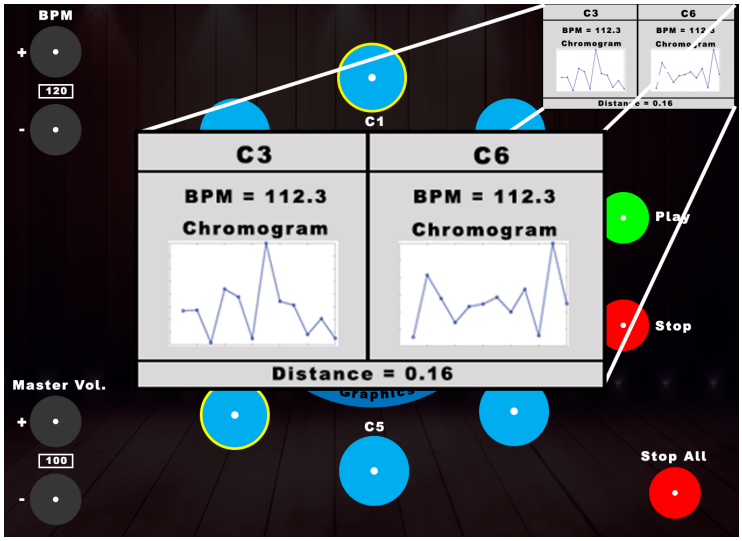
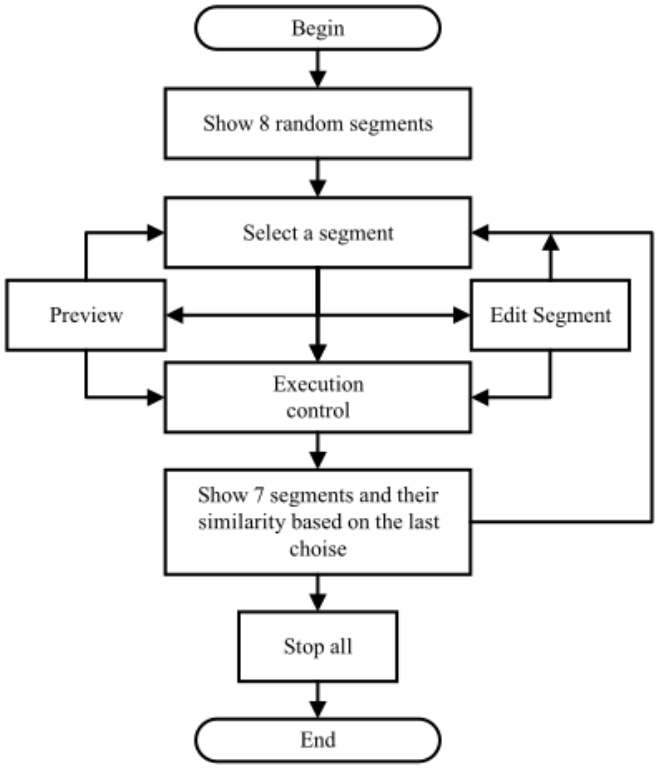


Figure 8 shows the procedure to use the gaze-controlled interface. Initially, the interface shows 8 randomly chosen segments, thus the user can start by choosing a segment, then he/she has options to edit, preview, and control the execution. When the segment is played, the interface shows 7 new segments similar to the last choice and their similarity levels. Afterward, the user can choose the next one and repeat it until the performance is over.

7. Experiments and Results

In this work, a small dataset of instrumental songs was chosen for testing. Thus, the 24 songs, shown in Table 1, were chosen from an instrumental Spotify playlist called Peaceful Piano[7]. The songs were divided into 2623 segments as described in Section 4.

FIGURE 8 – Flowchart with the user's procedure.



[7]https://open.spotify.com/playlist/37i9dQZF1DX4sWSpwq3LiO

TABLE 1 – Song list.

Max Richter - A Catalogue Of Afternoons	Aaron Lansing - Naive Spin
Alexandra Streliski - Changing Winds	Karin Borg - Norrsken
The Daydream Club - One Last Thought	Moux - Chasing Stars
M´elanie Laurent - D´ebut	Oskar Schuster - Maribel
Alexandra Streliski - Plus tˆot	Black Elk - Intro
Rhiannon Bannenberg - For Stormboy	Poppy Ackroyd - Strata
Robert Haigh - Portrait with Shadow	Tedosio - The Book of Jen
The Daydream Club - For the Lost Ones	Moux - Gaze
Peter Sandberg - Remove The Complexities	Phildel - Qi
Jean-Michel Blais - igloo - acoustique	Samuel Lindon - Tallis One
Peter Bradley Adams - Interlude For Piano	Bela Nemeth - This Moment
James Heather - Last Minute Change Of Heart	Poppy Ackroyd - Time

7.1. Experiment I: Harmonic Distance Influence

In order to verify the harmonic distance influence in segments combination, tests were done with high weight to this feature, according to

$$D(\underline{x}_1, \underline{x}_2) = 0.8D_c(\underline{x}_1, \underline{x}_2) + 0.2D_{tempo}(\underline{x}_1, \underline{x}_2). \quad (12)$$

The test, aiming to demonstrate segment compatibility, intends to create a new experimental song with 20 segments in sequence, following the Figure 4 procedure. In this experiment α was empirically defined as approximately 0.1386, thus, the fifth repetition has its distance doubled. The number of repetitions was counted only between the last 5 chosen segments. The results can be found in [https://bit.ly/2GDgKcU].

We also performed tests with HS explained in Subsection 6.2 with 20 segments and $\alpha = 0.1386$. However, the first segment (x_1) was chosen manually. The results can be found in [https://bit.ly/329nM3M].

7.2. Experiment II: Tempo Distance Influence

This experiment aims to observe the influence of the feature tempo distance for compatibility. Thus, tempo distance received high weight adopting the formula

$$D(\underline{x}_1, \underline{x}_2) = 0.2D_c(\underline{x}_1, \underline{x}_2) + 0.8D_{tempo}(\underline{x}_1, \underline{x}_2). \quad (13)$$

As in experiment I, the same tests were performed keeping the same concatenation, penalty parameters, and pivot segment. The results generated by the procedure in Figure 4 can be found in [https://bit.ly/2Gw1XQ9]. In addition, the results following the procedure using HS described in Subsection 6.2 can be seen in [https://bit.ly/2R8C12O].

7.3. Experiment III: Balance Between Harmonic and Tempo Distance

In this experiment, the main target is to verify the balance between harmonic and tempo distance. Therefore, the two features were weighted equally, that is, 0.5 for each one. The new distance formula was defined as

$$D(\underline{x}_1, \underline{x}_2) = 0.5D_c(\underline{x}_1, \underline{x}_2) + 0.5D_{tempo}(\underline{x}_1, \underline{x}_2). \quad (14)$$

This experiment was executed following the same idea used in experiment I with the same pivot segment, however, using the new distance formulation. The results for the tests following the procedure described in 6.1 can be found in [https://bit.ly/2ZtcGTU] and following the procedure represented in 6.2 can be seen in [https://bit.ly/2Zl7KSC].

Figure 9 depicts the convergence curves from experiment III using HS and the distance cost for the automatic concatenation process. Analyzing these curves, it is possible to conclude that the HS algorithm converged for the tests and, in some cases, it presented better results than the automatic process. Table 2 shows the distance results for all tests in which it is noticeable that the HS algorithm results are similar to the results from the automatic procedure described in Figure 4. However, the HS spent 30 minutes on average to give the segment list, showing incompatibility with the real-time interface proposed by this work. In contrast, the procedure shown in Figure 4 spent 1 second on average to give the segment list for each test. It is worth mentioning that the application of HS brings a good comparison base, showing that the procedure in Figure 4 presents satisfactory results since they both have similar values.

FIGURE 9 – Convergence curves of the HS optimization process.

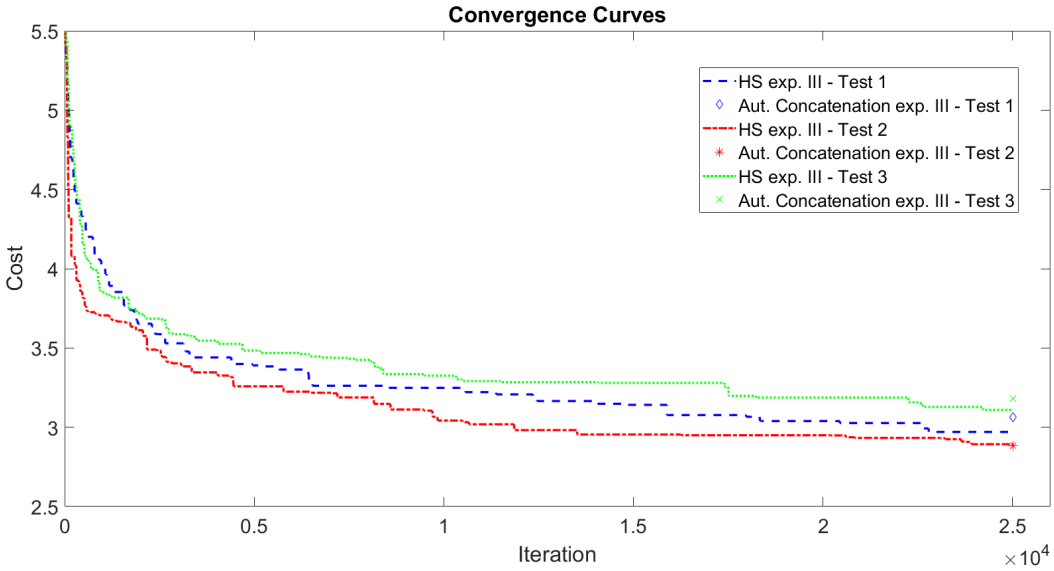


TABLE 2 – Results of HS and automatic concatenation process.

Experiment	Process	Test 1 (a.u)	Test 2 (a.u)	Test 3 (a.u)
I	Aut. concatenation	4,903	4,881	4,924
	HS	4,380	4,612	4,528
II	Aut. concatenation	1,226	1,153	1,225
	HS	1,454	1,193	1,669
III	Aut. concatenation	3,064	2,883	3,181
	HS	2,970	2,893	3,110

Another test was executed aiming to observe the application for user assistance. Thus, from a randomly chosen pivot segment, 15 segments closest to it were listed. Thereafter, the user could create new experimental music by choosing segments from that list. In addition, the user can edit segments to improve musical quality. The results of this test can be found in [https://bit.ly/2VVtSmi].

8. Conclusions

In this paper, we propose music information retrieval techniques for song segmentation and similarity extraction from segments applied to a gaze-controlled musical interface. We focus on the use of the segments and their similarity levels in an assistive interface that allows disabled users, especially non-musicians, to compose with musical segments, increasing the availability of this ubimus system. The system can also be seen as a tool to increase health care and well-being of users,

especially for people with motor limitations to play musical instruments. The chroma and beats per minute features were used to estimate the similarity level. A semi-automatic concatenation procedure was proposed for user performance. In addition, two automatic composition procedures, tested on a dataset containing instrumental songs, were proposed to evaluate the considered MIR techniques, where one of them uses the optimization algorithm harmony search.

The experimental result done by user choices from a list of audio segments is shown in [https://bit.ly/2VVtSmi] and the results using the harmony search algorithm can be seen in [https://bit.ly/2Zl7KSC]. They demonstrated an aesthetic sound analogous to concrete music and mashup. However, there are many challenges to be solved in the interface. For example, we observed that segmentation based on one bar results in short segments, which reduces the user's time to choose the next segment.

As future work, we suggest a study of segmentation based on more than one bar or musical phrase. Also, the two automatic procedures can be used to form larger segments. We also suggest studies related to the musical context in which a segment is removed from the original music, in order to understand combination possibilities. Afterward, the application of experimental tests, using the gaze-controlled interface, done by people with disabilities.

ACKNOWLEDGMENT

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior - Brasil (CAPES)* - Finance Code 001.

REFERENCES

AGRES, K.; SCHAEFER R.; VOLK A.; VAN HOOREN S.; HOLZAPFEL A.; DALLA BELLA S.; MÜLLER M.; DE WITTE M.; HERREMANS D.; RAMIREZ MELENDEZ R.; NEERINCX M.; RUIZ S.; MEREDITH D.; DIMITRIADIS T.; MAGEE WL. *Music, Computing, and Health: A Roadmap for the Current and Future Roles of Music Technology for Health Care and Well-Being*. Music & Science, 4, 1-32, 2021

BENWARD, Bruce.; SAKER, Marilyn. *Music in Theory and Practice volume 1*. New York, NY, USA: McGraw-Hill, 2008.

BROWN, Judith C. *Calculation of a constant Q spectral transform*. The Journal of the Acoustical Society of America, v. 89, n. 1, p. 425–434, 1991.

CAMPOREZ, Higor; FREITAS, Yasmin.; SILVA, Jair; COSTALONGA, Leandro; ROCHA, Helder . Features extraction and segmentation for an assistive musical interface. In: THE 10TH WORKSHOP ON UBIQUITOUS MUSIC, 10, 2020. Ubiquitous Music and Everyday Creativity. Porto Seguro: g-ubimus, 2020. 161-172. Available at: https://doi.org/10.5281/zenodo.4248230

CAMPOREZ, H. A. F. ; NETO,, A.F.; COSTALONGA, L.; ROCHA, H. *Interface Computacional para Controle Musical Utilizando os Movimentos dos Olhos*. Revista Vórtex, 6, 2, 1-17, 2018a. Available at: <http://vortex.unespar.edu.br/camporez_et_al_v6_n2.pdf>.

CAMPOREZ, Higor A. F et al. Olhar Musical: Uma Proposta de Interface para Expressividade Musical Voltada a Indivíduos com Deficiência Motora. In: THE 8TH WORKSHOP ON UBIQUITOUS MUSIC, 8, 2018b, São João del Rei – MG. São João del Rei: UFSJ, 2018. 76–85.

CANNAM C.; BENETOS E.; DAVIES M.E.P.; DIXON S.; LANDONE C.; LEVY M-; MAUCH M-; NOLAND K.; STOWELL D. MIREX 2014: Vamp Plugins from the Centre for Digital Music. In: PROCEEDINGS OF THE MUSIC INFORMATION RETRIEVAL EVALUATION EXCHANGE. 2014

CHOI, Young Mi; SPRIGLE, Stephen H. *Approaches for Evaluating the Usability of Assistive Technology Product Prototypes*. Assistive Technology, 23, 1, 36-41, 2011. Available at: <https://doi.org/10.1080/10400435.2010.541407>.

CORREA, Ana Grasielle Dionisio; DE ASSIS, Gilda Aparecida; NASCIMENTO, Marilena do; FICHEMAN, Irene; LOPES, Roseli de Deus. GenVirtual: An Augmented Reality Musical Game for Cognitive and Motor Rehabilitation. In: VIRTUAL REHABILITATION, 2007. IEEE: 2007. 1-6.

DAVANZO, Nicola; DONDI, Piercarlo; MOSCONI, Mauro; PORTA, Marco. Playing Music with the Eyes through an Isomorphic Interface. In: PROCEEDINGS OF THE WORKSHOP ON COMMUNICATION BY GAZE INTERACTION, 2018, New York. NY, USA: Association for Computing Machinery, 2018. 1-5. Available at: <https://doi.org/10.1145/3206343.3206350>.

DAVIES, M E P; PLUMBLEY, M D. A spectral difference approach to downbeat extraction in musical audio. In: 14TH EUROPEAN SIGNAL PROCESSING CONFERENCE, 14, 2006, Florence - Italy. IEEE, 2006. 1-4.

DORIGO, M; DI CARO, G. Ant colony optimization: a new meta-heuristic. In: THE 1999 CONGRESS ON EVOLUTIONARY COMPUTATION, 1999, Washington - USA. IEEE, 1999. 1470-1477.

FRID, Emma. *Accessible Digital Musical Instruments—A Review of Musical Interfaces in Inclusive Music Practice.* Multimodal Technologies and Interaction, 3, 3, 1-20, 2019. Available at: <https://www.mdpi.com/2414-4088/3/3/57>.

FUTRELLE, Joe; DOWNIE, J. Stephen. *Interdisciplinary Research Issues in Music Information Retrieval*: ISMIR 2000-2002. Journal of New Music Research, 32, 2, 121-131, 2003.

GEEM, Zong Woo. State-of-the-Art in the Structure of Harmony Search Algorithm. In: GEEM,

Zong Woo (Org.). *Recent Adv. Harmon. Search Algorithm*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. 1-10. Available at: <https://doi.org/10.1007/978-3-642-04317-8_1>.

GÓMEZ, Emilia. *Tonal description of polyphonic audio for music content processing*. INFORMS Journal on Computing, 18, 3, 294–304, 2006.

HORNOF, Anthony. The Prospects For Eye-Controlled Musical Performance. In: NIME, 2014, London, United Kingdom. Goldsmiths, University of London, 2014. 461-466. Available at: <http://www.nime.org/proceedings/2014/nime2014_562.pdf>.

HORNOF, Anthony J; SATO, Linda. EyeMusic: Making Music with the Eyes. In: NIME 2004, Singapore. Singapore: National University of Singapore, 2004. 185-188.

HUCKAUF, Anke; URBINA, Mario H. Gazing with pEYEs: Towards a Universal Input for Various Applications. In: ETRA '08, 2008, New York. NY, USA: ACM, 2008. 51-54. Available at: <http://doi.acm.org/10.1145/1344471.1344483>.

KANDPAL, Devansh; KANTAN, Prithvi Ravi; SERAFIN, Stefania. A Gaze-Driven Digital Interface for Musical Expression Based on Real-time Physical Modelling Synthesis. In: 19TH SOUND AND MUSIC COMPUTING CONFERENCE, 2022, France. France: 2022. 456-463.

KELLER, Damián; LAZZARINI, Victor; PIMENTA, Marcelo S (Org.). *Ubiquitous Music*. Springer International Publishing, 2014.

KELLER, Damián. *Challenges for a second decade of ubimus research: knowledge transfer in ubimus activities*. Revista Música Hodie, 18, 1, 148-165, 2018. Available at: <https://revistas.ufg.br/musica/article/view/53578>.

KIM, Juno; SCHIEMER, Greg; NARUSHIMA, Terumi. Oculog: playing with eye movements. In NIME, 2007, NY, USA. NY: ACM, 2007. 50-55.

LARSEN, Jeppe Veirum; OVERHOLT, Dan; MOESLUND, Thomas B. The Prospects of Musical Instruments For People with Physical Disabilities. In: NIME, 2016, Brisbane, Australia. Australia: Queensland Conservatorium Griffith University, 2016. 327-331. Available at: <http://www.nime.org/proceedings/2016/nime2016_paper0064.pdf>.

LINDENBAUM, Ofir; MASKIT, Shai; KUTIEL, Ophir; NAVE, Gideon. Musical features extraction for audio-based search. In: IEEE 26-TH CONVENTION OF ELECTRICAL AND ELECTRONICS ENGINEERS IN ISRAEL, 2010, Israel. Israel: IEEE, 2010. 87-91.

LOURO, V. S.; IKUTA, C. Y.; NASCIMENTO, M. *Música e deficiência: levantamento de adaptações para o fazer musical de pessoas com deficiência*. Arquivos Brasileiros de Paralisia Cerebral, 1, 2, 11–17, 2005.

MAJARANTA, Päivi; BULLING, Andreas. Eye Tracking and Eye-Based Human--Computer Interaction. In: FAIRCLOUGH, Stephen H; GILLEADE, Kiel (Org.). *Adv. Physiol. Comput*. London: Springer London, 2014. 39-65. Available at: <https://doi.org/10.1007/978-1-4471-6392-3_3>.

MCFEE B.; RAFFEL C.; LIANG D.; ELLIS D.P.W.; MCVICAR M.; BATTENBERGK E.; NIETO O. Librosa: Audio and music signal analysis in python. In: THE 14th PYTHON IN SCIENCE CONF, 2015. 2015. 18-25.

MIRJALILI, Seyedali; MIRJALILI, Seyed Mohammad; LEWIS, Andrew. *Grey Wolf Optimizer*. Advances in Engineering Software, 69, 46-61, 2014.

PAYNE, William; PARADISO, Ann; KANE, Shaun K. Cyclops: Designing an Eye-Controlled Instrument for Accessibility and Flexible Use. In: NIME, 2020, Birmingham, United Kingdom. Birmingham, United Kingdom: Birmingham City University, 2020. 576-580.

PISTON, Walter; DEVOTO, Mark. *Harmony*. 5a ed. New York: W. W. Norton & Company, 1987.

THOMPSON, William Forde. *Music, thought, and Feeling: Underst. Psychol. Music*. 2nd ed. New York, NY, US: Oxford University Press, 2015.

VALENCIA, S.; LAMB, D.; WILLIAMS, S; KULKARNI, H.S.; PARADISO, A.; RINGEL MORRIS, M.D. Accessible, Gaze-Operated Musical Expression. In: ASSETS, 2019, New York. NY, USA: Association for Computing Machinery, 2019. 513-515. Available at: <https://doi.org/10.1145/3308561.3354603>.

VAMVAKOUSIS, Zacharias; RAMIREZ, Rafael. Temporal control in the EyeHarp gaze-controlled musical interface. In: NIME, 2012, Ann Arbor. Ann Arbor: University of Michigan, 2012. 11-16.

VAMVAKOUSIS, Zacharias; RAMIREZ, Rafael. *The EyeHarp: A Gaze-Controlled Digital Musical Instrument*. Frontiers in Psychology, 7, 1-14. 2016. Available at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00906>.

VENTER, Gerhard. *Review of Optimization Techniques*. Encyclopedia of Aerospace Engineering: John Wiley & Sons, 2010. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470686652.eae495>.

VICKERS, Stephen; ISTANCE, Howell; SMALLEY, Matthew. EyeGuitar: Making Rhythm Based Music Video Games Accessible Using Only Eye Movements. In: THE 7TH INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTER ENTERTAINMENT TECHNOLOGY, 7, 2010. ACM: 2010. 36-39.

WHO, (WORLD HEALTH ORGANIZATION). Relatório mundial sobre a deficiência, 2012. Available at: <http://apps.who.int/iris/bitstream/handle/10665/70670/WHO_NMH_VIP_11.01_por.pdf?sequence=9>. Acesso em: 1 jun. 2018.

YOO, Jae-Chern; HAN, Tae Hee. *Fast normalized cross-correlation*. Circuits, systems and signal processing, 28, 6, 819–843, 2009.

ABOUT THE AUTHORS

Higor A. F. Camporez was born in Espírito Santo, Brazil. He received his B.S. degree in computer engineering in 2018 and his M.S. degree in electrical engineering in 2020 both from the Federal University of Espírito Santo (UFES), Espírito Santo, Brazil. He is a PhD student in electrical engineering at UFES since 2020. He is a member of the NESCoM research group that carries on Computer Music related research, especially on Ubiquitous Music. His research interest includes Ubimus, robotic musicians, optimization, artificial intelligence and telecommunication systems. ORCID:

https://orcid.org/0000-0003-2197-8588. E-mail: higorcamporez@gmail.com

Yasmin M. de Freitas is a master student of the Postgraduate Program in Art and New Media (PPGA - UFES). Graduated in Music Degree at the Federal University of Espírito Santo (UFES) and Member of the Sound Experimentation Group (GEXS) with partnership with the Spirit-Santense Core of Musical Computing (NESCOM). ORCID: https://orcid.org/0009-0002-8842-6948. E-mail: yasmarquesf@gmail.com

Jair A. L. Silva received his BS, MS, and PhD degrees in electrical engineering from the Federal University of Espŕito Santo (UFES), Vitória, Brazil, in 2003, 2006, and 2011, respectively. In 2012, he joined the Department of Electrical Engineering of UFES. His research interests include optical fiber communication, orthogonal frequency division multiplexing (OFDM), and passive optical network (PON). ORCID: https://orcid.org/0000-0003-2567-184X. E-mail: jair.silva@ufes.br

Leandro Costalonga has a Computer Science Degree with Masters (UFRGS/Brazil) and PhD (University of Plymouth/UK) in Computer Music. Associate professor at the Federal University of Espírito Santo (UFES/Brazil) where teaches on undergraduate programs in Computer Science and Computer Engineering and Graduate Program in Arts. Head of the NESCoM Research Group that carries on Computer Music related research, especially on Ubiquitous Music. Besides Computer Music, other research interest includes Human-Computer Interaction, Programming Languages and Artificial Intelligence. ORCID: https://orcid.org/0000-0002-0252-9624. E-mail: leandro.costalonga@ufes.br

Helder R. O. Rocha was born in Santo Antao, Cabo Verde. He received his B.S. degree in electrical engineering in 2002 and his M.S. and D.S. degrees in computing science in 2005 and 2010 from the Federal University of Fluminense (UFF), Rio de Janeiro, Brazil. In 2013, he joined the Department of Computing and Electronic of UFES and in 2018 the Department of Electrical Engineering of UFES. His research interest includes optimization, artificial intelligence in smart grids and telecommunication systems. ORCID: https://orcid.org/0000-0001-6215-664X. E-mail: helder.rocha@ufes.br