Sibilim: A low-cost customizable wireless musical interface

Helena de Souza Nunes
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brazil
helena.souza.nunes@ufrgs.br

Lydia Wohl Coelho
Porto Alegre, Brazil
lydiacoelho@gmail.com

Federico Visi
www.federicovisi.com
Hamburg, Germany
mail@federicovisi.com

Rodrigo Schramm
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brazil
rschramm@ufrgs.br

ABSTRACT

This paper presents the Sibilim, a low-cost musical interface composed of a resonance box made of cardboard containing customised push buttons that interact with a smartphone through its video camera. Each button can be mapped to a set of MIDI notes or control parameters. The sound is generated through synthesis or sample playback and can be amplified with the help of a transducer, which excites the resonance box. An essential contribution of this interface is the possibility of reconfiguration of the buttons layout without the need to hard rewire the system since it uses only the smartphone built-in camera. This features allow for quick instrument customisation for different use cases, such as low cost projects for schools or instrument building workshops. Our case study used the Sibilim for music education, where it was designed to develop the conscious of music perception and to stimulate creativity through exercises of short tonal musical compositions. We conducted a study with a group of twelve participants in an experimental workshop to verify its validity.

Author Keywords

cardboard instrument, instrument design, reconfigurable interface

CCS Concepts

•Human-centered computing → Interface design prototyping; •Applied computing → Sound and music computing; •Computing methodologies → *Image processing;*

1. INTRODUCTION

Mobile devices such as smartphones and tablets have been used for music applications for several years now. Music software developed for these devices include production tools that imitate and extend the functionalities of established desktop software (e.g. GarageBand[1] and other digital audio workstations), and emulations of hardware equipment (such as Korg's iMS-20[2]). In addition to emulating existing concepts, mobile computing has enabled a distinctive range of musical interactions that take advantage of the characteristic features of mobile devices, such as multi-touch screens and accelerometers. A notable example is "Ocarina" [13] in which the gesture vocabulary of flute-like instruments is reinterpreted by allowing the user to play the mobile device as if it was a wind instrument. More recently, custom augmentations for mobile devices such as passive amplification of the built-in speakers and alternative grips [4] have been proposed with the purpose of extending instrument design possibilities.

Mobile device augmentation using low cost components such as common cardboard has been used also for purposes other than music. Google Cardboard[3] is intended as an affordable way to experience virtual reality using a standard smartphone housed in a cardboard fold-out viewer equipped with plastic lenses. Another example of cardboard extensions designed for a mobile device are the Nintendo Labo[4] kits for the Nintendo Switch mobile videogaming console. These cardboards cut-outs are to be assembled in combination with the console display and controllers (the Joy-Con, which features several motion sensing technologies [12]). These kits were designed also for educational purposes, as they can also be used—along with the console and the accompanying software—to teach basic principles of physics, music, and other subjects. Notably, one of the kits consists in a working cardboard toy piano that holds the console display as a music stand.

The properties of cardboard have also been exploited to amplify sound signals. In projects such as the "Cardboard Speaker-Table"[5] cardboard is paired with transducers to acoustically amplify the audio signals coming from a mobile device. Small transducers have been used extensively in the sonic arts to transmit vibrations to objects of various shapes and materials with the purpose of creating alternative ways of producing and experiencing sound. For example, Attila

[1] `https://www.apple.com/lae/ios/garageband/`
[2] `https://www.korg.com/us/products/software/ims_20_for_ipad/`
[3] `https://vr.google.com/cardboard/`
[4] `https://labo.nintendo.com`
[5] `https://lh-hantrakul.com/2014/04/13/cardboard-speaker-table-mkii/`

NIME'19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

Faravelli's "Aural Tool #1: Trifoglio"[6] consists of transducers applied on a purposely designed object made of curved sheets. By playing back dedicated compositions loaded on a built-in digital player, listeners can experience the music through an omnidirectional sound field and acoustic vibrations felt by holding the object in their hands.

The instrument presented in this paper, the Sibilim, was designed to be a low-cost project that leverages on the possibilities afforded by the ubiquity mobile devices and the properties of cardboard. Sibilim acts as a smartphone cardboard resonance box. The smartphone video camera tracks the movement of push-buttons used to play chords and scales. The buttons can be mapped to a specific set of sounds or MIDI parameter and can be easily reconfigured without the need of hard rewiring the system. The sound of the instrument is generated by synthesis or sample playback, and can be amplified by means of a transducer placed on the cardboard panels. Sibilim is targeted at students learning basic music concepts and aims at encouraging the exploration of instrument design and customisation thanks to the simplicity and affordability of its components.

2. INTERFACE DESIGN

This section describes the material composition and the system architecture of Sibilim. The musical instrument has two main parts: a cardboard box and a smartphone device. The physical interaction with the instrument is made mainly through push buttons that work similarly to keyboard keys, switching a musical note to the *on* state when pressed, and to the *off* state when released. Nonetheless, the push buttons can also be mapped to several control parameters, allowing many sound possibilities. The smartphone is attached to the cardboard box and can receive input parameters also through its multitouch screen. A dedicated sound synthesis and sample playback application is used to generate sound.

Sibilin was designed so that assembling it is very simple, and fun for the user. Thus, the entire instrument body is made of cardboard, push buttons built of wood with metal springs, and an acrylic mirror. There is no need for hard wires in this instrument since the proposed interface uses optical information to detect the buttons activation. However, optionally, it is possible to connect an active transducer to the smartphone and produce a much louder sound thanks to the vibrations amplified by the cardboard panels. Figure 1 presents a prototype version of Sibilin.

2.1 Interface Architecture

The schematic of the system operation, is presented in Figure 2. The instrument body is a cardboard box (1), with several push buttons on the top (2) connected to markers inside the box (3). The image of each marker is reflected by the mirror (4) through a pinhole (5) and captured by the smartphone camera (6) placed on top of the box. The dedicated mobile app uses computer vision algorithms to track the position of all the markers[11].

The system architecture permits many push-button layouts, allowing rapid creation and customisation of new musical interfaces. The only limitation is that the marker of each push-button must be visible within the smartphone camera field of view.

2.1.1 System calibration

In order to start using the instrument, the user has to lay the smartphone over the box so that the camera is placed

[6] `http://shop.plattfon.ch/ATTILA-FARAVELLI-aural-tool-1-trifoglio`

Figure 1: Cardboard prototype of *Sibilim.*

on the small aperture located on the box cover. The first time it is used, the system runs a learning procedure to detect and store each marker position. Moreover, at this stage, the user can map each marker to a specific control parameter or musical note (using MIDI specifications[3]). After the learning stage, for each new system initialization, the marker positions and mappings are restored, and the vision algorithm performs an automatic calibration to fix and compensate any deviation in the placement of the smartphone.

2.1.2 Outliers removal

Ideally, the vision system should detect exactly the same number of markers that are present inside the box. However, shadows of markers or even dust in the mirror might cause spurious detection. The vision system uses several features extracted from the image of each marker to remove these outliers, such as colour, shape, size, convexity, and neighbourhood connectivity. These features are incorporated to a set of heuristics during the learning stage. Later, during the execution time, these heuristics are used to filter any marker detection that are likely to be a false positive.

2.1.3 Mapping of Controls

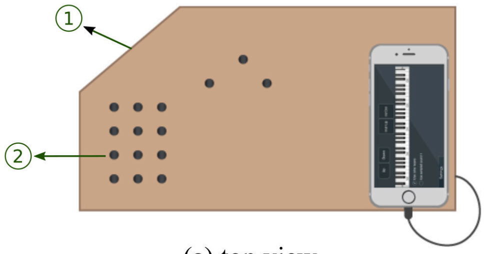
The discrete range of possible locations of each marker can be mapped to control parameters for sound generation and transformation. There is also the option for exclusively two state conditions. This option can be applied as an onset/offset concept, where each marker can be in state ON (attack phase of the ADSR envelope) or OFF (release phase of the ADSR envelope). Figure 3 illustrates the exclusively two state condition, which is implemented with the help of a fixed threshold. When a marker moves far enough from the resting position, crossing the threshold, the system changes the state from OFF to ON. When the user release the push button, the marker returns to the resting position, changing inversely the state from ON to OFF.

To simulate MIDI velocity, the Sibilim model computes the time difference obtained between the transition of marker positions from the OFF to the ON states. The velocity ΔV in pixels is then converted to sound amplitude A in the range $[0, 1]$ by the equation 1:

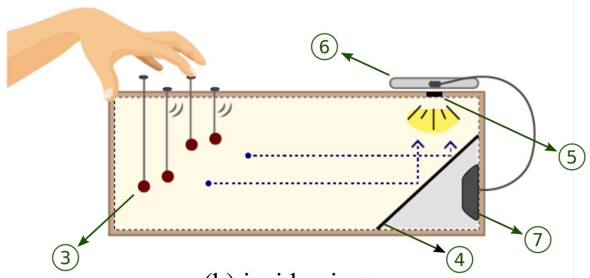
$$A = \frac{1}{\max\{\alpha \Delta V, 1\}}, \tag{1}$$

where α is a tuning parameter that depends on the layout of the instrument, size and distance of markers.

2.1.4 Computational feasibility



(a) top view



(b) inside view

Figure 2: Schematic of the optical push-button system.

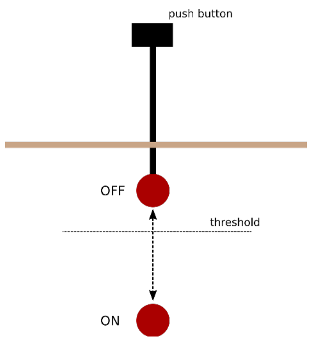


Figure 3: ON/OFF states model.

The vision system configuration implemented in this model captures 60 video frames per second. Despite this frame rate give us considerable latency (16.6ms) in the input of the system, we have empirically note that the push-buttons in format of pistons act masking, in some degree, part of the human perception of this latency. The computer vision algorithm can be configured to use lower frame rates in order to reduce computation cost. This is an important feature since a main concern about the use of smartphones is the durability of batteries. In section 4, we present an experiment designed to quantify the balance between the perceived latency and the computational need of the system.

2.1.5 *Sound generation*

Modern smartphones admits the use of several audio libraries and frameworks for implementation of techniques for sound synthesis and sample playback [9, 10, 5, 7]. The sound generation module of the Sibilim instrument uses the

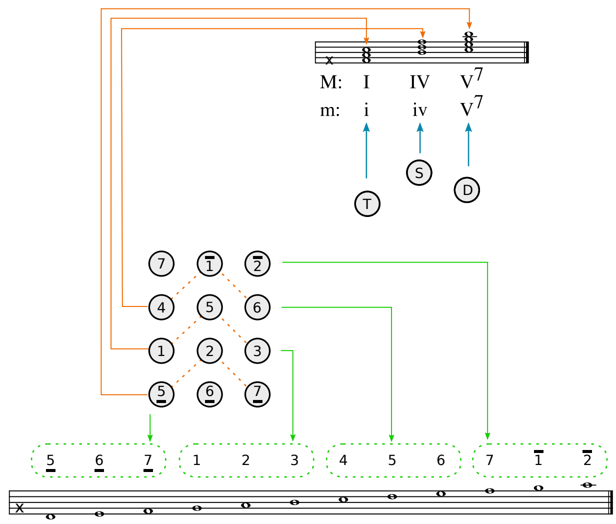


Figure 4: Tonal layout of push-buttons.

JUCE framework[7], which promotes rapid implementation of audio synthesis and sample-playback routines, and can deploy the Sibilim app to multiple platforms.

2.1.6 *Audio Amplification*

Optionally, the proposed instrument can be actively amplified to make it sound louder. In our model, the audio from the smartphone is amplified by a low voltage audio power amplifier (LM386 based module[8] with 9V battery), which drives the signal to a transducer (exciter speaker) attached to one internal side of the box. Figure 2-(7) shows this optional audio connection.

2.1.7 *Musical controls*

The app implements musical controls commonly found in commercial electronic keyboards, as octave and key shift, as well as instrument timbre selection. All these functions can be accessed through the smartphone touchscreen.

3. USE CASE: EDUCATIONAL DESIGN

Sibilim allows quick customisation for different uses. Here, we present the outcomes of the first use of Sibilim for music education. In this use case, the goal was to aid the study of harmonic perception and melodic sight-singing, based on the Musicalization of Adults through the Voice (MAaV) method [2]. The MAaV method is a musicalization approach that starts from the Tonal System and the Relative Solfege of the Scale Degrees. In this method, beginners prctice melodies with the extension of an octave and a half (from $\underline{5}$ to $\overline{2}$ degrees), and their sight-singing becomes accompanied by Tonic and Subdominant triads (I or i, and IV or iv, respectively) and Dominant seventh chord (V7) over the diatonic scales, in all major and minor-harmonic scales.

3.1 Buttons layout

Based on the beginners level of MAaV method, the Sibilim offers an specific layout for the push-buttons, defining, as a consequence, the instrument sound possibilities. The distance between the buttons was based on the measurement between the keys of a keyboard of general acoustic piano (2.2 cm); and the tension of the springs was determined, empirically. In the latter case, several users were asked to indicate the spring that brought them greater comfort for

[7] http://juce.com
[8] http://www.ti.com/product/LM386

the fingers. The researchers adopted the spring that received the highest positive score. All these measures were taken to ensure healthy ergonomic conditions for users.

As illustrated by gray circles in the Figure 4, the are two groups of buttons in this instrument architecture. The group formed by twelve buttons corresponds to the melodic notes and the group formed by only thee buttons are designed to activate chords (harmonic function).

3.1.1 Sound Production

The instrument offers two types of sounds: single notes, when the melodic buttons are pressed; and simultaneous pitches, in preset chords, when the harmonic buttons are pressed.

By pressing the melody buttons in sequence, from right to left and from bottom to top, the user listens to the interval relations of the major or minor harmonic scales (hence already with the seventh degree changed), starting at the fifth degree of the corresponding tonality. This sequence of notes contains twelve diatonic sounds, which corresponds to one and a half octaves, which is the range of many popular songbook melodies as well as the range that a person with little singing practice can emit effortlessly [2]. To extend the range of her/his voice, or to find a more comfortable *tessitura*, the user can transpose the selected part by choosing a new key in the App visual interface.

According to the mode and the tonality of each song or sight-singing exercise, the user can bind the exact pitches of the melody and chords buttons together. The user accesses the twelve buttons of the Sibilim's melodic keyboard with her/his left hand, and the three harmonic buttons with her/his right hand. In both cases, she/he uses only the ring, middle and index finger, decreasing technical skill requirements for the hands.

It is worth noting that the relative positions of the buttons on the keyboard have direct influence on the camera view since the image produced from the markers is reflected by the mirror and need to be fully captured by the camera of the smartphone. Several attempts have led the researchers to some relevant decisions about the locations of the markers, regarding the distance between them and the mirror, and the ergonomics of the keyboard.

3.2 Tonal approach

The tonality is structured, basically, on the tension versus rest relationship, established between the Dominant Seventh chords and the Tonic chords [1]. The Seventh chords of the Dominant are the same for both major tone and its minor harmonic homonym; the chords of three sounds built on the Tonic can be Major or Minor. The same happens with the Subdominant chords, which can either lead from rest to tension, or from tension to rest. Thus, by learning to play and listen to these three chords by placing them correctly on a melodic design, the user has an aunderstanfing of the tonal centre is able to follow simple tonal songs.

The idea of accompanied melody, which was born with the Continuous Bass (c. XVI) and evolved until the Lieder (c. XIX), is still widely employed in contemporary western popular music [1]. Our instrument prototype is built to explore this concept. The user can use the set of melodic buttons to learn a new song or to check an intonation pattern. Moreover, Sibilim permits an alternative format of musical execution, in which the melodic buttons activate bass notes in counterpoint with the melody that the singer is intoning. In this case, the performance may become more rewarding and rich if the users, with their right hand, activates chords through the harmonic buttons.

In summary, the user can use the same fingerings as well as the same set of buttons, to perform melodic lines and chords with different interval relationships and absolute pitches, as direct consequence of the previously selected scale in the App. These features allow the user to sing and play an accompaniment during the performance of a simple tonal repertoire. With this configuration, learners can sing and play with varied instrumental technical execution schemes, at a comfortable pitch for their voice and without the need of transposing key by changing fingerings. These features make the Sibilin a fun and easy instrument to play, both when practicing alone or with other musicians.

Video documentation and other supplementary materials can be found through the link `https://inf.ufrgs.br/~rschramm/projects/music/nime2019`.

4. EXPERIMENT

A group of twelve adults participated in an experimental workshop to verify the validity of the proposed musical interface. On January 25, 2019, between 14h and 16h, in the Auditorium Prof. Castilho of the Institute of Informatics of UFRGS.

Among the participants, 50% were women and 50% were men, whose ages range from 20 to 48, and professional characteristics were very different from each other. However, all of them had some connection with the process of creation and the utility of this new instrument: three music teachers of elementary school, one amateur musician, one electrical engineer, two software developers, one software engineer, three fellows without knowledge of music nor computer science, one production assistant with special cognitive needs.

The general goal of this experiment was to evaluate aspects related to the usability of Sibilim in the context of the MAaV method. The specific objectives were: to describe the reaction of the test group to the Sibilim; to get a set of user evaluations; to pair considerations and suggestions made by musicians and non-musicians; and to decide about any changes needed in the instrument's original design. The experiment was conducted using two prototypes of Sibilim. One instrument was used with an iPhone 8 and the other was used with an iPhone 6s.

4.1 Protocol

The scheduled meeting for data collection began with the presentation of the participants, most of whom did not know each other. After that initial moment, for about ten minutes, the researchers presented the instrument, talking about its history and construction aspects. They then organised the participants into two groups, according to their previous musical experiences. Each of the two groups received a prototype of the Sibilim. For about thirty minutes, participants were asked to explore it together spontaneously. After this free moment of exploration, the researchers conducted some musical activities, including melodic and harmonic elements, which employed the technical resources available in the instrument. While the activities took place, the researchers explained the instrument features and filled the participants' questionnaire. In the end, the twelve participants answered the questionnaire and collaborated with ideas and suggestions, in an open semi-structured debate for data collection.

In order to employ a consistent methodology for our experiments, we have used the DTA Discursive Textual Analysis [6, 8] technique in distinct phases of this study. DTA is an analytical tool for qualitative analysis of textual and discursive information widely used in education research. This protocol has three main steps: Unitarization, Categorization and Production of Metatext.

Table 1: Questionnaire / Answers

Assign points from 1 (minimally significant / poor) to 5 (very relevant / ideal)				
Evaluated Aspect		Details	Obtained Score	
			Avg	Std
Surprise factor	Impact on yourself		4.6	0.7
	School Impact Expectation		4.3	0.9
	Trade Impact Expectation		3.9	0.7
	Name		3.6	1.4
Ergonomics	Comfort on the touch		3.9	0.8
	Spring tension	Spring 1	4.2	1.0
		Spring 2	3.3	1.0
	Melodic button's layout		4.3	0.9
	Chord button's layout		4.5	0.9
System	Initialization		3.6	0.9
	Smartphone setting		3.2	1.3
	Number of features		3.7	1.1
	Quality of the timbres		3.6	0.7
Musical features	Sensitivity to dynamics		2.6	1.3
	Sound Intensity		4.2	0.8
	Latency	30fps	2.25	1.50
		40fps	2.33	1.52
		50fps	3.00	1.41
		60fps	4.20	0.83
Interface	Functionalities visualization		3.8	1.1
	Appearance of the box		4.0	0.7
	Friendly design		4.2	0.4

4.2 Discussion

The data collection instruments (questionnaire and semi-structured interview script) and the answers from all participants can be seen in the Table 1. This table shows the average and standard deviation of answers in the format of 5-point Likert-scale statements. In the most cases, the standard deviation is lower than 1.0, indicating a strong agreement among the participants.

All participants were enthusiastic about the simplicity of the instrument's construction and handling. They also stated their need for the autonomous, alone and independent assessment of the learning of harmonic perception and melodic musical reading. While referring to distinct aspects, they all stated that they felt good by exploring the possibilities of the instrument and trying to understand under what circumstances it could be used as a proper musical instrument. They used words such as curiosity, freedom, interest, fun, joy, motivation and even gratitude to express what they felt during the experiment. They also reported a general excitement about the instrument as all the participants all seemed anxious to play. A third of the participants stated that they had never seen any other similar instrument. Participant P1 has identified the possibility that each box could be customised by the user, using stickers and colours. P1 also pointed out the possibility of integrating languages and artistic expressions, supporting new creative ways and improving the educational process. Other manifestations that complement the characterisation of the type of response awakened by the Sibilim are: a desire to be part of a Sibilim orchestra; the willingness to create tutorials for their musical and multimedia work; and an impulse to spread it among their students.

The suggestions referred to the aspects of the physical construction of the instrument, the technological resources used, and the procedures pertinent to its use and handling. On the material construction of the Sibilim, the users suggested greater precision in the mechanisms of control and activation of the sounds. They also recommended enhancement of dynamics and enrichment of timber files. It could be a good idea to make a lateral opening in the box, so that the movements of the markers, in its interior, can be appreciated by the audience. In the same way, they recommended insertion of a latch, fold or double-sided adhesive tape to hold the phone to the carton. Regarding its technological requirement, the opinion was that the quality required for the mobile processor could limit its reach to the target audience.

The participants identified that the use of Sibilim brings the idea of Continuous Bass, implying reading and practising melodies and harmony. Thus, the effective use of the instrument requires knowledge of musical theory, besides notions of technology and motor skills. For example, one of the participants stated that they had not been able to test all the features because they did not know how to do it, suggesting that they might be excessive. Others said it took a long time to recognise all the possibilities. Reacting to these observations, the idea of opening a space for debates and exhibitions, associated with the distribution of the instrument, was representative. Some participants suggested a broadening of the concept about the target audience, not only targeting singers and adults; they also thought about children. Others also suggested the provision of training courses for teachers and educational videos on the Internet. Finally, they indicated the creation of a site for postings sent by users, to exchange experiences about their possibilities of use.

5. CONCLUSION

This paper presented a new musical interface named Sibilim. The mechanical and physical requirements for the construction of the Sibilim are elementary. It consists of a smartphone positioned on a cardboard box with a mirror and a transducer inside. On the upper face of this box, there are push-buttons attached to sticks and regulated by springs. These rods extend into the cardboard box and at the other end have markers. The smartphone capture and track the movement of these markers through its video camera, and a sample-playback algorithm generates sounds in response to such movement. The instrument layout includes twelve single note push-buttons for the left hand and three harmonic (chord) push-buttons for the right hand.

The design of this interface was also planned to follow the educational approach proposed by the MAaV method [2]. This method addresses the Tonal Music System and focuses on the Majors and Minors Harmonics modes, and the Relative sight-singing of the Scale Degrees. The teaching procedures of MAaV are based on the practice of songs accompanied by some harmonic instrument, and employ chords of the Dominant Seventh Chord (V7), Tonic (I and i) and Subdominant (IV and iv).

An experimental evaluation was conducted with twelve people representing different professions, ages and musical knowledge. This study evaluated the proposed interface regarding the feasibility of use and acceptance, as well as specific aspects: surprise factor, ergonomics, system configuration, musical features, and interface. All the participants in the study were enthusiastic and motivated to use it. Feedback from these participants was registered using a 5-point Likert-scale. The final report indicates that the prototype presents good controllability. However, there are still many challenges that should be addressed in order to improve the Sibilim and enhance user experience. Based on the answers from the twelve participants of this experiment, it is clear that the system initialisation must be improved, facilitating the smartphone configuration at the instrument initialisation stage. The main dissatisfaction during the experimental use of the prototype is related to the sensitivity of the dynamics. This means that the estimated MIDI velocity through Equation 1 is not accurate enough. Since the lack of controllability over the dynamics has direct influence on the ADSR envelope, we believe that complaints about the instrument timbre might be correlated. Latency is a critical factor in any digital music instrument, and our prototype has shown satisfactory latency when using the camera frame rate set up to 60fps. This is an important result since now we know the proposed optical interface is feasible in modern smartphones. However, as we increase the computational power, the consumption of the battery also increase. Thus, still the need for more efficient vision algorithms to make the batteries lasting longer.

As future work, we plan to focus our efforts on improving the control of dynamics and reducing the computational complexity of the algorithms used in the Sibilim interface.

6. REFERENCES

[1] J. Burkholder and D. Grout. *A History of Western Music: Ninth International Student Edition*. W. W. Norton, 2014.

[2] H. de Souza Nunes. *Musicalização de Adultos através da Voz - livro do professor*. UFRGS/CAEF, Porto Alegre, 2005.

[3] D. M. Huber. *The MIDI Manual*. Sams, Indianapolis, IN, USA, 1991.

[4] R. Michon, J. O. Smith, M. Wright, C. Chafe, J. Granzow, and G. Wang. Passively Augmenting Mobile Devices Towards Hybrid Musical Instrument Design. In *NIME 2017 Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 19–24, 2017.

[5] R. Michon, J. Smith III, and Y. Orlarey. Mobilefaust: a set of tools to make musical mobile applications with the faust programming language. In E. Berdahl and J. Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 396–399, Baton Rouge, Louisiana, USA, May 2015. Louisiana State University.

[6] R. Moraes and M. do Carmo Galiazzi. *Análise textual: discursiva*. UNIJUI, 2007.

[7] M. Mulshine and J. Snyder. Oops: An audio synthesis library in c for embedded (and other) applications. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 460–463, Copenhagen, Denmark, 2017. Aalborg University Copenhagen.

[8] A. Pedruzzi, E. Schmidt, M. d. C. Galiazzi, and T. Podewils. Análise textual discursiva: os movimentos da metodologia de pesquisa. *Atos de Pesquisa em Educação*, 10:584, 09 2015.

[9] M. Puckette. Pure data: another integrated computer music environment. In *in Proceedings, International Computer Music Conference*, pages 37–41, 1996.

[10] M. Robinson. *Getting Started with Juce*. Packt Publishing, Birbmingham, 2013.

[11] S. Suzuki and K. Abe. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.

[12] F. Visi and F. Faasch. Motion Controllers, Sound, and Music in Video Games: State of the Art and Research Perspectives. In D. Williams and N. Lee, editors, *Emotion in Video Game Soundtracking*. Springer International Publishing, 1 edition, 2018.

[13] G. Wang. Ocarina: Designing the iphone's magic flute. *Computer Music Journal*, 38:8–21, 07 2014.