UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Mecânica

YU TZU WU

Visualization for Han Character Learning and Musical Experience

Visualização em Aprendizado de Caracteres Han e Experiência Musical

CAMPINAS
2024

YU TZU WU

Visualization for Han Character Learning and Musical Experience

Visualização em Aprendizado de Caracteres Han e Experiência Musical

Thesis presented to the School of Mechanical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Mechanical Engineering, in the area of Mechatronics.

Tese apresentada à Faculdade de Engenharia Mecânica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutora em Engenharia Mecânica, na Área de Mecatrônica.

Orientador: Prof. Dr. Eric Fujiwara

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA YU TZU WU E ORIENTADA PELO PROF. DR. ERIC FUJIWARA.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca da Área de Engenharia e Arquitetura
Elizangela Aparecida dos Santos Souza - CRB 8/8098

W95v

Wu, Yu Tzu, 1992-
Visualization for Han character learning and musical experience / Yu Tzu Wu. – Campinas, SP : [s.n.], 2024.

Orientador: Eric Fujiwara.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Mecânica.

1. Visualização. 2. Análise de imagem. 3. Interação humano-máquina. 4. Tecnologia educacional. 5. Arte e música. I. Fujiwara, Eric, 1985-. II. Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Mecânica. III. Título.

Informações complementares

Título em outro idioma: Visualização em aprendizado de caracteres Han e experiência musical
Palavras-chave em inglês:
Visualization
Image analysis
Human-machine interaction
Educational technology
Art and music
Área de concentração: Mecatrônica
Titulação: Doutora em Engenharia Mecânica
Banca examinadora:
Eric Fujiwara [Orientador]
Artemis Maria Francelin Sanchez Moroni
Carlos Kenichi Suzuki
Ely Carneiro de Paiva
Antônio Ribas Neto
Data de defesa: 19-12-2024
Programa de Pós-Graduação: Engenharia Mecânica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-2623-2526
- Currículo Lattes do autor: http://lattes.cnpq.br/3068238637672696

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA

TESE DE DOUTORADO ACADÊMICO

Visualization for Han Character Learning and Musical Experience

Visualização em Aprendizado de Caracteres Han e Experiência Musical

Autor: Yu Tzu Wu

Orientador: Prof. Dr. Eric Fujiwara

A Banca Examinadora composta pelos membros abaixo aprovou esta Tese:

Prof. Dr. Eric Fujiwara, Presidente
Faculdade de Engenharia Mecânica, Unicamp

Profa. Dra. Artemis Maria Francelin Sanchez Moroni
Centro de Tecnologia da Informação Renato Archer

Prof. Dr. Carlos Kenichi Suzuki
Faculdade de Engenharia Mecânica, Unicamp

Prof. Dr. Ely Carneiro de Paiva
Faculdade de Engenharia Mecânica, Unicamp

Prof. Dr. Antônio Ribas Neto
Instituto Federal Catarinense

A Ata de Defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 19 de dezembro de 2024.

ACKNOWLEDGEMENTS

This study was financed in part by The Brazilian National Council for Scientific and Technological Development (CNPq), grant #142580/2018-0.

RESUMO

Os avanços da tecnologia multimodal abriram possibilidades para experiências sinestésicas em diferentes áreas de conhecimento. Neste trabalho, duas ferramentas foram propostas para auxiliar o aprendizado de caracteres Han, também conhecidos como hanzi, a estrutura fundamental da escrita da língua chinesa. A primeira ferramenta combina a realidade aumentada e os princípios de mnemônica para ajudar na memorização do hanzi, com a representação gráfica do objeto 3D e da pronúncia em chinês. O software foi demonstrado para cinco hanzi, mas atestou-se que ele consegue distinguir caracteres semelhantes e identificar aqueles que não estão inclusos na sua biblioteca. Além disso, ele oferece a pronúncia em japonês dos caracteres comuns às duas línguas, comprovando que pode ser adaptado para demais idiomas. A segunda aplicação explora os componentes morfológicos do hanzi e marca automaticamente o bloco chamado bùshǒu (radical de indexação) dentro do caractere por meio de um identificador constituído por uma rede neural convolucional de duas camadas. A importância dessa aplicação reside no fato de que os nativos confiam na bùshǒu para assimilar, distinguir e inferir o sentido dos hanzi. O software considerou o caractere como uma imagem, destituído do teor semântico como veículo de linguagem, e alcançou uma taxa de ~90,0% e ~76,1% para classificar as amostras de validação e de teste genérico conforme a bùshǒu, respectivamente, e conseguiu localizar o bloco da radical de indexação com precisão aceitável. Além do aprendizado de segundo idioma com a ajuda da tecnologia multimodal, este trabalho também explorou a possibilidade de vivenciar visualmente a música clássica. A abordagem focou em identificar o fraseado da peça de entrada, em especial a dinâmica, e utilizar a impressão auditiva para modular um texto usando diferentes tipografias. Dessa forma, buscou-se transferir a experiência do ouvinte a um texto escrito, oferecendo uma experiência semelhante aos usuários com deficiência auditiva. Por último, ainda no campo da música, este trabalho propôs uma interface de produção musical baseada em fibra óptica. O protótipo da interface é composto por uma fibra óptica multimodo apoiada sobre uma superfície rígida e com espumas uniformes distribuídas ao longo da sua extensão. As espumas simulam as teclas de um piano e, conforme o usuário as aperta, gera oscilações no padrão de espectrograma do laser que atravessa a fibra e é capturado por uma câmera CCD. Com a calibração apropriada das teclas e do padrão do espectrograma, é possível recuperar qual tecla o usuário pressionou e reproduzir a nota correspondente em tempo real no computador. O protótipo foi demonstrado

para a escala de dó maior, podendo ser adaptado para outras escalas, ou até mesmo simulando outros instrumentos.

Palavras-chave: Caracteres Han; aprendizado de idioma asiático; tecnologia educacional; visualização de música; projeto de interação homem-computador; realidade aumentada; classificação de imagens; reconhecimento em imagens; visualização de informação; processamento de som e música.

ABSTRACT

The advancements in multimodal technology allowed synesthetic experiences to be possible in several fields of knowledge. This work proposed two tools to aid the learning process of Han characters, also known as hanzi, the fundamental structure of the Chinese writing system. The first tool combines augmented reality and mnemonic principles to help the user memorize hanzi, providing the graphical representation of the 3D object and the Chinese pronunciation. The software was demonstrated for five hanzi, but it distinguished similar characters and unknown instances outside the bag of words. Moreover, it provided the Japanese pronunciation of the samples shared by both languages, proving to be extendable to other languages. The second tool explores the morphological components of hanzi and automatically marks the block known as bùshǒu (indexing radical) within the character using an identifier characterized by a two-layered convolutional neural network. The importance of this application builds on the fact that native speakers trust in the bùshǒu to assimilate, differentiate, and infer the meaning of hanzi. The software considered the characters as a simple image, without the inherent meaning as a communication media, and reached a rate of ~90.0% and ~76.1% to classify the validation and the general test samples according to their bùshǒu, respectively, and localized the indexing radical with acceptable precision. Besides second language learning with the aid of multimodal technology, this work also investigated the possibility of visually experiencing classical music. The adopted approach identified the phrasing of the input piece, especially its dynamics, and modulated a text using different typographies according to the aural impression. The main objective is to transfer the hearing experience into written text, providing a similar experience to users with hearing impairment. Lastly, still within the musical field, this work proposed an interface to produce musical notes based on optical fibers. The prototype comprises a multimode optical fiber laid over a rigid surface, with uniform foam pads distributed uniformly over the optical fiber. The foam pads simulate piano keys and, while the user presses them, the laser spectrogram inside the optical fiber oscillates and is finally captured by a CCD camera. With the appropriate calibration of the foam pads and the spectrogram patterns, it is possible to recover which pad the user pressed and reproduce the corresponding note in real time via computer. The prototype was demonstrated for the C major scale, and it can be further configured to other scales or even different instruments.

Keywords: Han character; Asian language learning; educational technology; music visualization; human-computer interaction design; augmented reality; image classification; image recognition; information visualization; sound and music computing.

LIST OF FIGURES

List of Tables

List of Acronyms and Abbreviations

AR	Augmented Reality
BB	Bounding Box
BPM	Beats per minute
CJK	Chinese-Japanese-Korean
CM	Center of Mass
CNN	Convolutional Neural Network
lit.	Literal meaning or translation
MSE	Mean Squared Error
OCR	Optical Character Recognition

List of Symbols

Latin letters

A_{avg}	Average area of all loci
A_i	Area of locus *i*
$height$	Height of the bounding box
I_t	Image in time *t*
n	Number of characters
$n_{samples}$	Number of samples in a class
t	Time *t*
x_{BB}	x-coordinate of the bounding box
x_{CM}	x-coordinate of the center of mass
$width$	Width of the bounding box
y_{BB}	y-coordinate of the bounding box
y_{CM}	y-coordinate of the center of mass
y_{pred}	Predicted value for *i*-th sample
y_i	True value of *i*-th sample

TABLE OF CONTENTS

1 INTRODUCTION

The advancements in multimodal technology allowed many multidisciplinary applications, such as virtual classrooms, augmented reality dress rooms, interactive and multimodal learning tools, serious games, entertaining health-oriented applications and rehabilitation-oriented games (ABDULRAHAMAN *et al.*, 2020; ABDALLAH *et al.*, 2023; CHANG *et al.*, 2023). In these multimodal applications, user interaction, graphical elements and audio components are intertwined to provide an interesting and attractive experience for the users.

Given the current scenario, this dissertation focused on exploring the applications of visualization in learning Chinese as a second language and in music-related entertainment.

Chinese differs from alphabetical idioms such as English or Portuguese because of its orthography composed by hanzi 漢字 ("Han characters"), which prioritizes the meaning rather than the pronunciation (SHEN, 2005; ZHAN AND CHENG, 2014). For example, the hanzi *rì* 日 for "sun" is not a written representation of how one reads it, unlike the English word *sun* or the Portuguese word *sol*. Moreover, hanzi are comparable to building blocks: by assembling or rearranging smaller blocks, one can create novel characters in function, meaning, and sound, e.g. *lín* 林 ("woods") and *sēn* 森 ("forest") are a combination of *mù* 木 ("wood") (XIA *et al.*, 1995; TAFT, ZHU AND PENG, 1999).

On the other hand, the Chinese language is also full of homophones, such as *xián* 閒 ("relaxed") and *xián* 賢 ("virtuous"). Coincidentally, these characters can combine with other hanzi to form new homophones, e.g. *xiánrén* 閒人 ("a relaxed person") and *xiánrén* 賢人 ("a virtuous person"). This simple example shows how the choice of hanzi completely changes the interpretation of a sentence (CHEN *et al.*, 2013; REN, SHI AND ZHOU, 2001).

The resemblance between the hanzi, either because the components are alike or because they are the same combined differently, makes it easy to mistake one for another when reading, writing, or even learning, creating common misspellings (KE, 1992). For students used to a phonetic language, memorizing hanzi is particularly challenging because the structural shape and components do not always inform the pronunciation. And even though phono-semantic characters give clues regarding their sound, one must have a decent knowledge of other hanzi to infer how to pronounce an unknown one (ZHAN AND CHENG, 2014; HO, NG AND NG, 2003).

Regarding hanzi assimilation, this dissertation focused on helping the memorization and morphological understanding of the Han characters from the etymology perspective. Since hanzi are semantic-oriented and morphologically alike, which principles does a native speaker use as a guide to learn and memorize a new hanzi, distinguish similar units, or even guess the pronunciation of an unknown hanzi? From the iconic dictionary Shuowen Jiezi 說文解字 (lit.: *Talking about texts and explaining words*) (XU, EASTERN HAN), one knows that hanzi are a combination and recombination of smaller and fundamental blocks, known as "radical" (bùjiàn). Among the bùjiàn composing a hanzi, there is always a special one, namely the "indexing radical" (bùshǒu). It is special because it defines the fundamental meaning of a hanzi, resolves eventual ambiguities in the case of homonyms, and is the key element to order and classify the hanzi in digital and physical dictionaries (HO, NG AND NG, 2003). For a native speaker, identifying the bùshǒu contributes in assimilating new hanzi, differentiating ambiguities, and establishing semantic information. Although material for second-language learning rarely focuses on bùshǒu, identifying the bùshǒu in a hanzi is like second nature for native speakers, and former studies have demonstrated its importance in assimilating and memorizing hanzi for second-language learners (LÜ *et al.*, 2015; SHEN AND KE, 2007). In the literature review chapter, the concept of bùjiàn and bùshǒu will be further explained, as well as their effect on learning Chinese, and, later on, this dissertation endeavors to develop an application to introduce these elements to an adult second-language student, helping them to understand the origins and the meaning behind the hanzi so that they will be meaningful blocks instead of just random clusters of strokes.

Speaking of second language learning, one must not forget memorization of vocabulary. With technological advances, it is now possible to incorporate multimodal learning into traditional strategies like rote memorization and repetitive writing (CHUANG AND KU, 2011; SYSON, ESTUAR AND SEE, 2012). For instance, one finds variations of flashcard apps with a mixture of audio-visual keys to help one learn new vocabulary (ROSETTA STONE, 1992; ANKI, 2006), games creating an entertaining way of learning hanzi, individually or collaboratively (XIAN *et al.*, 2021; TIAN *et al.*, 2010), software with pronunciation accompanying the hanzi text (LEE AND KALYUGA, 2011) or showing the etymological evolution of the characters (ZHAN AND CHENG, 2014). Considering the importance of acquiring new vocabulary and the fact that Chinese does not mix phonetic elements in its writing systems (unlike Japanese and Korean), a student of Chinese must learn the hanzi to be able to read a text or a message. Therefore, this dissertation also proposed a flashcard-like software, with augmented reality elements, to help memorize Chinese hanzi using graphical elements. Since the flashcards accept any form of "picture" as the key to display the virtual elements, the software can be extended to Japanese kanji, and even to Korean hangul.

Apart from tools to aid second language learning, this dissertation also comprehends two music-related projects for entertainment purposes. The first one involves the development of a musical visualization routine for classical music. Currently, musical visualization tools can focus simply on entertainment, with images, animation, or pulses that follow the beats (KHULUSI *et al.*, 2020), or can delve deeper and help musicians, composers, academics, and engineers to study music and create digital content (LIMA, DOS SANTOS AND MEIGUINS, 2021). This dissertation, in its turn, explores the possibilities of showing a classical music piece's form, allying phrasing and interpretation with the typography elements to create a graphical phrasing similar to the auditive experience. The challenges of performing classical music is the complexity of the phrasing. The phrasing creates a piece of art instead of a dry and flat string of notes, and can be obtained by the most basic dynamics variation or a more complex articulation of the notes (KNÖSCHE *et al.*, 2005). Here, this dissertation focuses on the tempo (how fast the piece feels like) and on the dynamics variation (how soft/loud are the notes). Once this information is extracted from the input piece, it is transferred to written text as parameters of font type, size and color, and the text is modulated in waves similar to the waves of dynamics variation in the piece. If successfully implemented, this application can create a simple entertaining synesthetic experience to appreciate classical music, and, at the same time, it can show the form of the auditive phrasing to hearing-impaired users, so they can see the evolution of the piece.

Finally, this dissertation proposes to develop a user interface for musical applications. Although some musical instruments are accessible to most people, it requires time, space and dedication to be mastered. On the other hand, many mobile applications and consoles nowadays simulate the experience with musical instruments, but the method of interaction can be lacking and non-intuitive to users (LYONS AND FELS, 2015). In this sense, this dissertation followed the direction of allying a more intuitive physical interface with a computer to generate music. For this application, the project adopted the principles of specklegram in optical fiber to monitor the user intent. When a laser runs through a multimode optical fiber, the inner interference of the modes creates a reproducible pattern at another end of the optical fiber, known as specklegrams. Any stimuli on the optical fiber or the environmental parameters can shift and alter the output specklegram (FUJIWARA *et al.*, 2018; WU AND FUJIWARA, 2023). This project explores this principle to monitor the stimuli over the fiber, dictated by the user, and links each output specklegram to a specific musical note. Then, by interacting with the optical fiber-based interface, the user can generate a string of musical notes. Besides, with the sensitivity of the specklegrams to spatial shifting of the pressure point on the optical fiber, this user interface can create some sound effects that are particular to string instruments, such as glissando and vibrato.

2 OBJECTIVES

This dissertation presents the development of four projects, each with its motivation and goal. For clarity, they are listed individually below:

2.1 Radical identifier

Based on the importance of the indexing radical composing Chinese logograms, the radical identifier aimed to indicate which and where the bùshǒu of a hanzi is. Currently, the project supports 30 indexing radicals with the most characters in the Chinese lexica. The final goal is to aid students of Chinese as a second language from Western-alphabet-based countries in understanding and remembering hanzi as meaningful and logical entities instead of random and highly similar blocks.

2.2 Augmented reality learning environment

Following the same motivation of aiding the learning process of Chinese as a second language, this multimodal project integrates the 3D virtual objects with virtual transliteration of selected hanzi in Chinese pinyin. Considering the similarity between Chinese hanzi and Japanese kanji, the Japanese romaji is offered as an alternative to transliteration, proving that the software can be adapted to different languages. The developed software functions as a multimodal flashcard platform, and the user can select what information will be displayed on the screen, and create rounds of learning new knowledge and reviewing a learned one.

2.3 Music visualization

This project integrates classical music with typography, modulating texts with selected color, font type, and font size to express graphically the phrasing in a classical music piece. The main purpose is to show the dynamic variations and the tempo in a ludic, intuitive way, aside from providing hearing-impaired users some experience to “view” the phrasing in the interpretation of the musicians.

2.4 User interface for music

A user interface for musical applications differs from traditional musical instruments in being versatile in the digital entertainment field since it supports multimodal functions. In this case, the developed user interface converts the mechanical input from the user via an optical fiber specklegram-based sensor into sound output. The nature of the optical fiber specklegram allows the construction of a simple, compact, and feasible setup for user interaction. The main goal is to explore optical fiber technology in developing a piano-like user interface, which can be further integrated into music-generating or similar entertainment multimodal applications.

3 LITERATURE REVIEW

3.1 Han characters

Hanzi 漢字 ("Han characters") are the basic units of the Chinese writing system, shared by Japanese and Korean, in which case they are known as *kanji* and *hanja*, respectively. Rather than strings of a phonetic alphabet, groups of dots and horizontal, vertical, and diagonal strokes combine to form basic components that further create individual characters as unique entities (CHU, NAKAZAWA AND KUROHASHI, 2012; CHEN *et al.*, 2014; ZHANG, 2014). Therefore, Han characters represent forms and ideas instead of the sound, unlike alphabetical words from Western languages (XIA *et al.*, 1995; HAN, 2020).

According to legend, a man called Cangjie (Cāngjié 倉頡) created symbols after observing the footprints of different animals. The first characters to be created reproduced the forms of nature and were called *wén* 文 ("words"); then the association between forms and sounds originated the *zì* 字 ("symbols") (XU, EASTERN HAN). Further on, the fundamental units of dots and strokes recombined to form novel characters, which are subdivided into six categories (*liùshū* 六書; lit.: six writings) depending on their application and the principle behind their creation (XU, EASTERN HAN; LINDQVIST, 1991; LIU, 2021):

a) Ideograms (*zhǐshì* 指事; lit.: indication):
   They indicate an idea or a concept with their structural composition, such as *shàng* 上 ("top") and *xià* 下 ("bottom"), as illustrated in Figure 3.1.
b) Pictograms (*xiàngxíng* 象形; lit.: form imitation):
   Their shape derives directly from the objects they represent, such as *rì* 日 ("sun"), *yuè* 月 ("moon"), and the examples in Figure 3.2.
c) Phono-semantic characters (*xíngshēng* 形聲; lit.: form and sound):
   These compound characters are a combination of two entities: one provides the phonetic and the other the semantic information, such as *jiāng* 江 ("large river") and *hé* 河 ("river").

d) Compound ideographs (*huìyì* 會意; lit.: joined meaning):

They are a combination of two or more ideograms or pictograms whose union point to a new meaning, such as *xìn* 信 (“credibility”) and *míng* 明 (“bright”).

e) Derivative cognate (*zhuǎnzhù* 轉注; lit.: reciprocal meaning):

They are mutually explanatory characters that share morphological, phonetic, or semantic similarities, such as *lǎo* 老 (“old age”) and *kǎo* 考 (“long life”).

f) Rebus or phonetic loan characters (*jiǎjiè* 假借; lit.: false borrowing):

Existing characters gain a diverse definition because both old and novel meanings share the same pronunciation, such as *lìng* 令 (“order”, borrowed to express an honorific title or a government position) and *cháng* 長 (“far away”, borrowed to express “respectable, elderly person” or “chief”).

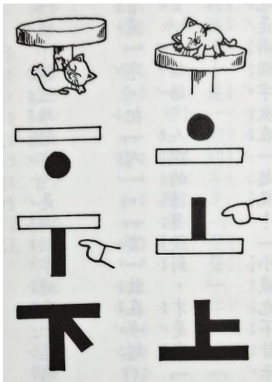

Figure 3.1: Illustrative explanation of how the ideograms shàng *上 (“top”) and* xià *下 (“bottom”) were created. From: Xia* et al. *(1995).*

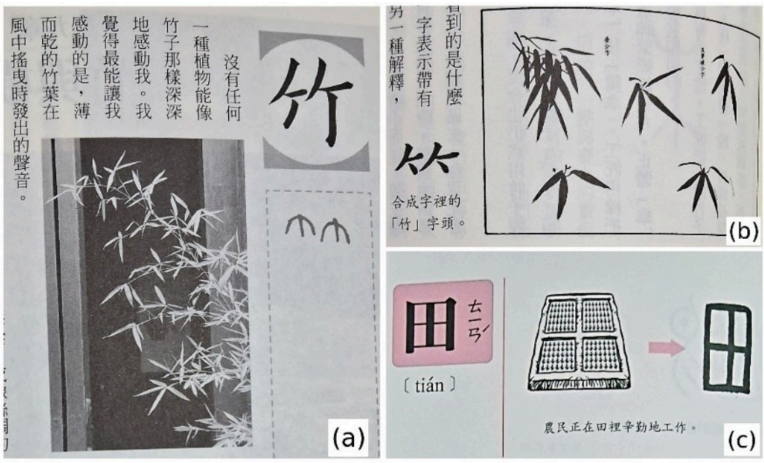

Figure 3.2: Examples of pictograms. The hanzi zhú *竹 (“bamboo”), in its (a) whole and (b) partial form in composed hanzi, resembles bamboo leaves. From: Lindqvist (1991). (c) The hanzi* tián *田 (“field”) reproduces the image of plantation fields. From: Wu (2017).*

Investigating the origins of hanzi shows that they are deeply involved in everyday life, cultural aspects, and the values of Chinese society, and each character has its definition (HAN, 2020; LINDQVIST, 1991). Classical Chinese (*wényánwén* 文言文), an ancient oral style that remained as a literary language until the 20th century, is very concise and compact precisely because it employs single hanzi to play any of the various syntactic roles within a sentence (XU *et al.*, 2021; LIU *et al.*, 2022). In comparison, the modern style, also known as written vernacular Chinese (*báihuàwén* 白話文), has longer phrases because most words nowadays have multiple characters (HUANG *et al.*, 2002; WU, ZHAO AND CHE, 2018; TSAI, WEI AND TSENG, 2021). Besides the usual number of hanzi composing syntactical terms, ancient grammar differs from modern grammar and allows omissions around terms in Classical Chinese (FENG, 2015). Consequently, it is even more challenging to interpret the ancient variation relying only on listening because many hanzi share the pronunciation but not the meaning, increasing the candidates of which single character word is in use.

Chinese is indeed full of homonyms. Compared to Japanese, Chinese has much fewer homographs, but the homonyms are abundant by contrast. Besides the homonyms, there are also common misspellings (*cuòbiézì* 錯別字; lit.: wrong or another character), composed by pairs of hanzi that are structurally, visually, grammatically, or phonetically similar (KE, 1992). The homonyms and misspellings are challenging not only for someone learning Chinese as a foreign language but for native speakers as well. For example, both *nǎo* 惱 ("distress, anger") and *nǎo* 腦 ("brain, mind") sound the same, but one should use them accordingly depending on the context. Then one has to remember to pronounce 好 as *hào* (fourth tone) instead of *hǎo* (third tone) when it acts as a verb. Besides the homonyms and common misspellings, there are also some hanzi that are almost visually identical, such as *rì* 日 ("sun") and *yuē* 曰 ("to speak"), *tǔ* 土 ("earth") and *shì* 士 ("scholar"). A closer look shows that each pair differs only in the proportion between the strokes, leading to common reading mistakes depending on the calligraphy or font type, and one will need the context to identify which is the correct candidate. With these simple examples, one notes that Chinese language users should have a certain amount of lexical knowledge to read and write appropriately (TIAN *et al.*, 2010).

3.2 Bùshǒu radical

The Han characters fall into any of the six groups of *liùshū* depending on their usage and origin; however, whichever their category is, all can be disassembled into smaller basic units called *bùjiàn* 部件 (lit.: parts, components). Bùjiàn can have an independent definition and pronunciation and rearrange to create different characters (Li and Chen, 1997; Xue *et al.*, 2021; Li and Zhou, 2007).

Taking the hanzi shown in Figure 3.3, for example, the first three (Figure 3.3(a-c)) exist individually with their own definition but cannot be further separated into smaller components. On the other hand, they can compose novel hanzi, such as the ones in Figure 3.3(d-f).

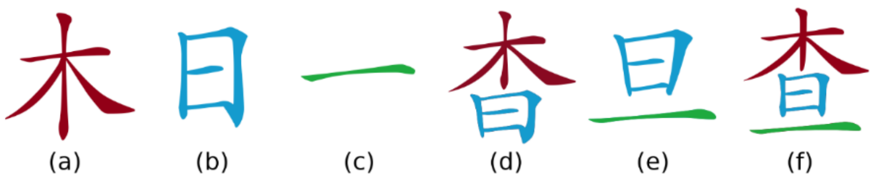

Figure 3.3: Examples of hanzi and their bùjiàn components. Hanzi (a) mù *木 ("wood"), (b)* rì *日 ("sun"), and (c)* yī *一 ("one") cannot be further segmented into smaller bùjiàn. As fundamental blocks, the three of them can recombine to form new hanzi, such as (d)* yǎo *杳 ("dark and quiet"), (e)* dàn *旦 ("dawn"), and (f)* chá *查 ("to investigate"). Identical blocks share the same color.*

Among the bùjiàn components, there is always one unique and special entity, known as *bùshǒu* 部首 (lit.: section header) or "indexing radical" (Li *et al.*, 2021). Its importance results from some reasons, among them (Wu, Fujiwara and Suzuki, 2022; Breen, 2004; Wang, 1998):

a) Although a hanzi can be composed of multiple bùjiàn components, it has only one bùshǒu radical.

   Returning to the six examples in Figure 3.3, the first three coincidentally are their own bùshǒu radical. The remaining examples have two or three bùjiàn components, but only one is indexing radical: the red block for Figure 3.3(d, f) and the blue block for Figure 3.3(e). Similarly, Figure 3.4 shows three examples whose bùshǒu radical is in red. Although the block *rì* 日 ("sun") comes in all three, it is only indexing radical in Figure 3.4(a), whereas it is just an ordinary component in the other two (in gray).

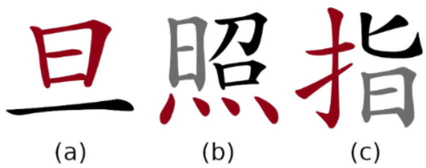

Figure 3.4: Examples of bùshǒu radical within hanzi, highlighted in red: (a) sun, (b) fire, and (c) hand. The red block in (a) and the gray blocks in (b-c) are the same as rì *日 (“sun”).*

b) The bùshǒu radical usually defines the meaning and application of the hanzi. The pair of homophones shown in Figure 3.5 takes the sound of the phonetic component on the right (in black) while the meaning is related to the bùshǒu radical on the left (in red). *Jìng* 境 (“land”) is indexed under radical *tǔ* 土 (“earth”) because the land is a terrain. On the other hand, *jìng* 鏡 (“mirror”) has *jīn* 金 (“metal”) as bùshǒu radical because ancient mirrors were polished metal surfaces with reflective properties.

c) The entries inside dictionaries are traditionally organized according to the bùshǒu and, subsequently, by the number of remaining strokes.

d) The bùshǒu radical establishes an entire family of characters with similar or related meanings.

   For example, all five hanzi in Figure 3.6 refer to female members in a family; therefore, they naturally have *nǚ* 女 (“woman”) as a semantic defining component and, consequently, the indexing radical.

Figure 3.5: Examples of phono-semantic homophones whose definition comes from their bùshǒu radical (in red): (a) land is earth-related, and (b) mirror is metal-related.

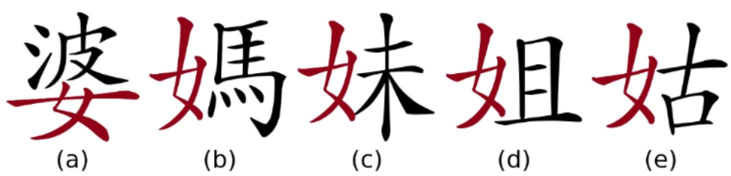

Figure 3.6: Examples of hanzi with indexing radical nǚ *女 (“woman”, in red) and respective English translation: (a) grandmother, (b) mother, (c) younger sister, (d) older sister, and (e) aunt.*

Please note that many English works in the literature refer to bùjiàn components as "radical" and the few works studying specifically bùshǒu radicals also refer to them as "radical". To avoid misunderstanding and ambiguity, bùjiàn will be referred to as "bùjiàn components" or "components" and bùshǒu as "bùshǒu radicals", "indexing radicals" or simply "radicals" along this work.

The linguist Xu Shen of the Eastern Han dynasty listed 540 bùshǒu radicals in the book Shuowen Jiezi 說文解字 (lit.: *Talking about texts and explaining words*) (XU, EASTERN HAN). Afterward, linguists of the Qing dynasty recompiled the characters in the Chinese language and composed the iconic Kangxi Zidian 康熙字典 (lit.: *Dictionary of Kangxi*), with 214 bùshǒu radicals, which currently remains (KANGXI ZIDIAN, c2005-2023). Figure 3.7 presents the basic morphology of all 214 bùshǒu radicals, ordered from the least to the greatest number of strokes and according to the default indexing order in dictionaries. The alternative forms, taken by some depending on their location within the hanzi, are not shown (LI *et al.*, 2021; WU, FUJIWARA AND SUZUKI, 2023).

1	一 1	丨 2	丶 3	丿 4	乙 5	亅 6	**2**	二 7	亠 8	人 9	儿 10	入 11	八 12	冂 13	冖 14	冫 15	几 16	凵 17	刀 18	力 19
勹 20	匕 21	匚 22	匸 23	十 24	卜 25	卩 26	厂 27	厶 28	又 29	**3**	口 30	囗 31	土 32	士 33	夂 34	夊 35	夕 36	大 37	女 38	子 39
宀 40	寸 41	小 42	尢 43	尸 44	屮 45	山 46	巛 47	工 48	己 49	巾 50	干 51	幺 52	广 53	廴 54	廾 55	弋 56	弓 57	彐 58	彡 59	彳 60
4	心 61	戈 62	戶 63	手 64	支 65	攴 66	文 67	斗 68	斤 69	方 70	无 71	日 72	曰 73	月 74	木 75	欠 76	止 77	歹 78	殳 79	毋 80
比 81	毛 82	氏 83	气 84	水 85	火 86	爪 87	父 88	爻 89	爿 90	片 91	牙 92	牛 93	犬 94	**5**	玄 95	玉 96	瓜 97	瓦 98	甘 99	生 100
用 101	田 102	疋 103	疒 104	癶 105	白 106	皮 107	皿 108	目 109	矛 110	矢 111	石 112	示 113	禸 114	禾 115	穴 116	立 117	**6**	竹 118	米 119	糸 120
缶 121	网 122	羊 123	羽 124	老 125	而 126	耒 127	耳 128	聿 129	肉 130	臣 131	自 132	至 133	臼 134	舌 135	舛 136	舟 137	艮 138	色 139	艸 140	虍 141
虫 142	血 143	行 144	衣 145	襾 146	**7**	見 147	角 148	言 149	谷 150	豆 151	豕 152	豸 153	貝 154	赤 155	走 156	足 157	身 158	車 159	辛 160	辰 161
辵 162	邑 163	酉 164	釆 165	里 166	**8**	金 167	長 168	門 169	阜 170	隶 171	隹 172	雨 173	青 174	非 175	**9**	面 176	革 177	韋 178	韭 179	音 180
頁 181	風 182	飛 183	食 184	首 185	香 186	**10**	馬 187	骨 188	高 189	髟 190	鬥 191	鬯 192	鬲 193	鬼 194	**11**	魚 195	鳥 196	鹵 197	鹿 198	麥 199
麻 200	**12**	黃 201	黍 202	黑 203	黹 204	**13**	黽 205	鼎 206	鼓 207	鼠 208	**14**	鼻 209	齊 210	**15**	齒 211	**16**	龍 212	龜 213	**17**	龠 214

Figure 3.7: List of the 214 official indexing radicals currently in use.

For the native speakers, the bùshǒu radical plays a vital role in the learning process during the school years and gives them cues regarding the semantic category of hanzi (TONG, TONG AND MCBRIDE, 2017; HO, NG AND NG, 2003). In the previous example of *jìng* 境 ("land") and *jìng* 鏡 ("mirror"), what defines the meaning is precisely the bùshǒu radical. Moreover, over 80% of hanzi are phono-semantic, so the awareness of both bùshǒu and bùjiàn helps natives to infer the meaning and the pronunciation of unknown characters (HO, NG AND NG, 2003; TONG AND YIP, 2015). Likewise, radical and component information has proven to be effective for learners of Chinese as a foreign language in memorizing and improving the reading of hanzi (WONG, 2017; LÜ *et al.*, 2015; SHEN AND KE, 2007). In particular, this knowledge does not provide only semantic and phonetic cues but also gives the feeling of the probable loci a lexical block should occupy. As a result, the radical and component sensitivity reduces the number of block combinations a learner should memorize and hones the intuition towards impossible structural compositions (CHEN *et al.*, 2013; TONG, TONG AND MCBRIDE, 2017).

Considering the importance of radicals and component blocks, researchers have been exploring them in different applications. Hu and Du (2012) demonstrated that indexing radicals are relevant clues in text mining and content extraction, and Tao *et al.* (2019) obtained similar results in text classification, relying only on the bùshǒu radical. On the other hand, Tan *et al.* (2021) proposed a framework to extract information from digital medical and doctor-patient Q&A records because sentences with terms from bùshǒu radical *chuáng* 疒 ("resting patient", "diseases") and *ròu* 肉 ("meat") are usually disease-related or body-related.

Although these works have already demonstrated that the indexing radical alone is enough for text classification, Ke and Hagiwara (2017) opted for the full decomposition of a hanzi or kanji for sentiment analysis in texts, which is closer to a component-level character recognition than a general semantic identification since a particular set of bùjiàn components can only refer to a unique character.

More complex tasks also require this much information to be conducted. Zhang, Du and Dai (2020) made use of all bùjiàn components of a hanzi in character recognition in natural scenes, and the same applies to handwriting recognition performed by Wang *et al.* (2019) and Shi, Gunn and Damper (2003). Similarly, style transfer (HUANG *et al.*, 2020) and character creation (XUE *et al.*, 2021) require the full decomposition of a hanzi as well. Considering the nature of these works, the requirement for more complex data than the bùshǒu radical is reasonable because one indexing radical can belong to more than just a few hundred different characters, while a specific set of bùjiàn components matches just one hanzi. Even if there are very similar hanzi to cause ambiguity, knowing the set of all components narrows down the candidates to just a few entities.

3.3 Multimodal learning and mnemonics

During the learning process of any language, one must acquire new vocabulary to develop passive understanding and enhance active production. Traditionally, natives learn hanzi through the repetitive exercise of the strokes after learning its definition, pronunciation, and common usages (ZHANG, 2014). A similar approach is applied for foreign learners and usually involves some degree of word-to-word translation to their native language. People already familiar with Han characters, such as Japanese or Korean students, can relate prior knowledge (of kanji or hanja) to new one (of hanzi), and vice versa (HAGIWARA, 2016). The learning process, however, is more challenging for someone from a phonetic orthography due to the lack of obvious relationships between shape and sound. Moreover, the low discriminability between characters encourages confusion among beginner students (SOEMER AND SCHWAN, 2012). Besides the traditional oral and written repetition, mnemonic strategies are also a viable option to memorize vocabulary (PAIVIO AND DESROCHERS, 1981; PUTNAM, 2015).

Mnemonics are strategies to aid memorization whose effectiveness in second language learning has been proven (COHEN, 1987; WANG AND THOMAS, 1992). Classical mnemonics in second-language learning involve picture-to-word translation, though it does not apply to some words, such as abstract concepts. Alongside images, animation and sound can stimulate memorization too (ZHAN AND CHENG, 2014; SOEMER AND SCHWAN, 2012).

Many language learning platforms explore this idea to fixate on vocabulary. For instance, Rosetta Stone (1992) offers digital flashcards with translation, pictures, texts, and audio for language learning. On the same line, Anki (2006) provides multimodal flashcard functionality along with an automatic review cycle to refresh the memory. Duolingo (2012) presents new vocabulary within each chapter with corresponding pictures and audio, showing the English translation only if the user selects the word, creating thus a beginner-friendly immersive environment. The platform How to Study Korean (c2023) offers a flashcard system with English translation, Korean audio, and hangul to fixate on new vocabulary.

Another common approach to learning vocabulary is through rhymes and storytelling, with some plot that employs the words of interest, or by diving deeper into the etymology of the word (AMY KUO AND HOOPER, 2004; PUTNAM, 2015; TIAN *et al.*, 2010). In the case of Chinese, understanding the etymology is especially helpful since the poorly distinguishable characters have a specific origin, often related to their definition or pronunciation

(Lindqvist, 1991; Han, 2020). For instance, pictograms originated from form imitation, meaning they are a written and simplified representation of the real object. Meanwhile, phono-semantic characters are compound hanzi with semantic and phonetic blocks, and ideograms indicate a concept or an idea with their structural composition (Xu, Eastern Han; Xia *et al.*, 1995). Throughout history, orthography evolved, and the creation principles of hanzi are not as evident as in the past. Therefore, analyzing the etymological roots will help one understand and memorize Han characters (Soemer and Schwan, 2012).

Nowadays, one can also employ educational games for memorizing Sino-Japanese characters. A series of gamified learning of Japanese kanji is available for Nintendo Switch, with puzzles and writing activities for the characters from Japanese elementary school, and each version of the game focuses on a specific grade (Media 5 Corporation, 2019). Xian *et al.* (2021) presented KANJI Write, an augmented reality (AR) mobile game for beginner learners of the Japanese language that incorporates elements of adventure games: the user selects a level with the corresponding AR card, then plays the stage to learn the strokes in the correct order and defeat enemies within time given. Plecher *et al.* (2018) combined adventure, role-playing elements, and Japanese culture to create Dragon Tale, another AR game for learning kanji, available for tablet devices, that explores the various aspects of kanji, such as stroke order, meaning, pronunciation, and compounds. Keehl and Melcer (2021) proposed Radical Tunes, a musical kanji-writing game that assigned a unique melody to specific stroke clusters inside a kanji, so the user creates a melodic link between the blocks one writes instead of the proper pronunciation of the word.

Similarly, multimodal resources are now usual in traditional classrooms to improve the memorization of hanzi strokes, pronunciation, and even content interpretation. With a character-centered approach, Lin, Lim and Wu (2022) proposed a learning game for primary school students in Singapore, providing an entertaining and collaborative environment to enhance and stimulate their interest in learning hanzi. Wong *et al.* (2013) developed a mobile app allowing students to disassemble and rearrange the components of hanzi. Wen (2018) proposed a similar learning tool with AR paper, combining the physical and the virtual worlds instead of a completely digital platform. On the other hand, Lee and Kalyuga (2011) focused on improving reading skills by supplying phonetic pinyin alongside hanzi within sentences to stimulate memory. Meanwhile, on a semantic-centered approach, Chen and Lin (2016) aid the interpretation of ancient Chinese poetry with immersive story simulation as opposed to the traditional verbal explanation of Classical Chinese content.

3.4 Music visualization

Music visualization has a long history in entertainment, tracing back to the vintage graphical effects on music players, that follow the sound beats with abstract points, waves, or lines to convey an aesthetically pleasing effect (POURIS AND FELS, 2012).

Apart from the entertainment field, music visualization shows its importance in technical professions for music academics, music engineers, and music composers. The specific purpose will define the type of input format: MIDI, common audio format, or others, and the medium to visualize the desired information: colors, shapes, abstract elements, pictures, score notation, etc. (LIMA, DOS SANTOS AND MEIGUINS, 2021; KHULUSI *et al.*, 2020).

For example, Chen *et al.* (2008) show photos to match the mood of the music. Soraghan *et al.* (2018) described timbre with semantic descriptors, such as bright-dark and dull-sharp. Fonteles, Rodrigues and Basso (2013) also described audio timbre, classifying the instrument families with colors. Ciuha, Klemenc and Solina (2010), on the other hand, showed the harmonic feature by representing intervals with colors and different saturations. Goto (2001) translates the beats of the music into the position and movement of virtual dancing dolls.

Han *et al.* (2023) visualizes the lyrics of the input music and modulates the characters' typography according to the rhythm of the music. For example, if the singer extends the sound, the character becomes wider, and if a part is faster, the character becomes thinner. Similar applications do not exist for classical music. Although classical music pieces do not have lyrics, they usually have a higher complexity in phrasing than pop music. The phrasing in classical music can be interpreted as a conversation or a speech from the instrument, with rises and falls in the intonation, and a cycle of softer and louder intensities to create a full picture of different sentences, weaving a whole piece.

This dissertation investigates the possibilities of showing the phrasing in classical music. The main questions are: what elements from the phrasing will be shown, which features should be extracted from the piece, and what are the most intuitive ways of showing them?

In musical education, the simplest way of creating phrasing is to weave cycles of piano-forte (i.e., softer and louder sounds) across notes and bars. The ascending and descending waves create depth in the music performance, avoiding a flat reproduction of just notes. Another fundamental characteristic of a piece of music is the tempo, i.e., how fast it sounds. This dissertation focuses on these two features to develop a music visualization routine that modulates a text accordingly, with colors and ascending and descending waves instead of a flat text with just one font type and size.

3.5 User interface for music

User interfaces for music applications are rich in variety and application. Some resemble traditional musical instruments, such as pianos, violins, and cellos, and some are peculiar and unusual in form and mechanism. No matter the nature of the user interaction principles, these interfaces consist of two blocks: the control space and the sonic space (LYONS AND FELS, 2015).

The control space can have a foundation on different principles, such as airflow, light level, the motion of body parts, proximity, neurophysiological signals, touchscreen, etc. (FRID, 2019). For example, Fels *et al.* (2004) proposed a wind instrument, activated by breath from two extremities and played by two people. Lyons and Tetsutani (2001) developed a musical interface controlled by facial expressions, acquired by video. Jordà (2010) created Reactable, a multimodal and intuitive tabletop interface for a tangible and visual musical experience, by exploring visual markers. Kang and Chien (2010) also proposed a camera-based user interface but chose calligraphy as the control parameter, i.e., the user plays music by writing traditional brush calligraphy. On the other hand, Wicaksono and Paradiso (2017) proposed FabricKeyboard, a fabric ribbon keyboard that supports multiple sensing modalities to control the parameters of the generated music: proximity, touch and electric field, pressure, stretch, and position. Hollinger *et al.* (2007) proposed a piano-like instrument with mechanical keys and optical sensing based on LED to evaluate the influence of playing musical instruments over neural activities, so the whole setup is compatible with the magnetic resonance imaging scanner environment. A similar user interface was developed, but this time simulating a cello, to study the neural stimuli of playing a string instrument (HOLLINGER AND WANDERLEY, 2015).

Following the control space, the sonic space generates the corresponding sounds and effects according to the commands from the user. Usually, the sonic space is composed of different digital sound generators, a mobile device, or a computer (LYONS AND FELS, 2015).

4 RADICAL IDENTIFIER

Related studies have proven the importance of radicals in the reading, learning, and comprehension of hanzi users. Understanding the probable loci a radical or component occupies within a character hones the instinct regarding orthography and simplifies the memorizing process of seemingly random and unrelated strokes. In particular, the indexing bùshǒu radicals are valuable clues for eliminating ambiguity before homonyms, context understanding, and overall meaning inference when one faces unknown characters. Although electronic dictionaries nowadays provide automatic search bars, and one can refer to Google to look up the meaning or pronunciation of unfamiliar hanzi, indexing radical contributes still for mastering Han characters.

In this sense, this dissertation proposes to extract the bùshǒu radical from images of hanzi, i.e., the input will be image data instead of communication symbols. The project opted for simple images of hanzi as input for practical reasons since most beginner learners of Chinese as a foreign language probably will find it difficult to type in the hanzi to consult it on Google or electronic dictionaries. On the other hand, they have easier access to a smartphone camera, so they can take a photo of the hanzi and consult its bùshǒu radical without further digitizing. This alternative would be more practical than running an optical character recognition (OCR) over the image to digitize the content, even though it would be simpler to consult the bùshǒu radical or even the meaning of the hanzi as a Unicode, due to the limitations of OCR over Asian script, despite this technology being effective in extracting text information from images for Western languages. The major challenge for OCR derives from the close similarity between hanzi and their structural complexity. Moreover, not all OCR software adjusts well to vertical text flows, different fonts, and content with combined Asian and alphabetical words (YIN *et al.*, 2019; KIM, KIM AND KIM, 2019).

Initially, the radical identifier supports only 30 radicals with the most entries registered in the unified Chinese-Japanese-Korean (CJK) Unicode table (THE UNICODE CONSORTIUM, 2021), as listed in Table 4.1, organized in descending order of entries instead of the official positioning of the 214 radicals (Figure 3.7).

Table 4.1: List of the bùshǒu radical supported by the radical identifier, with the class label, hanzi representation, English translation, and the number of entries in the CJK Unicode table.

Class Label	Radical	English Translation	Entries	Class Label	Radical	English Translation	Entries
0	水	Water	1079	15	竹	Bamboo	378
1	木	Wood	1016	16	山	Mountain	362
2	艸	Grass	981	17	肉	Meat	346
3	口	Mouth	756	18	玉	Jade	339
4	手	Hand	740	19	石	Stone	327
5	金	Metal	692	20	鳥	Bird	314
6	人	Person	645	21	衣	Clothes	283
7	心	Heart	581	22	日	Sun	267
8	女	Girl	477	23	足	Foot	248
9	虫	Insect	469	24	目	Eyes	237
10	土	Earth	460	25	疒	Convalescence	228
11	火	Fire	447	26	辵	To Walk	220
12	糸	Thread	423	27	犬	Dog	216
13	言	To speak	416	28	馬	Horse	192
14	魚	Fish	395	29	邑	Country	184

Radical identification combines two separate tasks: one of classification and another of localization. The classifier attributes the input hanzi to a bùshǒu radical while the localizer marks where it is within the character.

The proposed identification process is serialized, i.e., the input samples are classified according to the bùshǒu radical they likely belong to, and only then does the localizer searches for the radical block. This serialization allows the localizer to focus on just one instead of 30 (eventually 214) radical blocks. This simplification comes in handy when one notices that various bùshǒu radicals transfigure according to the position they occupy (as exemplified in Figure 4.1) besides bùjiàn components often coincide with existing indexing radicals (Figure 3.3 and Figure 3.4).

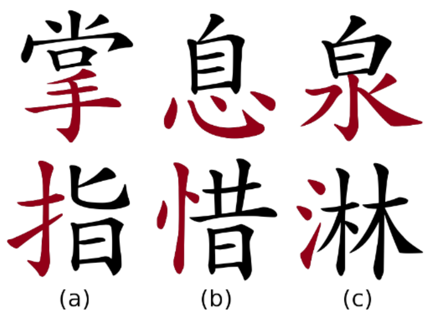

Figure 4.1: Examples of bùshǒu radicals with structural modification according to their placement within the hanzi. (a) Hand, (b) heart, and (c) water.

4.1 Classifier and localizer models

The classifier and the localizer were implemented in Keras with Tensorflow backend (CHOLLET. *et al.*, 2015).

As shown in Figure 4.2, both the classifier and the localizer take as input a 1-D grayscale image of 28 × 28 pixels. It holds a single hanzi within, wherein 0 corresponds to the background and 255 to the strokes, with intermediate values for the border areas where strokes and background pixels meet.

Considering the simplicity of the input samples, a convolutional neural network (CNN) with two hidden layers is proposed for the task. Each layer has 64 nodes, a stride of 2, 3 × 3 kernels, a Leaky ReLu activation function, and a dropout of 50%. The output of the hidden layers is flattened, and only then does the architecture of the classifier and the localizer differ.

The output layer of the classifier is a dense layer with a SoftMax activation function that returns a probability vector of 30 elements, each indicating the probability of the input sample belonging to any of the 30 classes in the training dataset. The final classification result matches the most probable class after the CNN.

On the other hand, the output layer of the localizer is a sigmoid function that returns a set of six floating numbers. These numbers tag the center of mass and the bounding box of the strokes from the radical block. The returned values range from 0.0 to 1.0 and are the coordinates normalized to the input image size. For instance, a center of mass in (0.5, 0.5) is right at the center of the image regardless of its nominal size.

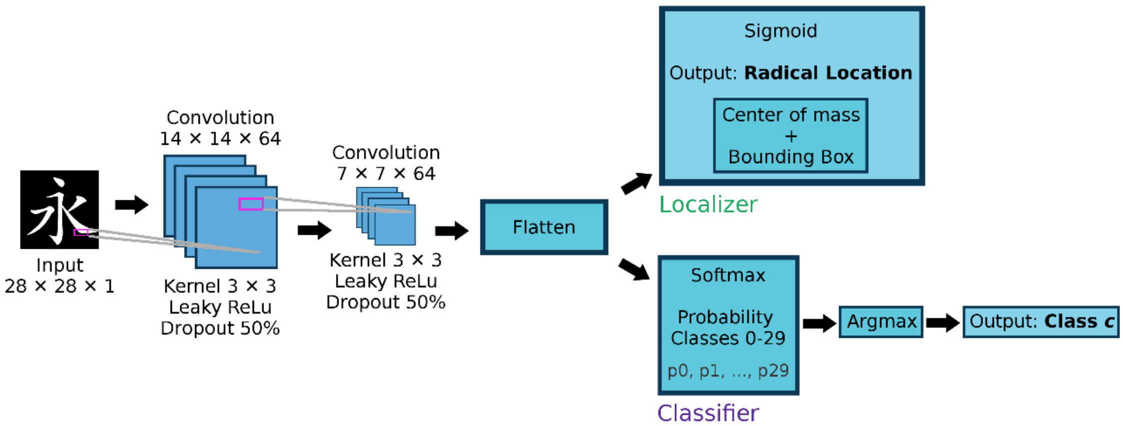

Figure 4.2: Diagram of the radical identifier architecture, separated into classifier and localizer.

Using a Windows 10, 64-bit machine with Intel® Core™ i5-7200U CPU running at 2.50 GHz, 8 GB of RAM, and without GPU, it took approximately 25 hours to train the classifier, and approximately 40 minutes to train the localizers. The training time discrepancy is attributed to the size of the training dataset (the classification dataset has 23 times more samples than the localization dataset, as explained in the following section).

4.2 Dataset

The dataset is machine-generated via a Python routine inspired by the zi2zi project (TIAN, 2017) and is stored as a NumPy array with the compressed .npz extension. Following the expected input data of the classifier and localizer models, the said NumPy array is a stack of 28×28 matrices that holds one hanzi each.

4.2.1 Basic dataset

The basic dataset has one instance of all hanzi indexed under the 30 bùshǒu radicals in Table 4.1. Therefore, it has 13,718 samples in total (the sum of the column "Entries" in Table 4.1).

It is worth mentioning that this dataset contains all instances in the CJK Unicode Table, including the traditional and simplified orthography, besides the main and the secondary forms of the radical block. Although the default morphology is usually a hanzi on its own, the alternative one generally does not exist outside the compound hanzi.

4.2.2 Classification dataset

The classification dataset reproduces the basic dataset in different font styles. The project selected 33 fonts from the HuaKang family, varying from handwriting, brush, and print fonts. Figure 4.3 provides an example of the same hanzi rendered in different font styles. One notices that the print font is square and boxy, while the handwriting is usually slanted and plays with proportions, with varying visual effects. The brush writing is an intermediate style and can be either regular or flourished, with clean or blended strokes.

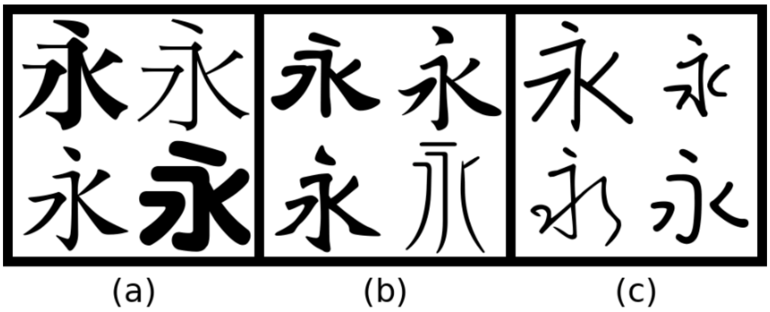

Figure 4.3: Instances of the same hanzi yǒng 永 *("forever"), rendered in different fonts from (a) basic print, (b) brush, and (c) handwriting styles HuaKang Family.*

In theory, rendering 13,718 samples 33 times should result in 452,694 entities; however, as mentioned before, the basic dataset contains all instances registered in the CJK table, some of which are not complete hanzi individually, and many are simplified variants or unusual hanzi. Although the built-in true type and other widely used font styles can render those instances, the less common font types do not support the incomplete or the unusual hanzi. In such cases, they all point to the system default font for Unicode characters or result in a blank image, meaning that the same hanzi would be rendered multiple times as a visually identical unit. The dataset generator checks and removes the empty or duplicated images, amounting to 316,309 samples in the classification dataset. Finally, all receive their class label, a number between 0 and 29, corresponding to the bùshǒu they belong to. Figure 4.4 provides some examples of the classification dataset with respective ground truth.

Figure 4.4: Samples from the classification dataset. The numbers above the hanzi refer to their indexing radical (ground truth).

4.2.3 Localization dataset

The localization dataset contains the basic dataset, rendered in the DF-Kai font type. As a built-in font type of Windows system, it supports all instances in the CJK table, even the incomplete blocks and the simplified variants.

Unlike the classifier, to whom the amalgamations of the inner strokes matter the most, the localizer requires the position of the radical block inside the hanzi body instead. Observing Figure 4.3 once more, one notes that, although the hanzi looks slightly different due to the calligraphy, the relative position of the elements within the character is well defined. For instance, stroke clusters on the top will not jump to the bottom just from a varying font style.

The radical block is marked individually for all data samples, indicating both the center of mass and the bounding box. Although it sounds redundant to tag both information to localize an entity, the center of mass is not always in the middle of the bounding box, as exemplified in Figure 4.5. Figure 4.6 provides some examples of the localization dataset with respective ground truth. Numerically, the ground truth is an array of six floating points: ($x_{CM}, y_{CM}, x_{BB}, y_{BB}$, *width, height*), where (x_{CM}, y_{CM}) correspond to the normalized coordinates of the center of mass and (x_{BB}, y_{BB}) to the normalized coordinates of the top-left corner of the bounding box, and (*width, height*) to the size of the bounding box. All positions and sizes fit within 0 and 1, corresponding to the normalized position inside the 28×28 area. In other words, (0, 0) is the bottom-left corner, and (1, 1) is the top-right corner of the image.

Since the proposed radical identifier employs a serialized procedure and the localizer considers the classification outcome before looking for a specific block, there are as many localization datasets as the number of class labels. Each localization dataset contains the number of samples as specified in the column "Entries" in Table 4.1.

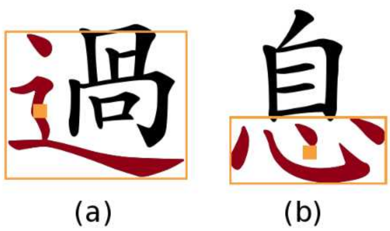

Figure 4.5. Examples of hanzi whose bùshǒu radical (red) has: (a) a center of mass (small orange square) dislocated from the center of its bounding box (outer orange box), and (b) a center of mass at the center of its bounding box.

Figure 4.6: Samples from the localization dataset. The true bounding box and center of mass of the bùshǒu radical block are in orange.

4.2.4 Validation dataset

The classification and the localization datasets are shuffled and randomly split into two subsets according to an 85%-15% ratio. The smaller set remains hidden from the training process, so it is unknown to the model for the subsequent validation.

4.2.5 General test

The initial excerpt from the historical novel "Romance of Three Kingdoms" by Luo Guanzhong (LUO, 1991) provided a random and uncontrolled set of characters to evaluate the performance of the radical identifier. The selected excerpt contains 743 characters, 272 (~36.6%) of which belong to one of the 30 radicals in the training dataset.

Given the amount of hanzi and their randomness inside the excerpt, generating the characters individually with the Python algorithm that produced the training and validation datasets became impractical. Instead, the radical identifier received a single JPEG image containing all 743 characters as input; however, with the input data specifications for the classifier and localizer, an additional image-processing routine is required to extract the content from the input image and prepare each character into 28 × 28 intensity image.

This routine explores the nature of machine Han characters: regardless of the number of strokes, all hanzi fit inside the same square box, also known as the full-width character size. The only difference is the percentage of empty space within, depending on the visual complexity and number of strokes. Therefore, the content in a text consisting only of hanzi, without alphabet letters and Arabic numerals, is evenly distributed within a mesh-like surface. The perpendicular lines defining the grid are easily identified by scanning the image from side to side, both vertically and horizontally: any row or column with nearly no foreground elements can be considered part of the grid.

With the grid established, the preprocessing routine computes the average size of the individual squares within the mesh (A_{avg}). Then, it scans the image and lists the valid hanzi separately. The validation is true only if the mesh locus meets the following criteria:

1. It contains foreground pixels;
2. Its area A_i is within the 30% error range of A_{avg}.

The first rule eliminates leading and trailing spaces, whereas the second guarantees that the square has a full-width character size. The threshold of 30% was empirically defined, based on the punctuation and line spacing. Some punctuation symbols, such as ellipses and em-dash, stretch across two full-width character spaces, and neighboring, big, complex hanzi can merge, eliminating the local grid. Such cases require special care: if A_i is twice or thrice the size of A_{avg}, the locus is split into two or three equal parts, and the resulting subregions are validated with the rules 1 and 2 above again. Lastly, all validated loci are rescaled and stacked into a $28 \times 28 \times n$ array, where n equals the number of characters within the image.

4.3 Results and discussion

4.3.1 Radical classification

Figure 4.7 shows the confusion matrix correlating predicted and true classes from the validation dataset, with known ground truth. One verifies in the main diagonal that most classes presented a high detection rate. The exception is the bùshǒu radical “sun”, with a hit rate of ~59%. Following the row of “sun”, ~11% of the samples were misclassified as belonging to “mouth”, which is a reasonable mistake because both “sun” and “mouth” occupy the same probable location as the indexing radical of a hanzi and they are similar in shape. Likewise, the samples belonging to the radical “eyes”, with a hit rate of ~70%, were most often misclassified as belonging to the bùshǒu “mouth” due to their alike shapes and frequently assumed position, as shown in Figure 4.8.

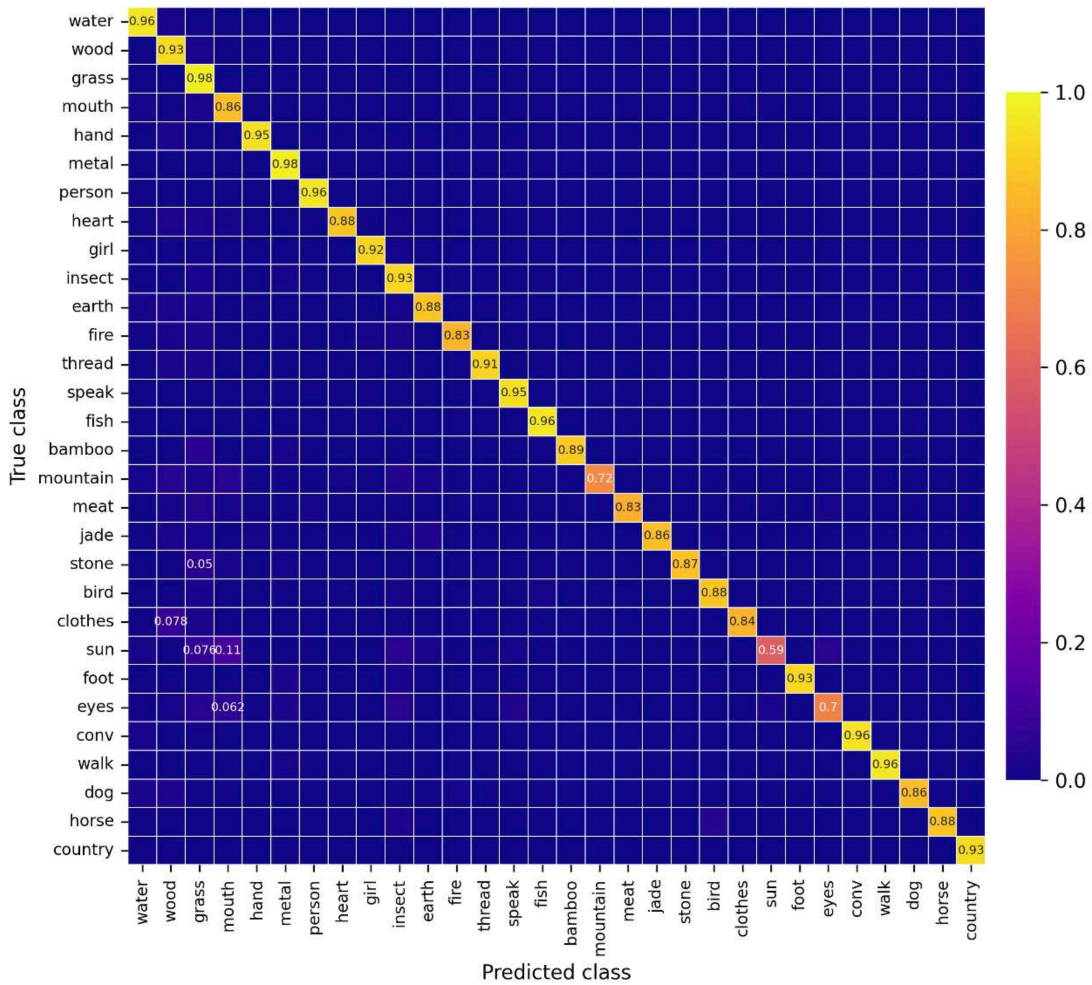

Figure 4.7: Confusion matrix of the classifier for the validation dataset.

Figure 4.8: Examples of hanzi from the bùshǒu radicals: (a) mouth, (b) sun, and (c) eyes.

Besides the hit rate for each class from the confusion matrix, the classification performance is further evaluated in terms of the F1-score, given by:

$$\text{F1-score} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} \tag{4.1}$$

where precision measures the percentage of relevant predictions among all predicted values, and recall measures the percentage of positively predicted instances among all positive samples (FAWCETT, 2006).

Figure 4.9 presents the classification F1-score in percentage for each class. The values corroborate the information in the confusion matrix, such that “sun” and “eyes” had the lowest scores (70.26% and 78.45%, respectively). Other classes showed a reasonable hit rate (above 80%), amounting to an average F1-score of ~90.0%.

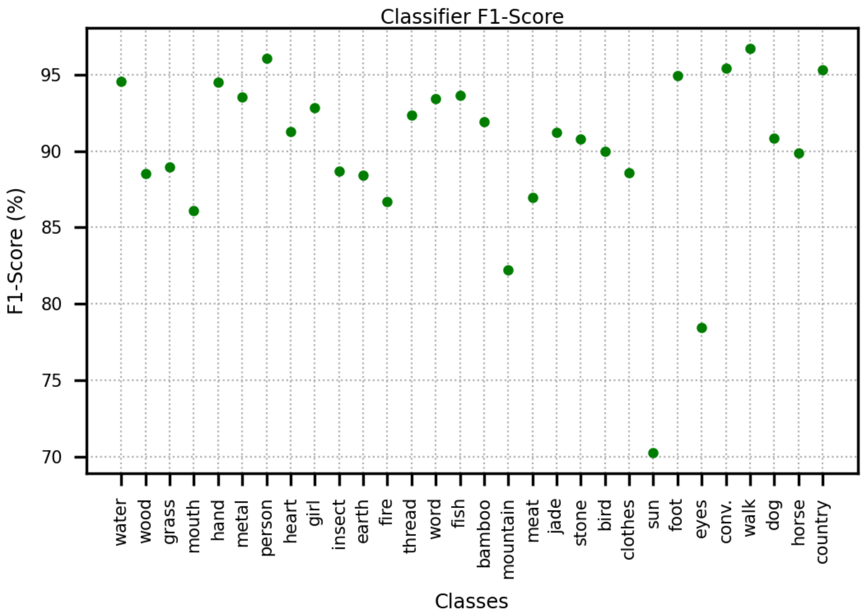

Figure 4.9: Classification F1-Score for the validation dataset.

4.3.2 Radical localization

The localizer returns an array of six floating points, normalized to the image size, where (0, 0) corresponds to the bottom-left corner and (1, 1) to the top-right corner. Therefore, its accuracy is evaluated in terms of the mean squared error (MSE), given by:

$$MSE = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \left(y_i - y_{pred}\right)^2 \tag{4.2}$$

where y_{pred} is the prediction for the i-th sample, y_i is the corresponding true value, and $n_{samples}$ is the number of samples in any given class (SCIKIT-LEARN, c2007-2023).

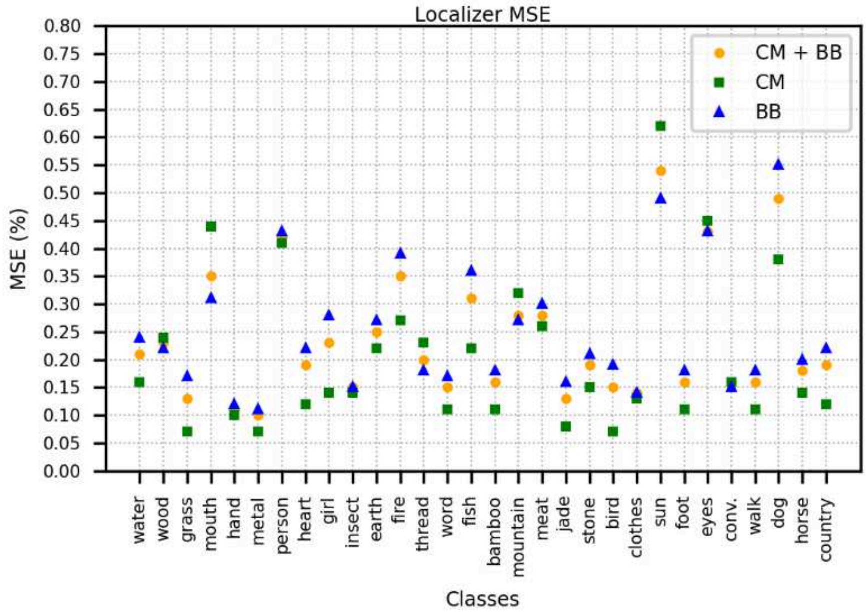

Figure 4.10: Localization mean squared error (MSE) for the validation dataset. BB: bounding box; CM: center of mass. From: Wu, Fujiwara and Suzuki (2022).

The MSE estimates how far the predicted location was from the actual one. In this experiment, its value was calculated for three scenarios, considering both the center of mass and the bounding box, only the bounding box, and only the center of mass. The values for each one are presented in Figure 4.10.

Most of the evaluated bùshǒu radicals presented lower MSE for the center of mass because it is a single point, whereas the bounding box surrounds an enclosed area. The error for the center of mass and bounding box combined is a value in between since it is averaged from the individual MSE of both, as confirmed in Figure 4.10.

Among the 30 bùshǒu radicals, the localizer had the poorest performance for “sun”, “dog”, “eyes”, “person”, and “mouth”, with over 0.35% of MSE for the center of mass. These five indexing radicals can fit into two groups: one preserves the structure but loses the default aspect ratio, and the other loses both the shape and the aspect ratio, depending on their placement within a character.

The radicals “sun”, “eyes”, and “mouth” belong to the first group and coincidentally have higher MSE for the center of mass than the bounding box. Considering the aspect ratio variation, as exemplified in Figure 4.11, plus occasional cases of the radical block repeating itself within a hanzi as bùjiàn components, it becomes more difficult to tag a single point instead of a rough area.

On the other hand, the radicals “dog” and “person” belong to the second group. Depending on their placement within the character, they can be either slim or bulky, as exemplified in Figure 4.11. Moreover, the number of strokes influences the inner proportions to preserve the overall visual harmony. Therefore, delimiting the box surrounding the radical block is more challenging than marking the center of mass.

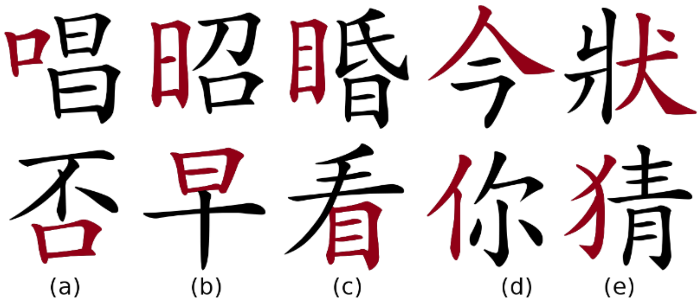

Figure 4.11: Examples of hanzi from the bushou radicals: (a) mouth, (b) sun, (c) eyes, (d) person, and (e) dog.

4.3.3 Effects of font style

A different dataset, with the basic press font, brush style, and handwriting calligraphy, isolated one from the other, was prepared to study the effect of the font style on the radical identifier. This set contains only the classes 0-14 and includes all instances of each category from the HuaKang Family instead of 33 assorted, selected fonts. Sample generation followed the same automatic procedure in Python, as previously described. The generated characters were then subdivided into training and validation samples, with an 85%-15% split ratio.

The resulting classifiers from the training samples received a name after their training font style (basic, brush, and handwriting) and were evaluated with all validation datasets in terms of F1-score, presented in Table 4.2.

As expected, they classified the best when validation samples share the training font style. From Table 4.2, the basic classifier presented acceptable results for the brush style samples (~83.0%); however, it showed a poor F1-score for handwriting (~41.8%). Conversely, the handwriting classifier achieved only ~30.9% of F1-score for the basic press font. These low values for the basic-handwriting pair are reasonable considering their characteristics. The press font is boxy, regular, and upright, whereas the handwriting, even though machine-generated, varies greatly, as shown in Figure 4.3. On the other hand, the brush style comprises both ends, combining steady, regular strokes with movement and personality. Consequently, the brush classifier had an intermediate performance for both basic (~69.4%) and handwriting (~71.7%) and was the most robust to different fonts among the three. The mix between regularity and fluidity of brush calligraphy also explains the better performance of the handwriting classifier for brush samples (~60.4%).

Table 4.2: Average F1-Score of the basic, brush, and handwriting classifiers for the validation dataset from different font styles.

Model	Validation Dataset		
	Basic	Brush	Handwriting
Basic	~98.6%	~69.4%	~30.9%
Brush	~83.0%	~96.8%	~60.4%
Handwriting	~41.8%	~71.7%	~91.8%

4.3.4 General Test

The general test data consists of 743 hanzi, but only ~36.6% belong to one of the 30 known radicals. Figure 4.12 shows the output of the radical identifier for the first hundred hanzi of the general test data. The numbers above the hanzi indicate the classification output. The blue and red numbers correspond respectively to the correct and wrong classifications. The prediction equals the ground truth for the former, so it shows only one value. On the other hand, the red numbers come in pairs: the first is the predicted class, and the second is the ground truth. The black numbers belong to instances of unknown bùshǒu radicals, so they replaced the second value, corresponding to the ground truth, with a question mark.

The radical identifier is evaluated separately from the classification and localization perspectives. The classifier reached an average F1-score of ~76.1% among the recognizable samples, and the localizer was evaluated case by case. Observing the localization results in Figure 4.12, one concludes that there are a few possible outcomes.

Figure 4.13 shows six correctly classified cases from the excerpt. The top row is clipped directly from Figure 4.12, and the bottom row shows the clear hanzi with the bùshǒu radical marked in red. Comparing the localization output with the ground truth, one notes that the localizer can tag the precise location of the indexing radical, as shown in Figure 4.13(a), but it can miss it, too. For instance, two hanzi in Figure 4.13(b) have the correct bounding box, but the center of mass is slightly off; thus, the localization output is still acceptable. However, the marking is completely off in the second example in Figure 4.13(b), considering that the localizer practically points to a bùjiàn component instead.

Figure 4.12: Radical identification results for the first hundred characters of the general test excerpt.

Figure 4.13: Case study of localization results for correct classifications. The radical block is marked (a) in its precise location or (b) with some offset.

Figure 4.14(a), in its turn, shows a correct classification and localization output. Coincidentally, a similar hanzi is in the excerpt and was misclassified. On a closer look, the character in Figure 4.14(b) belongs to class 3 ("mouth") but was misclassified into class 6 ("person") since its topmost bùjiàn component (in blue) is the same for "people", both as an individual hanzi and indexing radical. Given the serialized architecture of the bùshǒu identifier, the localizer will look for "person" in Figure 4.14(b) instead of "mouth", and the bounding box and the center of mass match the result in Figure 4.14(a) indeed.

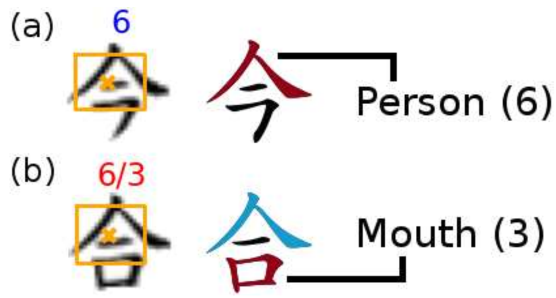

Figure 4.14: Case study of localization results for (a) correct and (b) incorrect classifications. The red strokes refer to the actual radical block, and the blue strokes mark the bùjiàn component "person" inside the misclassified character.

Whereas Figure 4.14 shows an example of characters whose components coincide with existing indexing radicals, Figure 4.15 illustrates misclassification deriving from a peculiar interpretation of bùjiàn components. Figure 4.15(a) shows the identification output, the clear version with the true radical in red, and the strokes resulting in the wrong class label in blue. Comparing the blue strokes with typical examples of hanzi from the mislabeled indexing radicals (Figure 4.15(b) and Figure 4.15(c)), one verifies that parts of the hanzi resemble the wrong class, and the localizer is efficient in finding what it is told to.

The same behavior repeats itself for the unknown bùshǒu radicals, exemplified in Figure 4.16. Again, the localizer tries to find the expected stroke configuration as informed by the classifier and succeeds in the task, as verified by comparing Figure 4.16(b) and Figure 4.16(c).

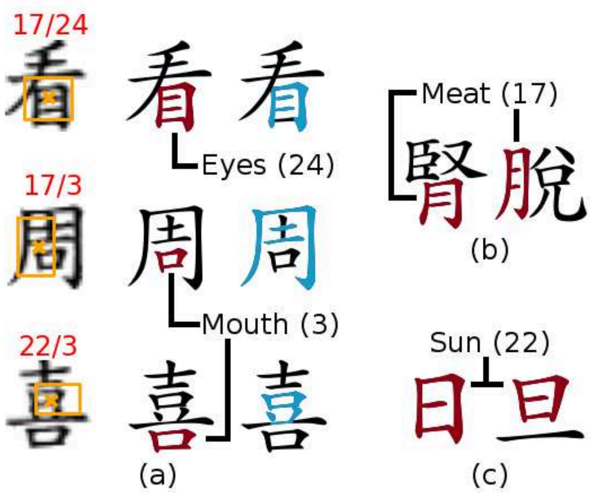

Figure 4.15: Case study of localization results for incorrect classifications. (a) The red strokes refer to the actual radical block, and the blue strokes mark the strokes misguiding the classifier, compared to examples of hanzi from bùshǒu radical (b) meat and (c) sun.

Figure 4.16: Case study of localization results for unknown radicals, with (a) the red strokes referring to the actual radical block, (b) the blue strokes marking the strokes misguiding the classifier, and (c) the corresponding predicted radical.

From these examples, one concludes that, although the localizer sometimes lacks in pinpointing the radical block, its performance is severely affected by the classification output. Therefore, enhancing the classifier is crucial to improve the radical identifier. Some options are training the identifier with more bùshǒu radicals to reduce the number of unknown cases or adding more hidden layers to the CNN model. Another possibility is to adapt the serial processing pipeline into a parallel one, so the localization becomes independent of the classification output, or to feed the detected position back to the classifier. The combined information of which and where the indexing radical is, matched against the possible loci it usually assumes, can be used to eliminate misclassifications. For instance, all cases in Figure 4.15 returned unnatural stroke clusters combined with broken components. One should not expect additional strokes inside the "meat" cluster or the "sun" to be sandwiched between other bùjiàn. This kind of common sense might help to validate an identification result if incorporated into the algorithm.

4.4 Final remarks

This dissertation proposed a bùshǒu classifier and localizer for hanzi interpreted as images. Currently, the software supports 30 indexing radicals with the most hanzi in the Unicode CJK table. The classification and the localization routines run sequentially instead of simultaneously. Although this choice is more time-consuming, it avoids errors when one of the bùjiàn inside the hanzi matches a bùshǒu in form and placement.

The experiments proved the influence of the font style over the classification results. So, adding more font styles in the training dataset will create a more robust classifier for different situations. However, if the working dataset is specific to one font style family, this mixture might be unnecessary noise in the training.

Considering that there are 214 indexing radicals in total, the classifier and the localizer fall behind with only 30 supported bùshǒu. In the future, the software will be updated to comprehend all 214 radicals. Given the morphological similarity of the hanzi and the unequal allocation of the samples across the radicals, more complex CNN models will be tested in comparison with the two models investigated so far, with the potential aid of GPU. Then, other measures, such as data augmentation, should be taken to balance the training step because some bùshǒu have less than ten hanzi in total.

5 AUGMENTED REALITY LEARNING ENVIRONMENT

Visual mnemonics are common and effective strategies in helping one to retain new vocabulary, linking the written word with the corresponding static or dynamic image, and reinforcing memory. In this sense, an AR application was developed using ARToolKit (KATO, 1999) to provide a dynamic learning environment correlating hanzi or kanji to the object they represent and their respective pronunciation.

The AR function requires a physical tag, upon which the software renders the appointed virtual object. In this project, the AR cards are black-and-white squares with 8.2 cm width total; the outer black frame of 2.1 cm width creates a smaller square of 4 cm at the center. The Han character is printed in black on the inner white space and is scaled to occupy most of the available area to maximize the software sensitivity (Figure 5.1). The tags are all linked to a virtual object, created in the software Blender (2018), along with the pronunciation (pinyin for Chinese and romaji for Japanese).

Initially, the framework employs a webcam of 640 × 480 pixels and 30 fps, supports five characters that exist in both Chinese and Japanese to recycle the physical cards, as listed in Table 5.1, with the respective translation to English and pronunciation. In Table 5.1, the pinyin follows the established rules for the tone symbols; however, the AR software rendered the tones as numerals since this simplification is common in many digital input mechanisms in computers, smartphones, and e-dictionaries. Figure 5.2 reproduces the original AR tag, along with the corresponding virtual object.

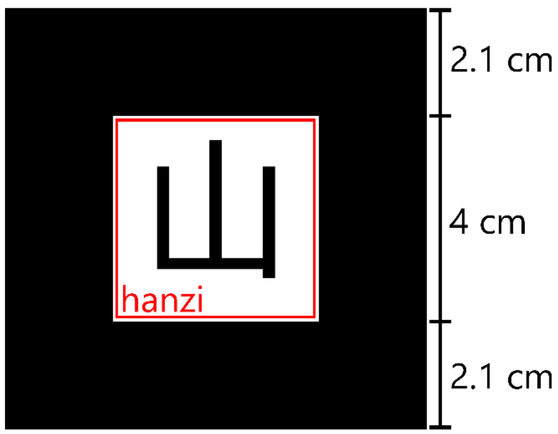

Figure 5.1. Specifications of the AR tag.

Table 5.1: Han characters supported by the AR application, with corresponding English translation, Chinese pinyin, and Japanese romaji.

Hanzi	Translation	Chinese Pinyin	Japanese Romaji
山	Mountain	*shān*	*yama*
花	Flower	*huā*	*hana*
猫	Cat	*māo*	*neko*
炎	Flame	*yán*	*honō*
犬	Dog	*gǒu*	*inu*

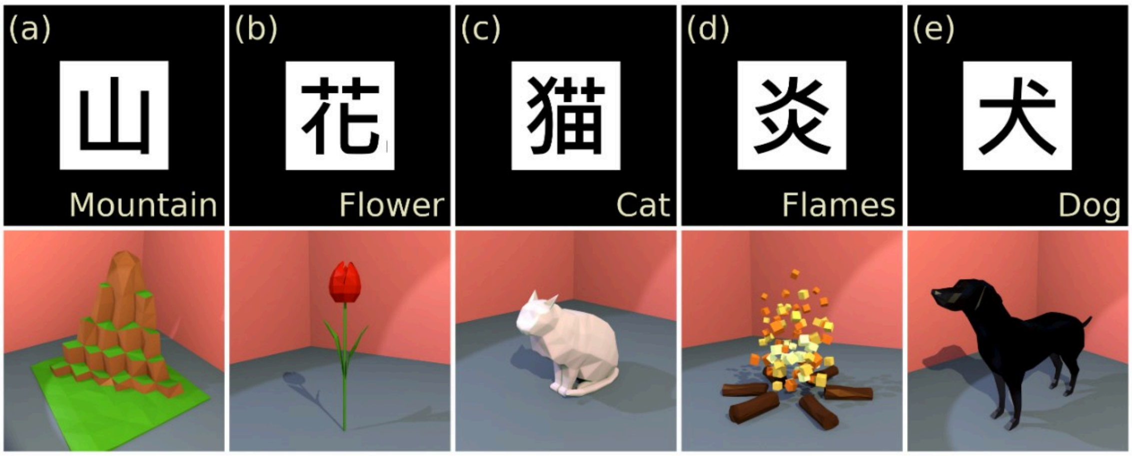

Figure 5.2: Han characters supported by the AR application, with the corresponding virtual object.

5.1 Supported functions

Characters are registered in the application as AR tags, with three virtual entities linked to them (object, pinyin, and romaji). The user runs the software with the webcam on, and the application renders the virtual object on top of the physical tag on the screen whenever it identifies a registered character. Additionally, the user can choose displayed elements among four options (Figure 5.3):

1. Only the raw image of the webcam;
2. The image of the webcam plus the virtual object on top of the character;
3. The image of the webcam plus the virtual object and the Chinese pinyin pronunciation on the bottom of the AR tag;
4. The image of the webcam plus the virtual object and the Japanese pronunciation on the bottom of the AR tag.

If the software should find an AR card but cannot recognize the character within, it renders a chromatic triangle for any of the three rendering options above (2-4). This measurement is necessary to indicate that the software correctly identified the card, though the character is unknown and thus has no associated virtual object or pronunciation.

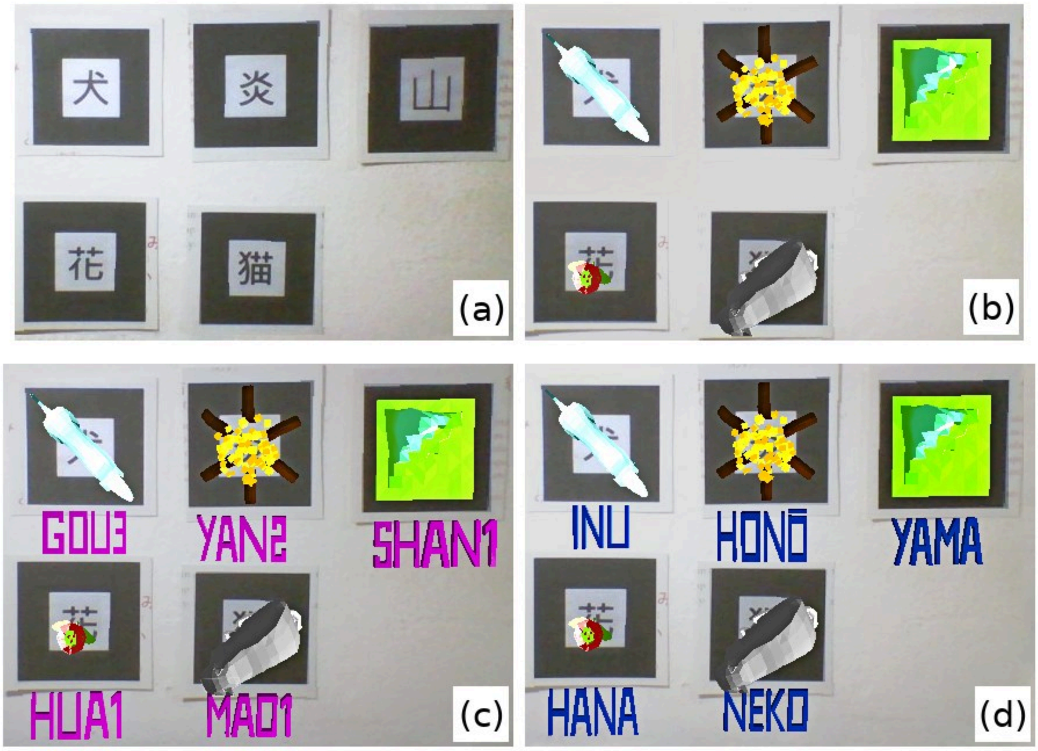

Figure 5.3: Display from the AR application: (a) with only the AR cards, (b) with the virtual objects on the cards, (c) with Chinese pinyin, or (d) Japanese romaji under the virtual objects on the cards.

5.2 Similar characters

Besides the five hanzi from Table 5.1, experiments with other characters evaluated if the software can identify unknown tags and how well it can distinguish similar characters. Table 5.2 lists the new set nested under the registered hanzi it resembles, along with the English translation, the pinyin, and the romaji. Some parts of the romaji are underlined, indicating the pronunciation of only the kanji when it needs other components (usually a hiragana) to form a common word. When it is usual to find the Japanese kanji per se without any hiragana constituents, the romaji has no underlining. Finally, Figure 5.4 shows the AR tag of the registered and corresponding unregistered similar characters.

Table 5.2: Set of hanzi resembling the supported bag-of-words, with their English translation, Chinese pinyin, and Japanese romaji.

Hanzi (Registered)	Hanzi (Similar)	Translation	Chinese Pinyin	Japanese Romaji
山	—	—	—	—
花	范 化	Model, example Transformation	*fàn* *huà*	*kata* *ka*
猫	苗 描	Sprout To trace (a drawing)	*miáo* *miáo*	*nae* *egaku*
炎	淡 火	Insipid, weak Fire	*dàn* *huǒ*	*awai* *hi*
犬	太 大	Very, extremely Big, great, elder	*tài* *dà*	*futoi* *oo*

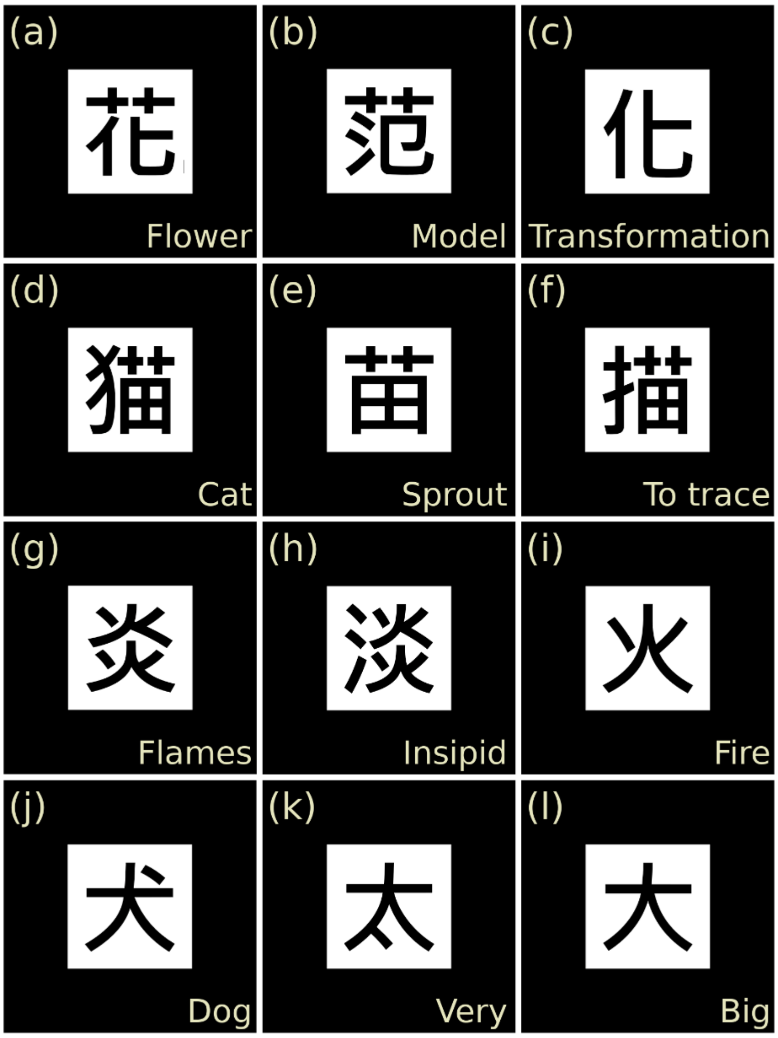

Figure 5.4: AR tags of the supported hanzi and corresponding similar characters.

5.3 Results and discussion

The software efficiently detected the AR cards and rendered the virtual elements accordingly. Moreover, the possibility of controlling what it shows on the screen, either just the flashcard with the logogram or accompanied by the virtual object with or without the pronunciation in either language, provides the flexibility of switching between learning, testing, and review sessions based on individual needs.

Figure 5.5 shows the tags “cat” and “flower” alongside the corresponding set of similar hanzi in random orientations. Similarly, Figure 5.6 and Figure 5.7 reproduce the output for “flames” and “dog” respectively. As one notes, the software detects the cards following the tag specifications and prior calibration, regardless of whether the hanzi inside is in the registered bag-of-words or not. Moreover, tag recognition does not depend on the scene orientation. However, it depends on the ambient lighting since the software employs a pixel-wise intensity threshold to identify candidates to AR cards defined as a square region with a dark outer frame and a light inner square. Also, there is a range of minimal and maximal distance between the cards and the camera for correct identification, which depends on the prior calibration to define camera and tag object parameters.

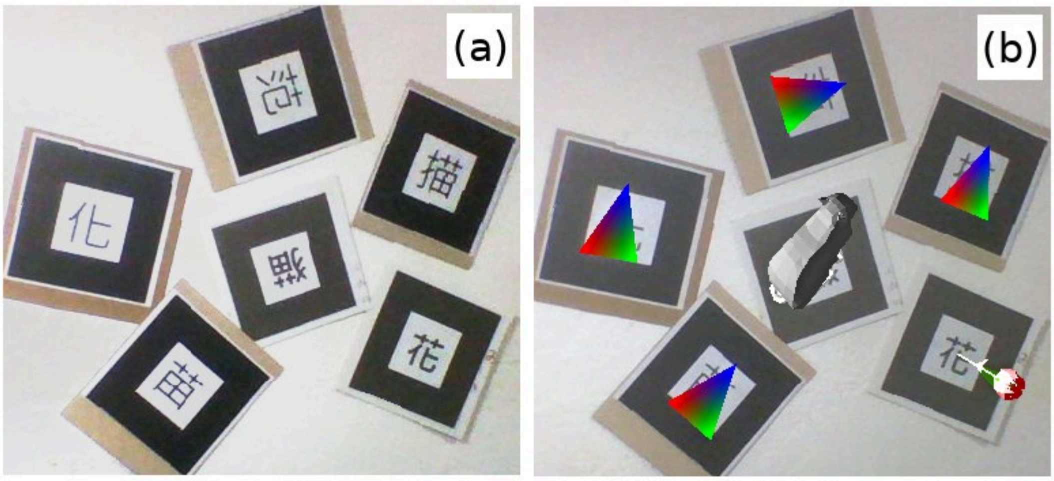

Figure 5.5: Display with (a) only the AR cards and (b) the virtual objects on AR cards, for the character set of “cat” and “flower”.

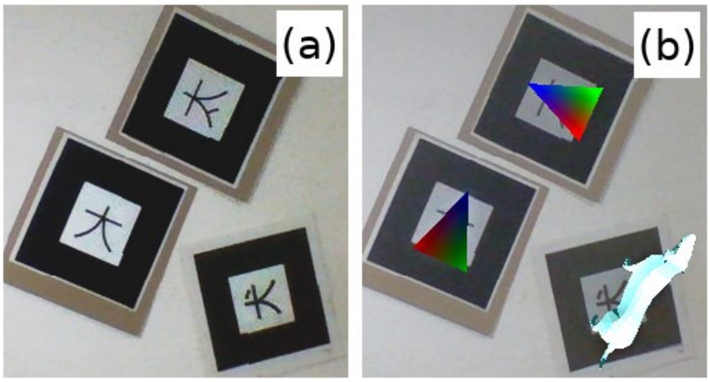

Figure 5.6: Display with (a) only the AR cards and (b) the virtual objects on AR cards, for the character set of "dog".

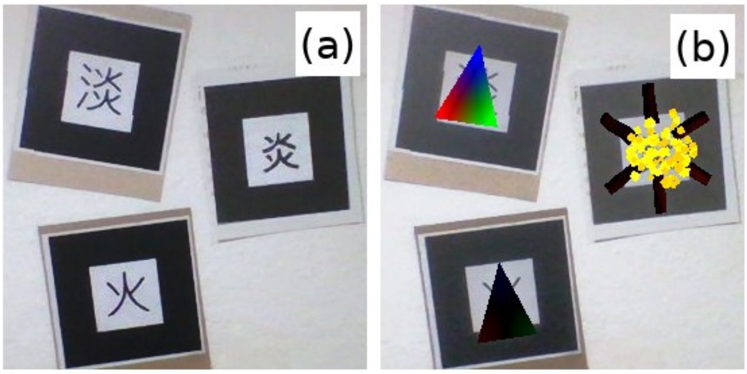

Figure 5.7: Display with (a) only the AR cards and (b) the virtual objects on AR cards, for the character set of "flames".

Both limitations are reasonable. The former is a common image-processing problem in object detection and is especially noticeable in color-based or intensity-based approaches. Ambient lighting will make the image darker or lighter for the same scenario, and the light source direction will create shades across the image. The software overcomes this problem by providing a user-input command to modify the threshold value used in the tag detection. This threshold is a single integer ranging from 0 (black) to 255 (white) and allows the user to control which gray levels are considered black. For instance, Figure 5.5(a) is converted into a grayscale intensity image first, and then all pixels below the selected threshold value are set to 0 and the remaining pixels to 255, converting thus the color image into a black-and-white binary picture. Depending on the lighting conditions and the ink level during the tag production, the black and white regions in the image will be effectively darker or lighter gray pixels. The added user-input command provides for manual adjustment of the thresholding gray level that defines them according to the scene from the webcam.

The second limitation relates to the minimal and maximal distance between the camera and physical cards and does not affect the software performance. On the contrary, it guarantees the appropriate functioning of the software. For instance, if the card is too close to the camera and the outer border becomes partially occluded, the framework will fail to recognize it. On the other hand, if it is too far from the camera, the hanzi strokes will merge due to camera resolution, so the framework will again fail to determine which AR pattern is inside the tag. Moreover, many hanzi are particularly stroke-wise busy, and the characters usually do not spam uniformly over a square area like an ordinary QR code. Instead, strokes grow outwardly from the center point preserving harmonious structural proportions so that even characters with many strokes end up converging to the same effective area inside the allocated space; thus, it is always better to give the camera a good, clear view of the patterns to maximize the detection performance.

Finally, the framework currently supports only five hanzi or kanji. Expanding the bag of words is required to provide a meaningful learning experience. Although many hanzi and kanji are structurally identical, false cognates are not rare (e.g., 木 means "wood" in Chinese and "tree" in Japanese), so the best alternative is to personalize a bag of words for each language. An option is to adopt the list of the thousand most common words or the vocabulary from the standardized proficiency tests for non-native speakers. In this sense, the physical card might need a redesign to contain multiple characters since modern Chinese mostly has compound words, and Japanese mixes kanji with hiragana. Then, the framework will need to learn the new parameters of the physical flashcards. The virtual objects, in their turn, must be individually created in Blender and registered in the AR library. Another prospect of multimodal learning is adding pronunciation recordings along with the visual representation of the sound in pinyin or romaji.

6 MUSIC VISUALIZATION

Classical music is open to interpretation by the performer. Different performers master the instrument differently and each instrument creates individual tones. Therefore, the same notes can cause different feelings depending on the instrument, the technique, the articulation, and the phrasing. Among them, phrasing is probably the most basic element to play a melody instead of just notes, and the simplest way of creating phrasing is playing with the dynamics.

Although classical music is much more complex than dynamics, considering its importance, this work proposed to create a music visualization software focused on classical music and, more specifically, dynamics. So, the task is to show graphically the dynamic variation of a piece. To accomplish this task, the software will extract the dynamic and the tempo impression of a classical piece to modulate a text, varying its font type, size, and color, similar to the study of typography to call different emotions on the reader.

This medium of displaying and visualizing music is innovative because most entertainment music visualizers focus on images, abstract graphical elements, or animations. Texts are rarely modulated to express the auditive behavior of the music, however, the field of typography is rich in expressing feelings. For example, blue is soothing, red is danger, capital letters express fury or shock, and so forth. Therefore, the proposed software explores the typography to translate the auditive feeling to a random text, so that the viewers can "read" the music.

6.1 Dynamics

The dynamics of a piece of music describe how loud are the notes executed. The main descriptions are forte and piano, for "loud" and "soft" intensities respectively. Affixes are attached to them to describe louder and softer levels. For example, fortissimo is louder than forte, and mezzo forte is slightly softer than forte, thus creating a range of intensity levels within a piece.

Within an intensity level, subtle variations add an extra flavor to the execution. For instance, a piano phrase might not share the same absolute soft intensity between the notes, i.e., a soft note can be slightly louder or softer than the other, as long as the overall effect remains as piano. However, increasing distinctively the intensity from piano to forte along the phrase creates a crescendo, and decreasing to pianissimo creates a decrescendo.

The same phrase executed as a crescendo or a decrescendo provides different impressions. Therefore, the composer usually specifies the dynamics in the score to indicate the desired effect in the execution. Common notations to indicate dynamics in a score and their respective meaning is specified in Figure 6.1.

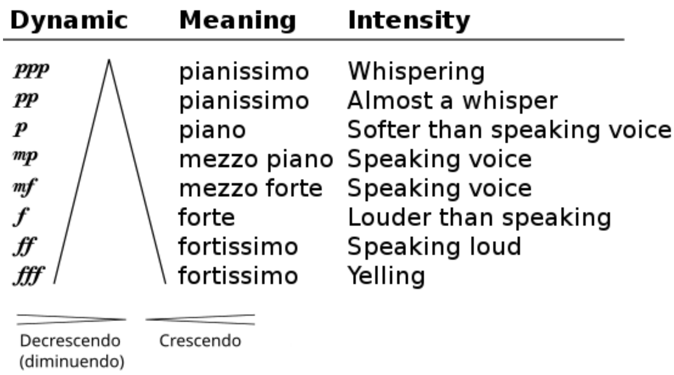

Figure 6.1. Common dynamic levels and respective variation in music notation.

6.2 Tempo

Usually, the tempo is constant and indicated at the top of the score. It marks the speed of a piece. For example, in the time signature 4/4, a tempo of 100 beats per minute (BPM) means there are 100 quarter notes within one minute, and the quarter note is considered a full-duration note. If this piece is entirely written using half notes, one gets the impression that it has half of its tempo, i.e. 50 BPM. On the other hand, if the piece is written using eighth notes, one gets the impression that it has doubled its tempo, i.e. 200 BPM. In either case, however, the default tempo is 100 BPM. From this observation, one concludes that, although the indicated tempo defines the global speed of a piece, the duration of individual notes (Figure 6.2) affects how one perceives the speed in different bars or parts within the piece.

Without prior knowledge of the established tempo and the note values in the score, one hears the music and feels its speed based on the duration of the separated notes one hears. If sequential notes share the same tune, the detached articulation marks the disconnection between them, so one perceives them as different notes. On the other hand, slurred notes require them to have different tunes to be perceived as separate notes.

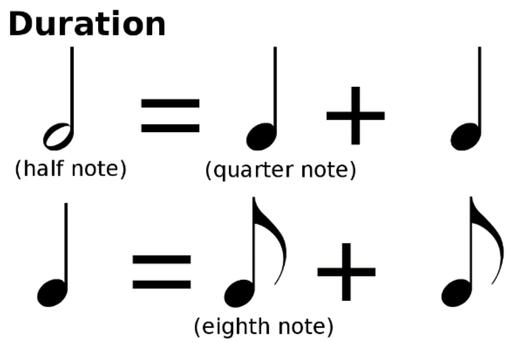

Figure 6.2. Duration of half, quarter, and eight notes.

6.3 Graphical elements

Typography causes different impressions on the readers and colors have a psychological effect on people. For instance, blue and green feel more harmonious while red and yellow are anxious or dangerous. Bold fonts and capital letters look like screaming or assertiveness. Sometimes, the handwriting gives away the personality of its owner.

In this work, typography and color are explored to express the dynamics and the tempo of classical music.

The dynamics (forte, piano, crescendo, decrescendo, and so forth) are similar to font boldness and size. In this sense, bolder and larger fonts are attributed to forte while thinner and smaller fonts are attributed to piano. Intermediate thicknesses and sizes are defined for mezzo forte and mezzo piano just like intermediate loudness is attributed to them between the forte and the piano thresholds (Table 6.1).

Table 6.1. Selected font type for each dynamic level (font size = 12 pt).

Dynamic	Font type
Forte	HELLO WORLD
Mezzo forte	Hello world
Mezzo piano	Hello world
Piano	Hello world

On the other hand, the tempo is related to the speed and the urgency of the rhythm, similar to the effect colors have over the mind. Although music pieces usually come with a predefined default tempo, this work considered the final impression of speed instead of calculating the default tempo. Faster pieces are colored in red, and slower pieces are colored in blue. This being said the estimated speed might differ from the default tempo.

6.4 Software

The proposed music visualization software was developed with Python. The audio processing element was created with the aid of the Python libraries Librosa (LIBROSA, c2013-2023) and Numpy (NUMPY, c2024), and the graphical display routine was created with the aid of the Python image library Pillow (PILLOW, c1995-2011), all of them are open source.

The input file is an MP4 audio file, which is extracted by Librosa as a vector of numbers. Subsequently, the relevant information is extracted from the numbers by manipulating them with Numpy functions.

Since the dynamic levels are relative to the maximum and minimum volume levels within the piece, the first step is to find the maximum audio intensity. However, a single point of absolute maximum value could be noise or some other form of anomaly. Besides, the forte should be a range of values instead of just a single absolute maximum, therefore the dynamic level of the forte is set at 75% of the absolute maximum, and then the threshold of the remaining three levels is computed relative to it, shown in the Table 6.2 along with the corresponding range.

Table 6.2. Intensity range of the four dynamic levels.

Dynamic level	Intensity threshold	Intensity range
Forte	$f_{default}$ = 75% of absolute maximum intensity	$mf_{max} < f$
Mezzo forte	mf_{max} = 75% of $f_{default}$	$mp_{max} < mf \leq mf_{max}$
Mezzo piano	mp_{max} = 45% of $f_{default}$	$p_{max} < mp \leq mp_{max}$
Piano	p_{max} = 15% of $f_{default}$	$0 \leq p \leq p_{max}$

The direct comparison of the instantaneous audio intensity with the dynamic level maximum and minimum thresholds gives the absolute dynamic level of that point. On the other hand, the crescendo and decrescendo behavior requires the information of neighboring points.

Therefore, the piece is divided into smaller windows. For each window, the standard deviation of the notes is computed. If the standard deviation is lower or equal to 0.04, an empirically set value, the software concludes that the notes have a similar intensity across the window frame, so it is roughly flat. For higher standard deviations, the software concludes that there was a significant difference in the dynamic level; then, it compares the intensity value at both extremities of the window. If there is an increase, the window is marked as a crescendo; otherwise, it is marked as a decrescendo.

The tempo is returned by the Librosa function "beat_track" (LIBROSA, c2013-2023). In this case, the returned tempo does not necessarily match the tempo marked in the music score but is rather the auditive impression of how many notes there are within a minute. The threshold between fast and slow is defined at 110 BPM.

For the text modulation, each one of the four dynamic levels is linked to a font type (Table 6.1), with two font sizes (48 pt and 96 pt). The forte has bolder strokes and a higher cap height, and the piano has thinner strokes and a smaller cap height. If there is a subtle dynamic variation, such as a crescendo within the piano level, the text will be printed with the piano font type, starting with the smaller font size and ending with the bigger font size. If the phrase is a continuous crescendo from piano to forte, the text will be modulated across all four font types, with both font sizes. Finally, the modulated font color will be red for a faster rhythm, and blue for a slower rhythm.

6.5 Results and discussion

The piece of orchestral music *In the Hall of the Mountain King* of Edvard Grieg is characterized by a very soft piano in the beginning, joined by other instruments and increasing in volume, culminating in a strong forte in the end. Therefore, this piece encompasses all four dynamic levels established in this work, with a continuous crescendo, allied with local crescendos and decrescendos within each dynamic level within smaller phrases. Because of this particularity, the developed music visualization routine was tested with this classical piece.

The modulated text could be any random text. If classical music had lyrics, using auditive phrasing to modulate the lyrics would be the most straightforward option. In the absence of the lyrics, the experiment was done with an excerpt from *Alice's Adventures in Wonderland* by Lewis Carroll.

In the Hall of the Mountain King has a 2:35 minute duration and the selected excerpt from *Alice's Adventures in Wonderland* has 141 words for demonstration.

Considering the duration of the piece, it was divided into 50 parts, with 3.1 s each, to analyze local crescendo and decrescendo.

Figure 6.3 shows the sound wave of the piece, and Figure 6.4 shows the excerpt modulated by it.

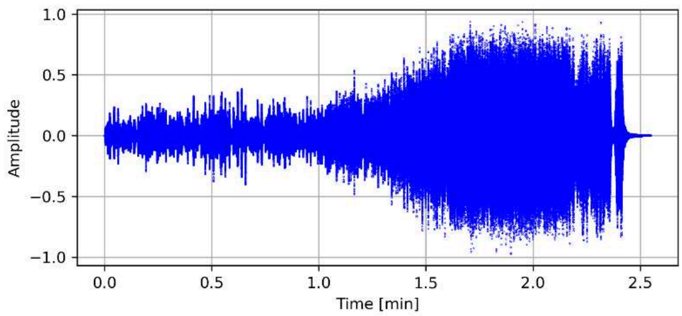

Figure 6.3. Sound wave of "In the Hall of the Mountain King".

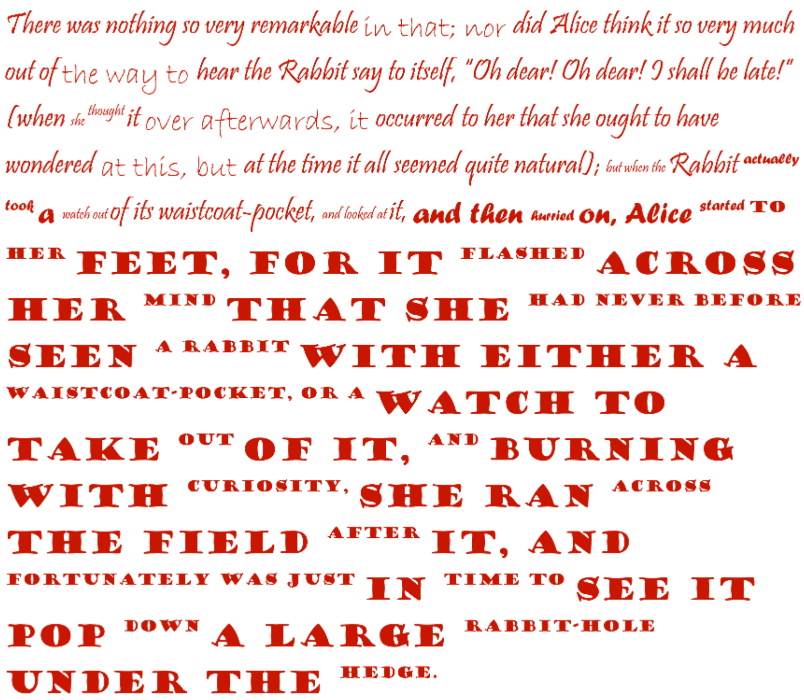

Figure 6.4. Excerpt of "Alice's Adventures in Wonderland", modulated by "In the Hall of the Mountain King".

As expected, the typography at the beginning is thinner and smaller, consistent with the font type and size attributed to the piano and the mezzo piano. As the piece progresses, the initial piano part experiences subtle crescendo and decrescendo, creating a variation between the mezzo piano and piano font types and ranging between 48 pt and 96 pt, back and forth (Figure 6.5).

Eventually, the piece becomes mezzo forte for a short while before entering the forte section (Figure 6.6). All the while, the piece has subtle crescendo and decrescendo within each section, which is reflected in Figure 6.7 as the font alternates between 48 pt and 96 pt.

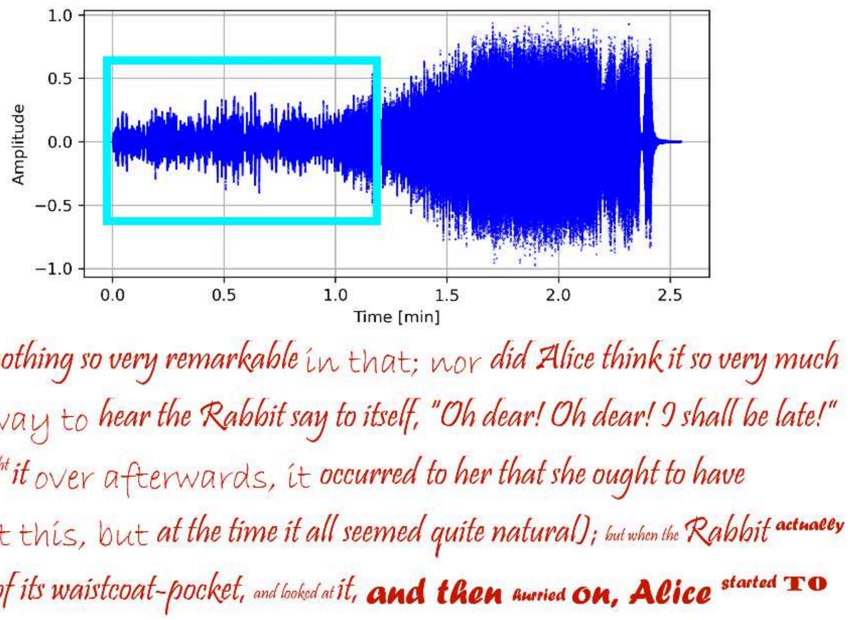

Figure 6.5. Text modulation for the piano and the mezzo piano section.

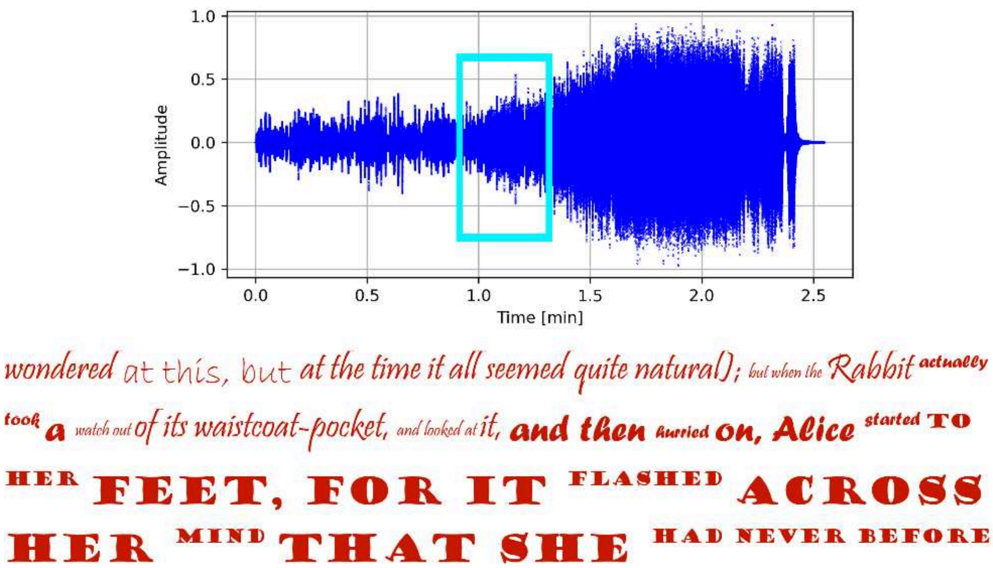

Figure 6.6. Text modulation for the mezzo forte section.

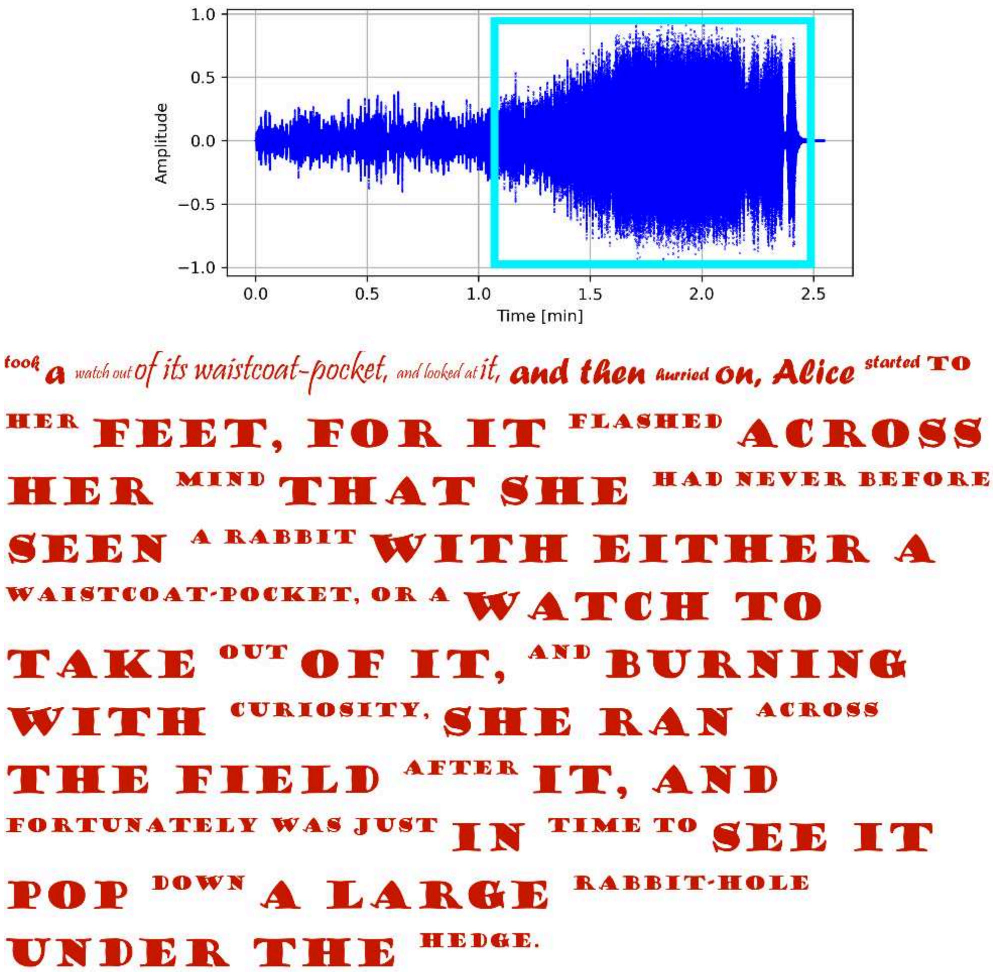

Figure 6.7. Text modulation for the forte section.

Finally, *In the Hall of the Mountain King* has a tempo of allegro and is generally performed at 120 – 156 BPM. Allegro is a fast tempo and is faster than the threshold of the software to define the font color in the text modulation (moderato, 110 BPM). Moreover, although the piece is written with a 4/4 time signature, most bars have eighth notes, meaning that they have only half of a whole duration note. The impression it creates is that the piece has twice as many notes or is twice as fast as the established tempo. Therefore, it is reasonable that the modulated text is red-colored, marking its faster nature.

From the Figure 6.4, the text modulation via classical music was successful. It can identify four dynamics levels, from piano to forte, in addition to continuous crescendo and decrescendo. However, the software has some limitations.

In classical music, a phrase usually spams across several bars and can even be subdivided into shorter phrases. Crescendos and decrescendos generally add flavor to the phrases and indicate their development. Currently, the software splits the whole input piece into smaller windows of ~3 s, regardless of the sound phrase, and checks subtle dynamic variations by comparing the intensity of the notes within each window. Although this approach can identify subtle crescendos and decrescendos, it depends on how fortunate the subdivisions are. Consider, for example, a forte phrase with a sequence of subtle crescendo and decrescendo. If this phrase is split exactly at the points where the intensity gradient is the most noticeable, the software will miss the dynamic gradient, and interpret the whole phrase as a flat forte. This problem can be solved by adopting a double-window system, with the cross-validation of the computed dynamic variation via larger and then smaller windows.

Dividing the pieces as a binary class of fast or slow, with just one threshold of 110 BPM is a little restrictive since there are many possibilities for tempo: lento (very slow), largo (slow), andante (at a walking pace), moderato (moderately), allegro (fast), presto (very quickly), and so forth. Although the beat-based measurement of tempo does not necessarily match the established tempo for the piece, the general impression of speed is proportional to the base tempo. A 4/4 largo piece with only eighth notes can sound faster than the “largo”, but the same impression applies to a 4/4 allegro piece with only eighth notes. So, more thresholds and colors can be added to the software to provide the same experience of different tempos.

Regarding the selected font types for each dynamic’s level, the stroke thickness and cap height can provide the piano-forte impression. However, the visual impression of a “loud piano” followed by a “soft mezzo piano” is not very natural, since the piano’s font in the bigger font size looks bigger than the mezzo piano’s font in the smaller font size. So, the impression of the piano evolving into a mezzo piano is not very evident. Therefore, one possibility is to explore more font types, selecting two font types per dynamics level, to create the impression of subtle crescendo and decrescendo within the dynamics level, but making sure that the bigger font size within a softer level is smaller than the smallest font size in the following level, so that the louder level will always look bigger than the softer level.

Some pieces are performed with just one or two dynamic levels, while others can specify more than four dynamic levels. However, the proposed software will always look for four dynamics levels by default. This approach is acceptable because the piano-forte definition is a relative concept. Forte is a louder sound and piano is a softer sound. Depending on the ability of the instrument to project the sound, the scenario where one is performing, the role of the musician, and the individual skills of the musician, the piano-forte varies. For example, the piano of a soloist will be louder than the piano of an orchestra, and a beginner musician will project less than an experienced musician. In this sense, the piano from a piece where the softest level is piano might be the pianissimo from a piece where the softest level is pianissimo. Therefore, it is reasonable for the visualization software to mark the intensity level variation: where are the softest or the loudest parts? And where are the levels in between?

6.6 Final remarks

The software converts the dynamic variation and the impression of tempo to text, using typography parameters. Currently, it extracts four dynamic ranges, different combinations of dynamic variation, and the impression of tempo. The identified tempo does not necessarily match the tempo in the score, because the routine to estimate the tempo is based on the number of beats, i.e., the number of individual and separated notes, which can be longer or shorter than the whole duration to count the tempo defined by the composer. In the future, more elements of classical music interpretation can be explored and implemented in the software, such as articulation, bowing techniques, etc., providing a synesthetic experience or aiding hearing-impaired audiences to feel the form of the melody.

7 User Interface for Music

7.1 Optical fiber specklegram

When coherent light travels through a multimode optical fiber, the different modes interfere with each other and create a granulated image at the output of the optical fiber. The output image is known as a “specklegram”. Any variations in the environmental parameters, such as temperature and vibration, or on the optical fiber itself, such as pressure or bending, can cause disturbances on the output specklegram. These disturbances are reversible if the original conditions on the optical fiber are recovered, and are reproducible if the optical fiber is submitted to the same set of external parameters again. This nature can be explored in creating a variety of sensors in industrial applications or user interfaces (YU, 2002).

In particular, the output specklegram varies depending on the point where the external force is applied over the optical fiber. This sensitivity is explored in this dissertation to create a user interface for musical applications using only one optical fiber. By pressing different points of the optical fiber, the output specklegram varies accordingly and can be correlated to a specific musical note, defined during a prior calibration routine.

7.2 Description of the experimental setup

The musical interface consisted of a multimode optical fiber of 3 m, laid over a rigid surface and with foam pads distributed over it. The foam pads are 2 cm × 4 cm × 0.4 cm and distributed evenly with 1 cm space in between (Figure 7.1). The optical fiber is connected to a He-Ne laser (~633 nm, 1 mW) and a webcam (15 fps). The laser runs through the multimode optical fiber and creates a specklegram pattern on another end, which is captured by the webcam and further processed to extract the state of the foam pads.

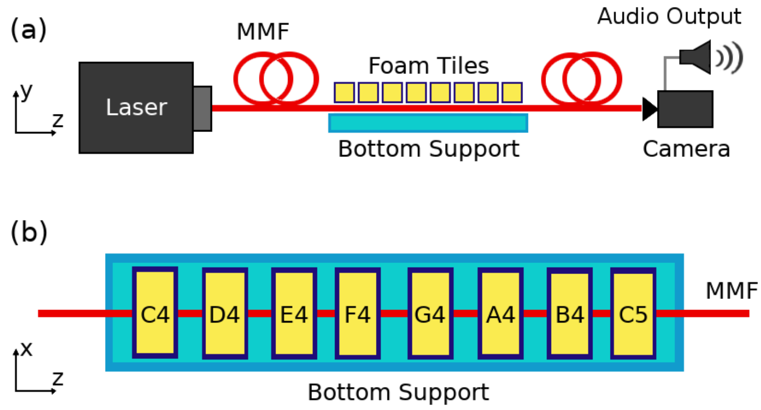

Figure 7.1. Design of the user interface for music: (a) full system, (b) placement of the foam tiles.

7.3 Selected range and instrument for experiments

The experiments focused on reproducing a keyboard-based instrument and worked with a full octave of the C major scale, varying from C4 to C5, plus a neutral state N. The notes and the corresponding frequencies are listed in Table 7.1.

Table 7.1. C major scale (C4 to C5) plus silent state (N), and their respective frequencies.

Note	Frequency (Hz)
N	0
C4	261.63
D4	293.67
E4	329.63
F4	349.23
G4	392.00
A4	440.00
B4	493.88
C5	523.25

7.4 Description of data processing software

The data processing software was developed with Python's open-source library Numpy (NUMPY, c2024), Pillow (PILLOW, c1995-2011), and Pysine (PYSINE, c2024).

The webcam captures the speckle at the end of the optical fiber. During the calibration routine, the software saves the speckle pattern at the end of the optical fiber for each state in

Table 7.1. To reduce the effects of noise and laser fluctuations, the reference image is averaged from 60 images. Once calibrated, the software is divided into two parallel routines: the image processing and the sound reproducing.

In the image processing routine, the instantaneous speckle I_t at the instant t is captured and compared with the calibration reference to evaluate the similarities between them. The similarity is given as a scalar number by:

$$r = \frac{\sum_i \sum_j (A_{ij} - \bar{A})(B_{ij} - \bar{B})}{\sqrt{\left(\sum_i \sum_j (A_{ij} - \bar{A})^2\right)\left(\sum_i \sum_j (B_{ij} - \bar{B})^2\right)}} \tag{7.1}$$

where A_{ij} is the intensity of the image A in coordinates (i, j), and $\bar{A}$ the average value of A (FUJIWARA *et al.*, 2018). In this case, I_t is the image A, and the calibration reference is the image B. According to this equation, the higher the value of r, the more similar the images A and B are. Therefore, the state with which I_t has the maximum similarity is considered the state of the fiber in the instant t.

With the state identified, it is sent to the parallel routine of sound generation and the software reproduces the corresponding frequency for 1 s or remains silent for the state N.

7.5 Results and discussion

Figure 7.2 shows the speckle output of a calibration routine and Figure 7.3 shows the similarity rate between any pair of notes from the calibration reference. Comparing them, note A4 is the least similar to other notes and the silence. Other pairs of notes share some similarities, but most of them are discernable, for the similarity rate is below 0.85. The exceptions are the pairs (N, F4) and (N, C5), with $r = 0.97$ and $r = 0.96$, respectively.

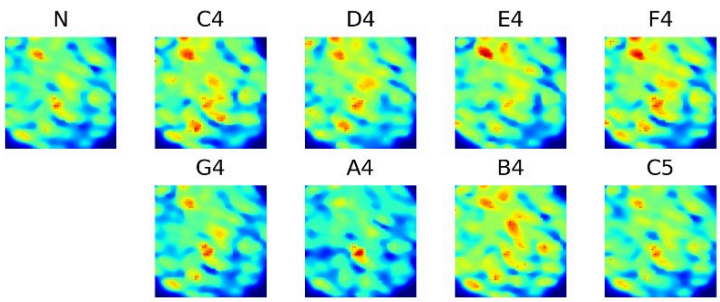

Figure 7.2. Reference image for the notes C4-C5 and N from a calibration routine.

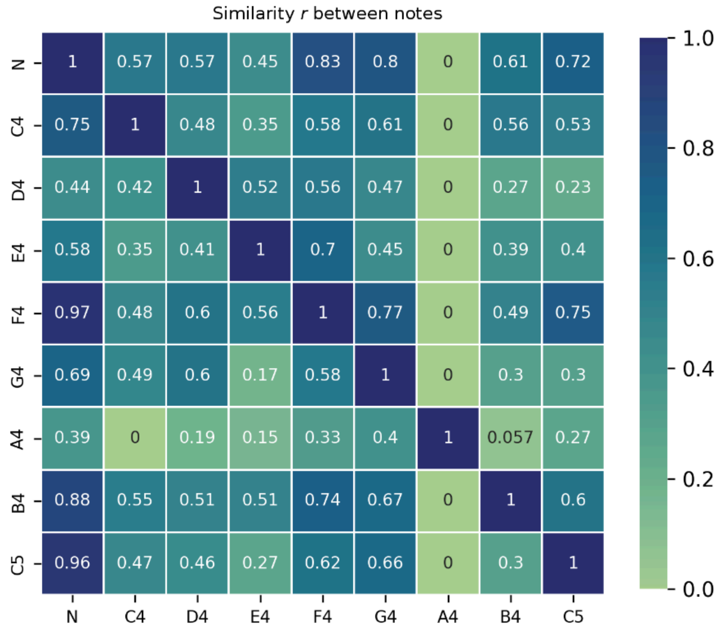

Figure 7.3. Similarity between pairs of refecence images.

Figure 7.4 shows the confusion matrix between the estimated note and the ground truth during a run of experiments. From the values in the main diagonal, most detections are true positives, with a spatial precision of 1 cm, except for the state N, 26% of which were mistaken for the note F4. Considering that the similarity rate between N and F4 is 0.97, this mistake makes sense. On the other hand, although N has a similarity rate of 0.96 with the note C5, only 6.5% of N were mistaken as C5. Therefore, the similarity between the output specklegrams is not the major reason for the wrong predictions.

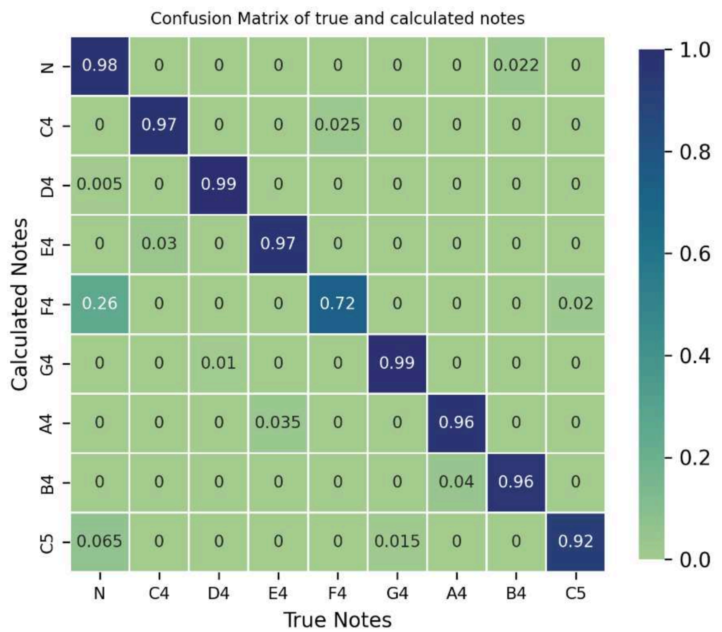

Figure 7.4. Confusion matrix of true and predited noted from similarity rate.

The mechanical factors also contribute to the false positives. Since the speckles are very sensitive to external stimuli and the optical fiber is very thin, displacements of the pressure point and variation of the force distribution on the foam pads during the calibration routine and during the actual run can result in unexpected speckle patterns, similar to or unlike any of the calibrated references. Nevertheless, the prediction will return the note with the highest similarity rate, resulting in a potential false positive.

Regarding the hardware setup, the foam pads are essential to evenly transfer and maximize the pressure to the fiber beneath them. Pressing the fiber directly over a flat surface does not cause great variations in the output speckles. Besides, given the width and transparency of the optical fiber, the users need external elements to indicate where the pressure points of each note are so that the results can be consistently reproduced. Empirically, the selected material for the foam pads can be neither too rigid nor too soft. Rigid foams hardly transfer the pressure to the optical fiber, and soft foams do not spread the pressure to maximize the sensitivity to different notes.

The foam material also affects the response time of the user interface. Figure 7.5 shows the similarity r of the images I_t when the foam E4 is pressed (blue curve) and when it is released (red curve). The red curve is shifted to match the origin of the blue curve. Considering that the foam tile has a step response to stimuli, the similarity values can be fitted to a step response to inquire about the response time. From the fitting, the current material takes 0.52 s to stabilize when pressed and 0.21 s when released, both of which are coherent with the 1 s duration of the notes in the sound generation routine.

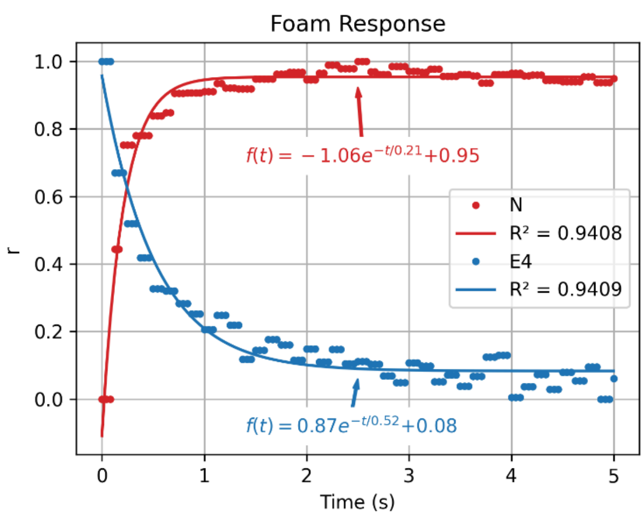

Figure 7.5. Response of the foam tile E4 when pressed (E4) and when released (N).

The sound reproducing routine runs simultaneously with the note estimation routine. The estimated notes are reproduced for 1 s, without interruptions, meaning that if the user changes the note within a 1 s interval, the software will have a delay in reaction. If the user plays three notes under 1 s, the software should miss at least the second of the three notes. This problem can be solved by reducing the note duration in the sound generation routine, but respecting the slowest response time of the foam tiles (0.52 s). To further decrease the note duration, the foam material should be changed.

7.6 Final remarks

The proposed musical interface was successfully demonstrated to eight notes of the C major scale for real-time application. The speckles are reproducible and can be used to track the input note in the form of external pressure stimuli. The note prediction is based on the computation and the direct comparison of similarity rate; therefore, it is suitable for real-time operation since the limiting factors on the response time are the note prediction, sound reproduction interval, the foam response, and the camera's frame rate. Currently, the demonstrated musical interface is able to operate with a sound resolution of 1 s, but this interval can be further reduced by software to 0.52 s at most due to hardware issues. The next step is to expand the frequency range, encompassing more notes, accidents, and other scales.

This prototype was inspired by a keyboard-based instrument, so the notes are detached. Given the string-like nature of the optical fiber and the sensitivity of the specklegrams to external stimuli, a prospect to explore this musical interface setup is to implement the vibrato and glissando, techniques of string instruments.

8 CONCLUSION

This dissertation explored the applications of visualization in second language learning and in music-related entertainment.

Regarding second language learning, Chinese was the language of choice. The first approach was the identification of the indexing radical of the Han characters because of their importance in the assimilation of the meaning. Someone unfamiliar with hanzi usually struggles to memorize the poorly connected blocks and often mistakes a character for another, especially if they are similar. Therefore, the project proposes to identify the bùshǒu radical from image files without a laborious extraction of a character into all bùjiàn components. The implemented radical identifier showed a classification rate of ~90.0% for the 30 bùshǒu radicals with the most entries in the Unicode CJK table within the validation dataset, which dropped to ~76.1% in the general test set. The localizer correctly identified stroke clusters corresponding to the bùshǒu radical in most cases; however, it depends entirely on the classification results since it looks specifically for the bùshǒu radical the classifier indicates. This identification process can be improved by training the identifier with all 214 official indexing radicals, reducing the number of unknown cases. Another possibility is to create a feedback flow, validating the classification results with the localized clusters and checking if they match the allowed placement of the bùshǒu radical. Improved, the proposed radical identifier can function as an auxiliary learning tool or enable image content organization without the laborious decomposition of the characters into all bùjiàn components.

The second approach to aid second language learning explores visual mnemonics, a common strategy to memorize new vocabulary. An AR framework developed with ARToolKit and Blender for computers simulates a flashcard-based system to learn and revise new Chinese hanzi or Japanese kanji. The user controls the displayed information on the screen: the AR cards alone or accompanied by their corresponding virtual objects and pronunciation in pinyin or romaji. The software can recognize the registered words and distinguish them from similar Han characters within its range of operation, which depends on the previous calibration routine to acquire camera parameters and learn the physical specs of the AR cards. Although visual mnemonics are suitable only for words with corresponding objects, they are popular among students. The framework, currently supporting only five hanzi or kanji, can be further expanded, even with characters specific to Chinese or Japanese, allowing the user to delve into each language. It will also provide a better learning experience if it combines the imagery with recordings of actual natives reading the characters instead of just displaying the corresponding pinyin or romaji pronunciation on screen.

Both applications need further improvement, especially a database expansion: the AR framework needs more entities inside its bag-of-words, and the radical identifier requires all 214 official bùshǒu. Even though some indexing radicals have less than ten entries in dictionaries, they all compose the daily lexicon and, thus, are equally important.

Regarding the operation platform, both applications currently run on a computer, so a promising perspective is to port them to mobile devices, given their convenience and widespread accessibility. Besides the fact that users across all ages are more likely to have a smartphone than a computer, implementing the radical identifier and the AR flashcards on Android or iOS allows the users to learn on the go, anytime, anywhere. This convenience can encourage more people to use the auxiliary learning tool from this dissertation to learn Chinese or a similar second language, accomplishing the final goal of helping the community.

In Brazil, the dominant language adopts the Roman alphabet, and the words describe their pronunciation instead of the meaning. Thus, the logographic nature of the Han characters is a challenge for Brazilians who are learning Asian languages such as Chinese and Japanese. A beginner student often thinks that the stroke clusters are random drawings, so similar hanzi become indistinguishable unless they are put side-by-side. The difficulty is aggravated because the surrounding characters might not be helpful for a beginner student to determine the true identity of the questionable hanzi within a multiple-character vocabulary, so mistakes are common. In this sense, the proposed radical identifier tags the indexing radical to increase the user's awareness of the fundamental bushou block. Thus, when encountering questionable and similar hanzi, the Brazilian user can learn to identify the bushou radical and deduce the meaning of the hanzi, just like native speakers do unconsciously.

On the other hand, the AR flashcard application explores the graphical objects, correlating images to hanzi, and extrapolating the word-to-word translation. The integration of visual arts and Chinese characters provides a ludic environment to memorize characters. Considering that learning and distinguishing Han characters is a challenging process for Brazilian students of Chinese, especially because the Chinese writing system rarely uses phonetic-based writing systems in text (unlike Japanese, which can be converted to hiragana or katakana, and Korean, which adopts mainly hangul), different forms of stimuli on the user's senses to memorize the hanzi are welcome. The proposed AR flashcard stimulates the sight with 3D objects and respective transliteration. In the future, the sound can be integrated to stimulate hearing too, so that Brazilian users can train reading and hearing simultaneously.

This dissertation also developed a classical music visualization software, endeavoring to show the phrasing of the input music piece. Currently, the dynamics and tempo information are extracted from the classical music piece to modulate a text with different typographies. The identified tempo is actually how fast or slow the piece sounds instead of the actual tempo marked in the score. The routine was demonstrated for four dynamics levels: piano, mezzo piano, mezzo forte, and forte. Each level is correlated to a font type, whose thickness and height cap are defined by the sound intensity. Variations from one level to another, such as piano to mezzo piano, will be visualized as a variation from a thinner font type to a bolder font type. Subtle dynamic variations, such as a softer piano to a louder piano, reflect themselves as font size variation, i.e., the same font type will increase in size. However, currently, there are only two font sizes in the visualization routine, thus the evolution from a softer to a louder intensity level in a crescendo section looks a little unnatural because the last word from the softer intensity level looks bigger than the first word from the louder intensity level. This problem can be resolved by attributing a specific pair of font sizes for each dynamic level, noting that the louder levels will always have bigger font sizes than the softer levels. On top of adding more font sizes, the visualization routine can take advantage of more speed thresholds to comprehend the various tempos in the classical music universe.

The last project this dissertation worked on was an optical fiber specklegram-based user interface to generate sound. The system was successfully demonstrated for the C major scale (C4 to C5), with a silent state (N). The optical fiber specklegram is sensitive to the geographical localization of the pressure point on the optical fiber, therefore the user interface supports multiple notes with a single optical fiber. For larger frequency ranges, the system can be updated to have more optical fibers or be extended to accept combined pressure points on the same optical fiber. On the same thought of the geographic sensitivity, the specklegrams are adequate to reproduce some effects of string instruments, such as vibrato and glissando. In the future, the project can be updated to support more pitches, timbres, and scales. Currently, the sound generation routine keeps an input note for 1 s, consistent with the response time of 0.52 s. For shorter notes, a possibility is to decrease the generated sound’s duration or to modify the system hardware to reduce the response time.

REFERENCES

ABDULRAHAMAN, M. D.; FARUK, N.; OLOYEDE, A. A.; SURAJUDEEN-BAKINDE, N. T.; OLAWOYIN, L. A.; MEJABI, O. V.; IMAM-FULANI, Y.O.; FAHM, A. O.; AZEEZ, A.L. Multimedia tools in the teaching and learning processes: a systematic review. **Heliyon**, v. 6, n. 11, e05312, 2020. DOI: https://doi.org/10.1016/j.heliyon.2020.e05312.

ABDALLAH, M.; PRABHAKARAN, B.; CAI, W.; HSU, C.-H. Recent advances in immersive multimedia. **IEEE Multimedia**, v. 30, n. 2, p. 5-7, 2023. DOI: https://doi.org/10.1109/MMUL.2023.3280408.

AMY KUO, M. L.; HOOPER, S. The effects of visual and verbal coding mnemonics on learning Chinese characters in computer-based instruction. **Educational Technology Research and Development**, v. 52, n. 3, p.23-34, 2004. DOI: https://doi.org/10.1007/BF02504673.

ANKI. 2006. Available online: https://apps.ankiweb.net. Accessed on: July 12, 2023.

BLENDER. Version 2.79b. 2018. Available online: https://www.blender.org. Accessed on: Nov. 20, 2018.

BREEN, J. Multiple indexing in an electronic kanji dictionary. *In*: WORKSHOP ON ENHANCING AND USING ELECTRONIC DICTIONARIES, 2004, Geneva. **Proceedings** […]. Stroudsburg: Association for Computational Linguistics, 2004. p. 1-7. DOI: https://doi.org/10.3115/1610042.1610044.

CHANG L. S.; KUO H. C.; SUEN J. J.-B.; YANG P. H.; HOU C.-P.; SUN H. R.; LEE Z. M.; HUANG Y.-H. Multimedia mixed reality interactive shared decision-making game in children with moderate to severe atopic dermatitis, a pilot study. **Children**, v. 10, n.3, p. 574, 2023. DOI: https://doi.org/10.3390/children10030574.

CHEN C. H.; WENG M. F.; JENG S. K.; CHUANG Y. Y. Emotion-based music visualization using photos. *In:* SATOH, S.; NACK, F.; ETOH, M. (eds) **Advances in Multimedia Modeling.** MMM 2008. Lecture Notes in Computer Science, vol 4903. Heidelberg: Springer, 2008. DOI: https://doi.org/10.1007/978-3-540-77409-9_34.

CHEN, H. C.; HSU, C. C.; CHANG, L. Y.; LIN, Y. C.; CHANG, K. E.; SUNG, Y. T. Using a radical-derived character e-learning platform to increase learner knowledge of Chinese characters. **Language Learning & Technology**, v. 17, n. 1, p. 89-106, 2013. DOI: https://doi.org/10125/24511.

CHEN, M. P.; WANG, L. C.; CHEN, H. J.; CHEN, Y. C. Effects of type of multimedia strategy on learning of Chinese characters for non-native novices. **Computers & Education**, v. 70, p. 41-52, 2014. DOI: https://doi.org/10.1016/j.compedu.2013.07.042.

CHEN, H. R.; LIN, Y. S. An examination of digital game-based situated learning applied to Chinese language poetry education. **Technology, Pedagogy and Education**, v. 25, n. 2, p. 171-186, 2016. DOI: https://doi.org/10.1080/1475939X.2015.1007077.

CHOLLET, F. *et al.* **Keras**. 2015. Available online: https://keras.io. Accessed on: Mar. 10, 2022

CHU, C.; NAKAZAWA, T.; KUROHASHI, S. Chinese characters mapping table of Japanese, traditional Chinese and simplified Chinese. *In*: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 8th, 2012, Istanbul. **Proceedings** […]. Istanbul: European Language Resources Association, 2012. p. 2149-2152.

CHUANG, H. Y.; KU, H. Y. The effect of computer-based multimedia instruction with Chinese character recognition. **Educational Media International**, v. 48, n.1, p.27-41, 2011. DOI: https://doi.org/10.1080/09523987.2011.549676.

CIUHUA, P.; KLEMENC, B.; SOLINA, F. Visualization of concurrent tones in music with colours. *In:* ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 18th, 2010, Firenze. **Proceedings** […]. New York: ACM, 2010. p. 1677-1680. DOI: https://doi.org/10.1145/1873951.1874320.

COHEN, A. D. The use of verbal and imagery mnemonics in second-language vocabulary learning. **Studies in Second Language Acquisition**, v. 9, n. 1, p. 41-61, 1987. DOI: https://doi.org/10.1017/S0272263100006501.

DUOLINGO. 2012. Available online: https://en.duolingo.com. Accessed on: July 12, 2023.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861-874, 2006. DOI: https://doi.org/10.1016/j.patrec.2005.10.010.

FELS, S.; KAASTRA, L.; TAKAHASHI, S.; MCCRAIG, G. Evolving Tooka: from experiment to instrument. *In:* INTERNATIONAL CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 2004, Hamamatsu. **Proceedings** […].p. 1-6. DOI: https://doi.org/10.5281/zenodo.1176595.

FENG, S. Prosodically constrained localizers in classical and modern Chinese. *In:* XU, D.; FU, J. (eds) **Space and quantification in languages of China**. Cham: Springer, 2015. p.17-35. DOI: https://doi.org/10.1007/978-3-319-10040-1_2.

FONTELES, J. H.; RODRIGUES, M. A. F.; BASSO, V. E. D. Creating and evaluating a particle system for music visualization. **Journal of Visual Languages & Computing**, v. 24, n. 6, p. 472-482, 2013. DOI: https://doi.org/10.1016/j.jvlc.2013.10.002.

FRID, E. Accessible digital musical instruments: a review of musical interfaces in inclusive music practice. **Multimodal Technologies and Interact**, v. 3, n. 3, p. 57, 2019. DOI: https://doi.org/10.3390/mti3030057.

FUJIWARA, E.; RI, Y.; WU, Y. T.; FUJIMOTO, H.; SUZUKI, C. K. Evaluation of image matching techniques for optical fiber specklegram sensor analysis. **Applied Optics**, v. 57, n. 33, p. 9845-9854, 2018. DOI: https://doi.org/10.1364/AO.57.009845.

GOTO, M. An audio-based real-time beat tracking system for music with or without drum-sounds. **Journal of New Music Research**, v. 30, n. 2, p. 159-171, 2001. DOI: https://doi.org/10.1076/jnmr.30.2.159.7114.

HAGIWARA, A. The role of phonology and phonetics in L2 kanji learning. **The Modern Language Journal**, v. 100, n. 4, p. 880-897, 2016. DOI: https://doi.org//10.1111/modl.12350.

HAN, J. Hanzi: the foundation of Chinese culture. *In*: HAN, J. **Theorising culture:** a Chinese perspective, 1st ed. Cham: Palgrave Pivot, 2020. p. 35-52. DOI: https://doi.org/10.1007/978-3-030-23880-3_3.

HAN, K.; YOU, W.; SHI, S; SUN, L. Hearing with the eyes: modulating lyrics typography for music visualization. **The Visual Computer**, v. 40, p. 8345-8361, 2024. DOI: https://doi.org/10.1007/s00371-023-03239-5.

HO, C. S. H.; NG, T. T.; NG, W. K. A "radical" approach to reading development in Chinese: the role of semantic radicals and phonetic radicals. **Journal of Literacy Research**, v. 35, n. 3, p. 849-878, 2003. DOI: https://doi.org/10.1207/s15548430jlr3503_3.

HOLLINGER, A.; STEELE, C.; PENHUNE, V.; ZATORRE, R.; WANDERLEY, M. fMRI-compatible electronic controllers. *In:* CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 2007, New York. **Proceedings** […]. New York: ACM. p. 246-249. DOI: https://doi.org/10.1145/1279740.1279790.

HOLLINGER, A. D.; WANDERLEY, M. M. The design, implementation, and testing of a sensorized MRI-compatible cello. **IEEE Sensors Journal**, v. 15, n. 11, p. 6125-6134, 2015. DOI: https://doi.org/10.1109/JSEN.2015.2449876.

HOW to study Korean. Copyright 2023. Available online: https://www.howtostudykorean.com. Accessed on: July 12, 2023.

HU, H.; DU, X. Radical features for Chinese text classification. *In*: INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY, 9th, 2012, Chongqing. **Proceedings** […]. Chongqing: IEEE, 2012. p. 720-724. DOI: https://doi.org/10.1109/FSKD.2012.6234029.

HUANG, Y.; HE, M.; JIN, L.; WANG, Y. RD-GAN: few/zero-shot Chinese character style transfer via radical decomposition and rendering. *In*: EUROPEAN CONFERENCE ON COMPUTER VISION, 16th, 2020, Glasgow. **Proceedings** […]. Cham: Springer, 2020. p. 156-172. DOI: https://doi.org/10.1007/978-3-030-58539-6_10.

HUANG, L.; PENG, Y.; WANG, H.; WU, Z. Statistical part-of-speech tagging for classical Chinese. *In*: INTERNATIONAL CONFERENCE ON TEXT, SPEECH AND DIALOGUE, 5th, 2002, Brno. **Proceedings** […]. Brno: Springer, 2002. p. 115-122. DOI: https://doi.org/10.1007/3-540-46154-X_15.

JORDÀ, S. The Reactable: tangible and tabletop music performance. *In:* CONFERENCE ON HUMAN FACTORS AND INTERACTIONS: EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2010, Atlanta. **Proceedings** […]. New York: ACM. p. 2989-2994. DOI: https://doi.org/10.1145/1753846.1753903.

KANG, L.; CHIEN, H. Y. Hé: calligraphy as a musical interface. *In:* CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 2010, Sydney. **Proceedings** […]. p. 352-355. DOI: https://doi.org/10.5281/zenodo.1177819.

KANGXI zidian wangshangban [KANGXI dictionary (web version)]. Copyright 2005-2023. Available online: https://www.kangxizidian.com. Accessed on: July 12, 2023.

KATO, H. **ARToolKit**. 1999. Available online: https://www.hitl.washington.edu/artoolkit.html. Accessed on: July 12, 2023.

KE, X. X. **Wo bu zai du cuobiezi** [I do not misspell anymore]. Taipei: Zhongwen Publishing House, 1992. 253 p.

KE, Y.; HAGIWARA, M. Radical-level ideograph encoder for RNN-based sentiment analysis of Chinese and Japanese. *In*: ASIAN CONFERENCE ON MACHINE LEARNING, 9th, 2017, Seoul. **Proceedings** […]. Seoul: PMLR, 2017. p. 561-573. DOI: https://doi.org/10.48550/arXiv.1708.03312.

KEEHL, O. G.; MELCER, E. F. Music to my ears: developing kanji stroke knowledge through an educational music game. **Multimodal Technologies and Interaction**, v. 5, n. 12:83, 2021. DOI: https://doi.org/10.3390/mti5120083.

KHULUSI, R.; KUSNICK, J.; MEINECKE, C.; GILLMANN, C.; FOCHT, J.; JÄNICKE, S. A survey on visualizations for musical data. **Computer Graphics Forum**, v. 39, n. 6, p. 82-110, 2020. DOI: https://doi.org/10.1111/cgf.13905.

KIM, C.; KIM, J. S.; KIM, U. J. A study on features for improving performance of Chinese OCR by machine learning. *In*: HIGH PERFORMANCE COMPUTING AND CLUSTER TECHNOLOGIES CONFERENCE, 3rd, 2019, Guangzhou. **Proceedings** […]. New York: ACM, 2019. p. 51-55. DOI: https://doi.org/10.1145/3341069.3342991.

KNÖSCHE, T. R.; NEUHAUS, C.; HAUEISEN, J.; ALTER, K.; MAESS, B.; WITTE, O. W.; FRIEDERICI, A. D. **Human Brain Mapping**, v. 24, n. 4, p. 259-273, 2005. DOI: https://doi.org/10.1002/hbm.20088.

LEE, C. H.; KALYUGA, S. Effectiveness of on-screen pinyin in learning Chinese: an expertise reversal for multimedia redundancy effect. **Computers in Human Behavior**, v. 27, n. 1, p. 11-15, 2011. DOI: https://doi.org/10.1016/j.chb.2010.05.005.

LI, H.; CHEN, H. C. Processing of radicals in Chinese character recognition. *In:* CHEN, H. C. (ed) **Cognitive processing of Chinese and related Asian languages**. Hong Kong: Chinese University Press, 1997. cap. 8, p. 141-160.

LI, J.; ZHOU, J. Chinese character structure analysis based on complex networks. **Physica A: Statistical Mechanics and its Applications**, v. 380, p. 629-638, 2007. DOI: https://doi.org/10.1016/j.physa.2007.02.059.

LI, S. P. D.; LAW, S. P.; LAU, K. Y. D.; RAPP, B. Functional orthographic units in Chinese character reading: are there abstract radical identities? **Psychonomic Bulletin & Review**, v. 28, p. 610-623, 2021. DOI: https://doi.org/10.3758/s13423-020-01828-2.

LIBROSA Version 0.10.2 Documentation. Copyright 2013-2023. Available online: https://librosa.org/doc/latest/index.html. Accessed on: December 15, 2024.

LIMA, H. B.; DOS SANTOS, C. G. R.; MEIGUINS, B. S. A survey of music visualization techniques. **ACM Computing Surveys**, v. 54, n. 7, p. 1-29, 2021. DOI: https://doi.org/10.1145/3461835.

LIN, Y. S.; LIM, J. N.; WU, Y. S. Developing and applying a Chinese character learning game app to enhance primary school students' abilities in identifying and using characters. **Education Sciences**, v. 12, n. 3, p. 189, 2022. DOI: https://doi.org/10.3390/educsci12030189.

LINDQVIST, C. **China: empire of living symbols**. Translator: Joan Tate. Reading: Addison-Wesley, 1991. 424 p. ISBN 978-0201570090.

LIU, H. C. Using eye-tracking technology to explore the impact of instructional multimedia on CFL learners' Chinese character recognition. **The Asia-Pacific Education Researcher**, v. 30, n. 1, p. 33-46, 2021. DOI: https://doi.org/10.1007/s40299-020-00512-2.

LIU, M.; XIANG, J.; XIA, X.; HU, H. Contrastive learning between classical and modern Chinese for classical Chinese machine reading comprehension. **ACM Transactions on Asian and Low-Resource Language Information Processing**, v. 22, n. 2, p. 1-22, 2022. DOI: https://doi.org/10.1145/3551637.

LUO, G. **Sanguo yanyi** [Romance of the three kingdoms]. Tainan: ACME Cultural, 1991 (reprint). 929 p. ISBN 957-517-105-5.

LÜ, C.; KODA, K.; ZHANG, D.; ZHANG, Y. Effects of semantic radical properties on character meaning extraction and inference among learners of Chinese as a foreign language. **Writing Systems Research**, v. 7, n. 2, p. 169-185, 2015. DOI: https://doi.org/10.1080/17586801.2014.955076.

LYONS, M. J.; TETSUTANI, N. Facing the music: a facial action controlled musical interface. *In:* CONFERENCE ON HUMAN FACTORS AND INTERACTIONS: EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2001, Seattle. **Proceedings** […]. New York: ACM. p. 309-310. DOI: https://doi.org/10.1145/634067.634250.

LYONS, M.; FELS, S. How to design and build new musical interfaces. *In:* SIGGRAPH ASIA 2015 COURSES, 2015, Kobe. New York: ACM, 2015. p. 1-137. DOI: https://doi.org/10.1145/2818143.2818145.

MEDIA 5 CORPORATION. **Kanji no Yakata to Obake-tati** [KANJI mansion and ghosts]. 2019. Available online: https://www.media-5.co.jp/switch/gureko/kanji. Accessed on: July 12, 2023.

NUMPY Version 2.2.0. Copyright 2024. Available online: https://numpy.org/. Accessed on: December 15, 2024.

PAIVIO, A.; DESROCHERS, A. Mnemonic techniques in second-language learning. **Journal of Educational Psychology**, v. 73, n. 6, p. 780-795, 1981. DOI: https://doi.org/10.1037/0022-0663.73.6.780.

PILLOW (PIL Fork) Version 11.0.0. Copyright 1995-2011. Available online: https://pypi.org/project/pillow/. Accessed on: December 15, 2024.

PLECHER, D. A.; EICHHORN, C.; KINDL, J.; KREISIG, S.; WINTERGERST, M.; KLINKER, G. Dragon tale - a serious game for learning Japanese kanji. *In*: ANNUAL SYMPOSIUM ON COMPUTER-HUMAN INTERACTION IN PLAY, 2018, Melbourne. **Proceedings** […] **Companion Extended Abstracts**. New York: ACM, 2018. p. 577-583. DOI: https://doi.org/10.1145/3270316.3271536.

POURIS, M.; FELS, D. I. Creating an entertaining and informative music visualization. *In:* MIESENBERGER, K.; KARSHMER, A.; PENAZ, P.; ZAGLER, W. (eds) **Computers helping people with special needs.** ICCHP 2012. Lecture Notes in Computer Science, vol 7382, p. 451-458. Heidelberg: Springer, 2012. https://doi.org/10.1007/978-3-642-31522-0_68.

PUTNAM, A. L. Mnemonics in education: current research and applications. **Translational Issues in Psychological Science**, v. 1, n. 2, p. 130-139, 2015. DOI: https://doi.org/10.1037/tps0000023.

PYSINE Version 0.9.2. Copyright 2024. Available online: https://pypi.org/project/pysine/. Accessed on: December 15, 2024.

REN, F.; SHI, H.; ZHOU, Q. A hybrid approach to automatic Chinese text checking and error correction. *In*: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS: E-SYSTEMS, AND E-MAN FOR CYBERNETICS IN CYBERSPACE, 2001, Tucson. **Proceedings** […]. Tucson: IEEE, v. 3, 2001. p. 1693-1698. DOI: https://doi.org/10.1109/ICSMC.2001.973529.

ROSETTA Stone. 1992. Available online: https://www.rosettastone.com. Accessed on: July 12, 2023.

SCIKIT-LEARN Version 1.1.1 Documentation. Copyright 2007-2023. **Metrics and scoring: quantifying the quality of predictions**. Available online: https://scikit-learn.org/stable/modules/model_evaluation.html. Accessed on: July 01, 2022.

SHEN, H. H. An investigation of Chinese-character learning strategies among non-native speakers of Chinese. **System**, v. 33, n. 1, p. 49-68, 2005. DOI: https://doi.org/10.1016/j.system.2004.11.001.

SHEN, H. H.; KE, C. Radical awareness and word acquisition among nonnative learners of Chinese. **The Modern Language Journal**, v. 91, n. 1, p. 97-111, 2007. DOI: https://doi.org/10.1111/j.1540-4781.2007.00511.x.

SHI, D.; GUNN, S. R.; DAMPER, R. I. Handwritten Chinese radical recognition using nonlinear active shape models. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 25, n. 2, p. 277-280, 2003. DOI: https://doi.org/10.1109/TPAMI.2003.1177158.

SOEMER, A.; SCHWAN, S. Visual mnemonics for language learning: static pictures versus animated morphs. **Journal of Educational Psychology**, v. 104, n. 3, p. 565-579, 2012. DOI: https://doi.org/10.1037/a0029272.

SORAGHAN S.; FAIRE F.; RENAUD, A.; SUPPER, B. A new timbre visualization technique based on semantic descriptors. **Computer Music Journal**, v. 41, n. 1, p. 23-36, 2018. DOI: https://doi.org/10.1162/comj_a_00449.

SYSON, M. B.; ESTUAR, M. R. E.; SEE, K. T. ABKD: multimodal mobile language game for collaborative learning of Chinese hanzi and Japanese kanji characters. *In*: IEEE/WTC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, 2012, Macau. **Proceedings** […]. Macau: IEEE, v. 3, 2012. p. 311-315. DOI: https://doi.org/10.1109/WI-IAT.2012.65.

TAFT, M.; ZHU, X.; PENG, D. Positional specificity of radicals in Chinese character recognition. **Journal of Memory and Language**, v. 40, n. 4, p. 498-519, 1999. DOI: https://doi.org/10.1006/jmla.1998.2625.

TAN, H.; YANG, Z.; NING, J.; DING, Z.; LIU, Q. Chinese medical named entity recognition based on Chinese character radical features and pre-trained language models. *In*: INTERNATIONAL CONFERENCE ON ASIAN LANGUAGE PROCESSING, 2021, Singapore. **Proceedings** […]. Singapore: IEEE, 2021. p. 121-124. DOI: https://doi.org/10.1109/IALP54817.2021.9675274.

TAO, H.; TONG, S.; ZHAO, H.; XU, T.; JIN, B.; LIU, Q. A radical-aware attention-based model for Chinese text classification. ***In:*** AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 33rd, 2019, Honolulu. **Proceedings** […]. Palo Alto: AAAI Press, v. 33, n. 1, 2019. p. 5125-5132. DOI: https://doi.org/10.1609/aaai.v33i01.33015125.

TIAN, F.; LV, F.; WANG, J.; WANG, H.; LUO, W.; KAM, M.; SETLUR, V.; DAI, G.; CANNY, J. Let's play Chinese characters: mobile learning approaches via culturally inspired group games. *In*: SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2010, Atlanta. **Proceedings** […]. New York: ACM, 2010. p. 1603-1612. DOI: https://doi.org/10.1145/1753326.1753565.

TIAN, Y. zi2zi: master Chinese calligraphy with conditional adversarial networks. 2017. Available online: https://kaonashi-tyc.github.io/2017/04/06/zi2zi.html. Accessed on: Mar. 10, 2022.

TONG, X.; YIP, J. H. Y. Cracking the Chinese character: radical sensitivity in learners of Chinese as a foreign language and its relationship to Chinese word reading. **Reading and Writing**, v. 28, n. 2, p. 159-181, 2015. DOI: https://doi.org/10.1007/s11145-014-9519-y.

TONG, X.; TONG, X.; MCBRIDE, C. Radical sensitivity is the key to understanding Chinese character acquisition in children. **Reading and Writing**, v. 30, n. 6, p. 1251-1265, 2017. DOI: https://doi.org/10.1007/s11145-017-9722-8.

TSAI, J. C.; WEI, C. C.; TSENG, S. P. Discovering the research issues of classical Chinese segmentation via modern Chinese segmentation system. *In*: INTERNATIONAL CONFERENCE ON ORANGE TECHNOLOGY, 9th, 2021, Tainan. **Proceedings** […]. Tainan: IEEE, 2021. p. 1-3. DOI: https://doi.org/10.1109/ICOT54518.2021.9680653.

THE UNICODE CONSORTIUM. **The Unicode standard**. Version 14.0.0. Mountain View: Unicode Consortium, 2021. ISBN 978-1-936213-29-0. Available online: http://www.unicode.org/versions/Unicode14.0.0.

WANG, A. Y.; THOMAS, M. H. The effect of imagery-based mnemonics on the long-term retention of Chinese characters. **Language Learning**, v. 42, n. 3, p. 359-376, 1992. DOI: https://doi.org/10.1111/j.1467-1770.1992.tb01340.x.

WANG, S. C. A study on the learning and teaching of hanzi-Chinese characters. **Working Papers in Educational Linguistics**, v. 14, n. 1, p. 69-101, 1998.

WANG, T.; XIE, Z.; LI, Z.; JIN, L.; CHEN, X. Radical aggregation network for few-shot offline handwritten Chinese character recognition. **Pattern Recognition Letters**, v. 125, p. 821-827, 2019. DOI: https://doi.org/10.1016/j.patrec.2019.08.005.

WEN, Y. Chinese character composition game with the augment paper. **Journal of Educational Technology & Society**, v. 21, n. 3, p. 132-145, 2018.

WICAKSONO, I.; PARADISO, J. A. FabricKeyboard: multimodal textile sensate media as an expressive and deformable musical interface. *In:* CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 2017, Copenhagen. **Proceedings** […]. p. 348-353. DOI: https://doi.org/10.5281/zenodo.1176278.

WONG, L. H.; HSU, C. K.; SUN, J.; BOTICKI, I. How flexible grouping affects the collaborative patterns in a mobile-assisted Chinese character learning game? **Journal of Educational Technology & Society**, v. 16, n. 2, p. 174-187, 2013.

WONG, Y. K. The role of radical awareness in Chinese-as-a-second-language learners' Chinese character reading development. **Language Awareness**, v. 26, n. 3, p. 211-225, 2017. DOI: https://doi.org/10.1080/09658416.2017.1400039.

WU, S. Y. (ed) **Tujie "shuowen jiezi"** – huashuo hanzi: 1000 hanzi de gushi ["Discussing writing and explaining characters" illustrated — talking of Han characters through pictures: the story of 1000 Han characters]. Taipei: Yueyou Cultural, 2017. 521p. ISBN 978-986-95364-8-6.

WU, X.; ZHAO, H.; CHE, C. Term translation extraction from historical classics using modern Chinese explanation. *In*: CHINESE COMPUTATIONAL LINGUISTICS AND NATURAL LANGUAGE PROCESSING BASED ON NATURALLY ANNOTATED BIG DATA, 2018, Changsha. **Proceedings** […]. Cham: Springer, 2018. p. 88-98. DOI: https://doi.org/10.1007/978-3-030-01716-3_8.

WU, Y. T.; FUJIWARA, E. Keyboard-type optical fiber musical interface. **Optical Engineering**, v. 63, n. 1, p. 016101, 2023. DOI: https://doi.org/10.1117/1.OE.63.1.016101.

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. Identification of the radical component from images of Chinese characters. *In*: SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 35th, 2022, Natal. **Proceedings** […]. Natal: IEEE, 2022. p. 55-60. DOI: https://doi.org/10.1109/SIBGRAPI55357.2022.9991786.

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. Image-based radical identification in Chinese characters. **Applied Sciences**, v. 13, n. 4:2163, 2023. DOI: https://doi.org/10.3390/app13042163.

XIA, W. Y.; TAN, W. N.; XU, Y. M.; CAI, J. L. **Wenzi xiaopinpan** [Little puzzles with characters]. Taipei: Shinmiao Cultural, 1995. 168p. ISBN 9578942079.

XIAN, L.; IBRAHIM, N.; HIDAYAH, N.; SAARI, E. M.; MOHD RAZALI, F. The development of an augmented reality game KANJI Write for beginners. **Journal of ICT in Education**, v. 8, n. 2, p. 79-91, 2021. DOI: https://doi.org/10.37134/jictie.vol8.2.8.2021.

XU, S. **Shuowen jiezi [Discussing writing and explaining characters]**. Eastern Han. Available online: https://ctext.org/shuo-wen-jie-zi/zh. Accessed on: May 11, 2022.

XU, S.; LIU, Y.; YI, X.; ZHOU, S.; LI, H.; WU, Y. Native Chinese reader: a dataset towards native-level Chinese machine reading comprehension. *In*: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 35th, 2021, Virtual. **Track on datasets and benchmarks**. Virtual: Curran, v. 1, 2021. DOI: https://doi.org/10.48550/arXiv.2112.06494.

XUE, M.; DU, J.; ZHANG, J.; WANG, Z. R.; WANG, B.; REN, B. Radical composition network for Chinese character generation. *In*: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 16th, 2021, Lausanne. **Proceedings** […]. Cham: Springer, 2021. p. 252-267. DOI: https://doi.org/10.1007/978-3-030-86549-8_17.

YIN, Y.; ZHANG, W.; HONG, S.; YANG, J.; XIONG, J.; GUI, G. Deep learning-aided OCR techniques for Chinese uppercase characters in the application of Internet of Things. **IEEE Access**, v. 7, p. 47043-47049, 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2909401.

YU, F. T. S. Fiber specklegram sensors. *In:* Yu, F. T. S.; YIN, S. (eds) **Fiber optic sensors.** New York: Marcel Dekker, 2002. cap. 6, p. 235-293.

ZHAN, H.; CHENG, H. J. The role of technology in teaching and learning Chinese characters. **International Journal of Technology in Teaching and Learning**, v. 10, n. 2, p. 147-162, 2014.

ZHANG, H. A review of stroke order in hanzi handwriting. **Language Learning in Higher Education**, v. 4, n. 2, p. 423-440, 2014. DOI: https://doi.org/10.1515/cercles-2014-0022.

ZHANG, J.; DU, J.; DAI, L. Radical analysis network for learning hierarchies of Chinese characters. **Pattern Recognition**, v. 103, p. 107305, 2020. DOI: https://doi.org/10.1016/j.patcog.2020.107305.

APPENDIX A

LIST OF PUBLICATIONS

A.1 Journal article

WU, Y. T.; FUJIWARA, E. Keyboard-type optical fiber musical interface. **Optical Engineering**, v. 63, p. 016101, 2024. DOI: http://dx.doi.org/10.1117/1.oe.63.1.016101. (Contribution: System implementation, testing, and experimental results analysis.)

VIT, F. F.; **WU, Y. T.;** FUJIWARA, E.; CARVALHO, H. F.; LA TORRE, L. G. Microfluidic chip for synergic drugs assay in 3D breast cancer cell. **Microfluidics and Nanofluidics**, v. 28, p. 26, 2024. DOI: http://dx.doi.org/10.1007/s10404-024-02724-0. (Contribution: Software coding.)

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. Image-based radical identification in Chinese characters. **Applied Sciences**, v. 13, p. 2163, 2023. DOI: https://doi.org/10.3390/app13042163. (Contribution: Software design, implementation, and testing.)

FUJIWARA, E.; **WU, Y. T.;** SUZUKI, C. K. Oscillating optical fiber speckle patterns: modeling and application. **Optics Continuum**, v. 1, p. 2490, 2022. DOI: http://dx.doi.org/10.1364/optcon.477064. (Contribution: Software coding.)

FUJIWARA, E.; **WU, Y. T.;** CORDEIRO, C. M. B. Entropy analysis of optical fiber specklegram sensors. **Results in Optics**, v. 5, p. 100155, 2021. DOI: https://doi.org/10.1016/j.rio.2021.100155. (Contribution: Software coding.)

VIT, F. F.; NUNES, R.; **WU, Y. T.;** SOARES, M. C. P.; GODOI, N.; FUJIWARA, E.; CARVALHO, H. F.; DE LA TORRE, L. G. A modular, reversible sealing, and reusable microfluidic device for drug screening. **Analytica Chimica Acta**, v. 1185, p. 339068, 2021. DOI: https://doi.org/10.1016/j.aca.2021.339068. (Contribution: Software coding.)

FUJIWARA, E.; RODRIGUES, M. S.; GOMES, M. K.; **WU, Y. T.;** SUZUKI, C. K. Identification of hand gestures using the inertial measurement unit of a smartphone: a proof-of-concept study. **IEEE Sensors Journal**, v. 21, p. 13916-13923, 2021. DOI: https://doi.org/10.1109/JSEN.2021.3071669. (Contribution: Software coding.)

WU, Y. T.; GOMES, M. K.; DA SILVA, W. H. A.; LAZARI, P. M.; FUJIWARA, E. Integrated optical fiber force myography sensor as pervasive predictor of hand postures. **Biomedical Engineering and Computational Biology**, v. 11, p. 117959722091282, 2020. DOI: https://doi.org/10.1177%2F1179597220912825. (Contribution: Experimental results analysis and paper writing.)

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. Evaluation of optical myography sensor as predictor of hand postures. **IEEE Sensors Journal**, v. 19, p. 5299-5306, 2019. DOI: https://doi.org/10.1109/JSEN.2019.2905229. (Contribution: Software design, implementation, and testing.)

FUJIWARA, E.; DA SILVA, L. E.; CABRAL, T. D.; DE FREITAS, H. E.; **WU, Y. T.;** CORDEIRO, C. M. B. Optical fiber specklegram chemical sensor based on a concatenated multimode fiber structure. **Journal of Lightwave Technology**, v. 37, p. 5041-5047, 2019. DOI: https://doi.org/10.1109/JLT.2019.2927332. (Contribution: Software coding.)

FUJIWARA, E.; RI, Y.; **WU, Y. T.;** FUJIMOTO, H.; SUZUKI, C. K. Evaluation of image matching techniques for optical fiber specklegram sensor analysis. **Applied Optics**, v. 57, p. 9845, 2018. DOI: https://doi.org/10.1364/AO.57.009845. (Contribution: Software coding.)

A.2 Conference article

WU Y. T.; FUJIWARA, E. Optical fiber speckle-based musical interface. *In*: INTERNATIONAL CONFERENCE ON OPTICAL MEMS AND NANOPHOTONICS AND SBFOTON INTERNATIONAL OPTICS AND PHOTONICS CONFERENCE, 2023, Campinas. **Proceedings […]**. Campinas: IEEE , 2023. p. 1. DOI: http://dx.doi.org/10.1109/omn/sbfotoniopc58971.2023.10230966. (Contribution: System implementation, testing, and experimental results analysis.)

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. Identification of the radical component from images of Chinese characters. *In*: SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 35th, 2022, Natal. **Proceedings** […]. Natal: IEEE, 2022. p. 55-60. DOI: https://doi.org/10.1109/SIBGRAPI55357.2022.9991786. (Contribution: Software design, implementation, and testing.)

FUJIWARA, E.; GOMES, M. K.; **WU, Y. T.;** SUZUKI, C. K. Identification of dynamic hand gestures with force myography. *In*: INTERNATIONAL SYMPOSIUM ON MICRONANOMEHATRONICS AND HUMAN SCIENCE, 32nd, 2021, Nagoya. **Proceedings** […]. Nagoya: IEEE, 2021. p. 1. DOI: https://doi.org/10.1109/MHS53471.2021.9767134. (Contribution: Software coding.)

FUJIWARA, E.; **WU, Y. T.;** SUZUKI, C. K. Assessment of Hand Posture and Grip Force by Optical Fiber Force Myography Sensor. *In:* IEEE GLOBAL CONFERENCE ON LIFE SCIENCES AND TECHNOLOGIES, 2nd, 2020 Kyoto. **Proceedings** […]. Kyoto: IEEE, 2020. p. 15-16. DOI: https://doi.org/10.1109/LifeTech48969.2020.1570613711. (Contribution: Software coding.)

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. AR flashcards for Asian language learning. *In*: IEEE GLOBAL CONFERENCE ON LIFE SCIENCES AND TECHNOLOGIES, 2nd, 2020 Kyoto. **Proceedings** […]. Kyoto: IEEE, 2020.p. 409-410. DOI: https://doi.org/10.1109/LifeTech48969.2020.1570619196. (Contribution: Software design, implementation, and testing.)

FUJIWARA, E.; **WU, Y. T.;** GOMES, M. K.; SILVA, W. H. A.; SUZUKI, C. K. Haptic interface based on optical fiber force myography sensor. *In*: IEEE CONFERENCE ON VIRTUAL REALITY AND 3D USER INTERFACES, 2019, Osaka. **Proceedings** […]. Osaka: IEEE, 2019. p. 931. DOI: https://doi.org/10.1109/VR.2019.8797788. (Contribution: Software coding.)

FUJIWARA, E.; **WU, Y. T.;** SILVA, L. E.; FREITAS, H. E.; ARISTILDE, S.; CORDEIRO, C. M. B. Optical fiber 3D shape sensor for motion capture. *In*: IEEE CONFERENCE ON VIRTUAL REALITY AND 3D USER INTERFACES, 2019, Osaka. **Proceedings** […]. Osaka: IEEE, 2019. p. 933. DOI: https://doi.org/10.1109/VR.2019.8798003. (Contribution: Software coding.)

FUJIWARA, E.; DA SILVA, L. E.; DE FREITAS, H. E.; **WU, Y. T.;** CORDEIRO, C. M. B. Optical fiber chemical sensor based on the analysis of fiber specklegrams characteristics. *In*: SBFOTON INTERNATIONAL OPTICS AND PHOTONICS CONFERENCE, 2018, Campinas. **Proceedings** […]. Campinas: IEEE, 2018. p. 1. DOI: https://doi.org/10.1109/SBFoton-IOPC.2018.8610923. (Contribution: Software coding.)

FUJIWARA, E. **WU, Y. T.;** SUZUKI, C. K.; DE ANDRADE, D. T. G.; RIBAS NETO, A.; ROHMER, E. Optical fiber force myography sensor for applications in prosthetic hand control. *In*: INTERNATIONAL WORKSHOP ON ADVANCED MOTION CONTROL, 15th, 2018, Tokyo. **Proceedings** […]. Tokyo: IEEE, 2018. p. 342. https://doi.org/10.1109/AMC.2019.8371115. (Contribution: Software coding.)

WU, Y. T.; FUJIWARA, E.; SUZUKI, C. K. Optical myography sensor for gesture recognition. *In*: INTERNATIONAL WORKSHOP ON ADVANCED MOTION CONTROL, 15th, 2018, Tokyo. **Proceedings** […]. Tokyo: IEEE, 2018. p. 342. DOI: https://doi.org/10.1109/AMC.2019.8371116. (Contribution: System design, implementation, and testing.)

A.3 Conference Poster

WU, Y.T.; FUJIWARA, E. Form of music: visualizing musical phrasing with calligraphy. *In:* IEEE PACIFIC VISUALIZATION SYMPOSIUM, 2024, Tokyo. **PacificVis 2024 Posters**. Tokyo: IEEE, 2024. (Contribution: Feature design, implementation, and testing.)

FUJIWARA, E.; **WU, Y. T.** Enhancing softness sensation in optical fiber tactile sensors. *In:* IEEE PACIFIC VISUALIZATION SYMPOSIUM, 2024, Tokyo. **PacificVis 2024 Posters**. Tokyo: IEEE, 2024. (Contribution: Virtual feedback coding.)

A.4 Registered computer software

FRACAROLLI, J. A.; FALCÃO, A. X.; PAVARIN, F. F. A.; **WU, Y. T. Software para captura de pontos RGB-D por Kinect Xbox 360 para aplicação em biospeckle.** 2019. Patent: Computer Software. Register number: BR512019002155-0. Registration date: 08/10/2019. Title: “Software para captura de pontos RGB-D por Kinect Xbox 360 para aplicação em biospeckle”. Registration institute: INPI - Instituto Nacional da Propriedade Industrial. (Contribution: Software coding.)