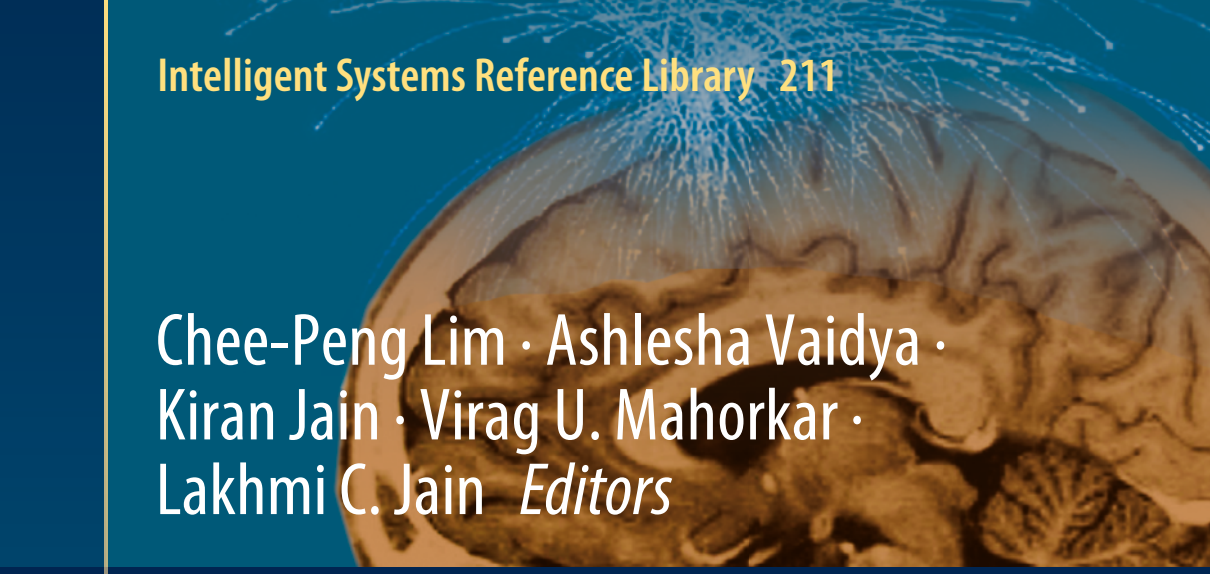
Intelligent Systems Reference Library 211



Chee-Peng Lim · Ashlesha Vaidya · Kiran Jain · Virag U. Mahorkar · Lakhmi C. Jain *Editors*

Handbook of Artificial Intelligence in Healthcare

Vol. 1—Advances and Applications

Springer

Intelligent Systems Reference Library

Volume 211

Series Editors

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

Lakhmi C. Jain, KES International, Shoreham-by-Sea, UK

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included. The list of topics spans all the areas of modern intelligent systems such as: Ambient intelligence, Computational intelligence, Social intelligence, Computational neuroscience, Artificial life, Virtual society, Cognitive systems, DNA and immunity-based systems, e-Learning and teaching, Human-centred computing and Machine ethics, Intelligent control, Intelligent data analysis, Knowledge-based paradigms, Knowledge management, Intelligent agents, Intelligent decision making, Intelligent network security, Interactive entertainment, Learning paradigms, Recommender systems, Robotics and Mechatronics including human-machine teaming, Self-organizing and adaptive systems, Soft computing including Neural systems, Fuzzy systems, Evolutionary computing and the Fusion of these paradigms, Perception and Vision, Web intelligence and Multimedia.

Indexed by SCOPUS, DBLP, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/8578

Chee-Peng Lim · Ashlesha Vaidya · Kiran Jain ·
Virag U. Mahorkar · Lakhmi C. Jain
Editors

Handbook of Artificial Intelligence in Healthcare

Vol. 1—Advances and Applications

Springer

Editors
Chee-Peng Lim
Institute for Intelligent Systems Research and Innovation
Deakin University
Waurn Ponds, VIC, Australia

Kiran Jain
St. Joseph's Hospital and Medical Center
Stockton, CA, USA

Lakhmi C. Jain
KES International
Shoreham-by-Sea, UK

Ashlesha Vaidya
Royal Adelaide Hospital
Adelaide, SA, Australia

Virag U. Mahorkar
Avanti Institute of Cardiology
Nagpur, Maharashtra, India

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-030-79160-5 ISBN 978-3-030-79161-2 (eBook)
https://doi.org/10.1007/978-3-030-79161-2

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Chapter 3
Autistic Verbal Behavior Language Parameterization



Daniela López De Luise, Ben Raúl Saad, Tiago Ibacache, Christian Saliwonczyk, Pablo Pescio, and Lucas Soria

Abstract In severe degrees of ASD (Autistic Spectrum Disorder), patients are not able to produce or understand natural language, and they also have social disorders that make it difficult the communication with other people. Their natural language presents different degrees of alteration, reaching in some cases the impossibility of speaking. This chapter presents an approach to model the patient's behavior by processing recordings during the therapy. Video and audio data provide certain hidden patterns as we already presented in previous work. By using Machine Learning, it is possible to obtain a customized model that makes it possible to evaluate the individual's performance during his interaction with other people. The model inputs a specific set of stereotyped responses collected in a systematic way, labeled as patterns. Those movements and sounds, represents how patterns in audio and video relate to stimuli from the environment. Findings allow to discriminate when and how there is a reaction, an autistic verbal behavior.

Keywords Autistic spectrum disorder · Natural language processing · Linguistics · Sound processing

3.1 Introduction

Autism is a neuron-developmental disorder characterized by a triad of symptoms that can be observed since the first years of life. They mainly consist of impaired language development, stereotyped behaviors associated with restricted interests, and social interaction disorder [1]. About 1 in 54 children born in 2008 have been identified with

D. L. De Luise (✉)
CI2S Lab, Buenos Aires, Argentina
e-mail: mdldl@ci2s.com.ar

T. Ibacache · P. Pescio · L. Soria
IDTI Lab—Facultad de Ciencia y Tecnología—UADER, Concepción del Uruguay, Argentina

B. R. Saad · C. Saliwonczyk
CAETI—Universidad Abierta Interamericana, Buenos Aires, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

C.-P. Lim et al. (eds.), *Handbook of Artificial Intelligence in Healthcare*,
Intelligent Systems Reference Library 211,
https://doi.org/10.1007/978-3-030-79161-2_3

Autistic Spectrum Disorder (ASD) according to estimates from Centers for Disease Control and Prevention, CDC's Autism and Developmental Disabilities Monitoring (ADDM) Network [2]. ASD is more than 4 times more common among boys than among girls [3]. About 1 in 6 (17%) children aged 3–17 years were diagnosed with a developmental disability, as reported by parents, during a study period of 2009–2017. These included autism, attention-deficit/hyperactivity disorder, blindness, and cerebral palsy, among others [4]. The World Health Organization changes those numbers to one in 160 children, but based on a study of Mayada et al. in 2012, that use a survey instead of statistical inference of patients' records [5]. In recent years there have been plenty of studies providing data about how many people in the world are being affected by ASD. Nowadays results are yet not conclusive due to various reasons, among others samples used are from different records, diagnostic procedures, experimental designs, ages ranges, etc. Despite all, it can be said that at least 60–70 people in 10,000 are diagnosed with traditional autism. Many others remain in a borderline, not diagnosed or even misdiagnosed. Despite the imprecision with the numbers, the community agrees in its increasing rate in countries where autism were tracked.

Children with autistic disorders can have significant social, communication, and behavioral problems. The Diagnostic and Statistical Manual of Mental Disorders (DSM), is the most relevant diagnostic classification system. The World Health Organization (WHO) considers autism as pervasive developmental disorders, characterized by a wide variety of clinical and behavioral expressions that are the result of multiple factors related to developmental dysfunctions of the central nervous system [2].

On the other hand, children with these types of disorders show significant cognitive deficits in different areas [7]. Many of the characteristics of children with Generalized Developmental Disorder (PDD) and autism could present alterations in their attention processes [8]. Autistic people have particular difficulties in interpreting socially relevant information, since socially significant stimuli are physically complex. Problems with social interaction, are combined with cognitive and physical symptoms, but their timings, severity, and nature may be quite different [9].

The alteration of verbal language can reach the complete impossibility of speech [10]. It is associated with Sensory processing disorder (SPD), a condition in which a person does not respond normally to sounds, smells, textures, and other stimuli. Many times autistic is associated to both hearing loss or hyperacusis, that can be general or focused on a certain sound. Other typical manifestation is the "stimming", a short for self-stimulatory behavior, sometimes also called "stereotypical" behavior. In a person with autism, stimming usually refers to specific behaviors that include hand-flapping, rocking, spinning, or repetition of words and phrases.

It is important to note that subtler forms of stimming are also a part of most people's behavior patterns. The main differences between autistic and typical stimming are the type, quantity, and obviousness of the behavior. In general, behaviors are described as "stims" when they go beyond what is culturally tolerated. Sometimes they can be quite extreme and are legitimately upsetting or even frightening to typical people. There are cases where autistic people stim by making loud noises that can sound

threatening or scary [11]. In other cases they hit themselves with their hands, or bang their heads against the wall. Those manifestations that go beyond certain limits, are problematic for a variety of reasons [12].

People with ASD, find it difficult to stop stimming. Stim appears when the individual is excited, happy, anxious, overwhelmed, or because it feels comforting. It helps the patient to handle overwhelming sensory input (too much noise, light, heat, etc.) [13].

It's not fully understood why stimming usually goes along with autism. Many experts believe it is a tool for self-regulation and self-calming. It could be an outgrowth of the sensory processing dysfunction that often goes along with autism [14].

In this project it is important, since the autistic stim relates to feelings of anxiety, fear, anger, excitement, anticipation, and other strong emotions [15]. They also stim to help themselves handle overwhelming sensory input (too much noise, light, heat, etc.). Therefore, the external manifestations could serve as a clue to understand the individual's status and thoughts.

In certain cases, stimming are with certain behaviors that are a sort of rudimentary communications but are understood as a distraction. The distinction in those cases is very hard. The parameterization of verbal behavior, proposed in this work is useful to do that discrimination.

There are some extra features in the body expression of people suffering autism. For example "echolalia". It consists of the precise repetition, or echoing, of words and sounds. Echolalia can be a symptom of various disorders, but it is most often associated with autism. It is one of the first ways in which children with ASD uses to communicate, and represents a relevant tool for speech-language therapy [16].

But it is important to note that echolalia has no communicative meaning at all in severe cases of autism. It may simply be a self-calming tool working as good as any other like hand-flapping or rocking. In those cases it is an imitation of human speech without grasping the meaning behind those sounds. So, it can be another way to calm themselves when they're anxious or cope with overwhelming sensory challenges. Thus, echolalia may work as a form of self-stimulation or stimming.

Other people on the autism spectrum use "prefabricated" phrases and scripts to communicate ideas when it is too difficult for them to formulate their own novel speech patterns. From this perspective, echolalia is an important first step toward more typical forms of spoken communication.

Any case, autistic verbal behavior can provide very useful information to understand what is happening inside the mind of the patient. As every individual has its own set of stimmings, phrases and scripts, a model can work as a dual translator machine to and from the environment.

This article follows this approach, certain studies that process utterances to determine whether a patient has ASD or not, and combine them with own findings on stimming and audio processing.

Authors started using five different musical notes are composed which are described by: pitch, intensity and features of timbre: frequency, amplitude, harmonic composition or waveform [17]. Different tests were carried out with autism spectrum

patients using those notes as stimuli. Experiences were filmed and the reactions they had to each note (movements of hands, arms, head, feet, sounds, etc.). The protocol included filming, and hand-made descriptions on sheets. All the information was used to perform statistics and develop a set of main features. It should be noted that this type of evaluation contrasts with the proposal, where the patient's sound production is analyzed.

In the next stage, simple visual information is added. To do that, the study performed by Cheol-Hong and Ahmed [18] was considered. It presents a new method for automatic detection of stereotyped behavior patterns using data from an accelerometer. Orthogonal subspaces are extracted from the sensor data, which are used to generate a clustering of the dictionary and in turn for the representation of signals. The algorithm was improved for the detection of new events that were unknown. In [19] a study carried out at the University of Southern California (USA) of the design of an emotionally directed interactive agent for boys with autism, called Rachel, is presented. In it, an incorporated conversation agent (RCT) was developed to achieve the interaction of patients, previously diagnosed with autism, who participated together with their parents, interacting with the RCT. The agent produces different scenarios that represent emotions such as angry, sad, happy and scared; with the aim of allowing a controlled evaluation of the communication skills of children with ASD. Although this work is extensive, it lacks of statistics and result validations.

The current work, takes some of the above approach. Instead of a specific interaction, the current therapies are considered. This is because the routines are very important for this type of patients. Changing the activities and/or the therapist affects the study. The processing is also different: the audio and video recordings are processed in parallel, with statistical evaluation and an automated model that serves as a tool for later evaluation of the activity. Both models (audio and video) will be afterwards related using the timestamp. Previous publications introduced some preliminary findings [20]. In this work, only audio processing is considered.

The rest of this paper organizes as follows: Sect. 3.2 presents some of the main working considerations about ASD, Sect. 3.3 introduces the proposal, materials and methods, Sect. 3.4 describes a testing protocol and preliminary considerations of the use case, Sect. 3.5 is the test for audio recordings of the use case, and Sect. 3.7 has conclusions and future work.

3.2 Considerations About the Autistic Spectrum Disorder

This section presents main aspects of the project. First, it introduces main definitions of ASD, starting with the three degrees of autism, and their stims. They are important since the article focuses in the severe conditions, which is the third degree. Then it introduces Skinner's verbal behavior, definition that is used to determine the model's variables.

3.2.1 Degrees of Autism

The DSM-V [21] of the American Psychiatric Association, establishes a classification guide, a coding used by much of the scientific community to diagnose different mental disorders. According to the authors [22] those who suffer from ASD experience a series of symptoms based on Wing's triad, these disorders refer to social interaction, communication and lack of flexibility in reasoning and behavior. In turn, there are classifications of different degrees in the guides and manuals, which leads specialists [22] to affirm that despite Wing's classification, no person suffering from ASD is similar to another in terms of their observable characteristics.

People who are diagnosed with Autism Spectrum Disorder ASD have deficiencies in communication and social interaction. The DSM-V guide determines the level of ASD on a three-grade scale, detailed below.

- **Grade 1**. Need Help

In this stage, there is still social communication. The patient can be without on-site help, but presents deficiencies in social communication causing major problems. Typically has difficulty initiating social interactions and there are clear examples of atypical or unsatisfactory responses to the social openness of other people. May appear to have little interest in social interactions. For example, a person who is able to speak in complete sentences and who establishes communication but whose extensive conversation with other people fails and whose attempts to make friends are eccentric and usually unsuccessful.

Typically have restricted and repetitive behaviors. The behavioral inflexibility causes significant interference with performance in one or more contexts, with difficulty alternating activities. The organizational and planning problems make autonomy difficult.

- **Grade 2**. Needs Notable Help

Patient has very reduced social communication, but with notable deficiencies in verbal and non-verbal social communication skills; apparent social problems even with on-site help; limited initiation of social interactions; and reduced response or abnormal responses to the social openness of other people. For example, a person who emits simple sentences, whose interaction is limited to very specific special interests, and who has very eccentric non-verbal communication.

As in grade 1, presents restricted and repetitive behaviors, with behavioral inflexibility, difficulty coping with changes, or other restricted/repetitive behaviors often appear clearly to the casual observer and interfere with functioning in various contexts. Also presents anxiety and/or difficulty in changing the focus of action.

- **Grade 3**. Needs Very Notable Help

This degree is the most severe condition with almost none social communication. Presents severe deficits in verbal and nonverbal social communication skills cause severe functional impairment, very limited initiation of social interactions,

and minimal response to other people's social openness. For example, a person with few intelligible words who rarely initiates interaction and who, when he does, performs unusual strategies only to meet needs and only responds to very direct social approaches. Behaviors are restricted and repetitive, with inflexibility, extreme difficulty coping with changes, or other restricted/repetitive behaviors significantly interfere with functioning across the board. Intense anxiety/difficulty in changing focus of action.

3.2.2 Verbal Behavior

In DSM-V there is a mention of verbal and non-verbal social communication, when indicating one of the problems that patients with ASD present. Skinner [23] carried out studies on verbal behavior, indicating that vocal language is only a part or subset, arguing that the emission of sounds or certain actions such as gestures are verbal even though they are not part of an organized language, since they provoke a reaction in the listener or observer similar to the emission of vocal language and therefore must be part of the verbal behavior.

Repetitive and stereotyped behaviors are considered an important symptom of autism spectrum disorders (ASD). Jumping, turning, and other rhythmic body movements were described in the first patients originally described in Kanner's series [24, 25].

Leo Kanner conducted an evaluation in the case study with 8 boys and 3 girls, describing various repetitive movements such as: rhythmic jumping, clapping, sucking sounds, walking on tiptoe, not looking directly at the face, continuous repetition of words and/or phrases. Later, in 1979, Wing and Gould [25] determined that people with Autism Spectrum Disorder have problems in three areas: (a) They cannot develop skills in reciprocal social interaction. (b) They have difficulties in their verbal and non-verbal communication and (c) they have restricted patterns of behavior and interests. These symptoms are known as the Wing Triad.

3.3 Materials and Methods

This section describes how the use case is being designed, the hardware and software used for data collection, and certain preliminary considerations for the testing.

3.3.1 Hardware

For the use case, a therapy session with a patient diagnosed with severe ASD (degree 3) is considered. The patient's age is 13 years old, native from Argentina, with

parents authorization. The activity analyzed cover several typical tasks: identification, concentration, pointing, etc. The interaction was registered in a video recording, with a Sony Handycam HDR-CX250, with 5 sound channels.

3.3.2 Protocol

The video has a total duration of stimmings, phrases and scripts 22:00 min. As the scope of this article covers only audio, it was extracted as a wav file.

All the procedure was also registered in printed forms. The use case consists of a set of sounds and movements of the original video, that are part of the patient's verbal behavior and allows to identify the different actions. From them it is expected to derive the audio and visual stimmings, phrases and scripts, as introduced in Sect. 3.1. The initial set of clues was manually registered in an online Google spreadsheet.

The worksheet was divided into the following columns: Time, Action, Characteristics, Comments. That original worksheet was extended to add two new columns with the timestamp before the starting of the visual/audio clue, of a possible stimulus, and a similar column with the duration of the behavior. The Action column, which is the stim observed (a body movements, or certain noise or sound produced by the patient). A short description of these movements is recorded in the Characteristics column; and finally, in the Comments column, there are the actions and expressions that the therapist performs at each opportunity.

The next step to prepare the use case is to convert this sheet into a list of stims with every repetition during the recorded session, and relevant information for its automatic detection. Part of the resulting document is in Table 3.1.

While the registration has been traditionally performed manually according to the actions of the patient and the indications of the therapist, the actions are those that were described in the introduction section. There are many papers explaining them, and for the aim of this project, they were added to the protocol. An excerpt of that long list of detectable movements and sounds are in Table 3.2 and is based on the work of Riviere [26].

The last step to prepare the use case is to adapt Table 3.1 to use the listing in Table 3.2. This way, any clue detected represents a stimming, or specific utterance.

3.3.3 Software

For processing, five versions of a prototype were written in Python, and five versions in Octave. Both languages are frequently used for signal processing, statistical analysis and certain mining procedures [27–33].

The final version of the prototype, consists of a python module with several functions. The visual and audio signals required the Python and Octave library LTFAT,

Table 3.1 Time-sheet of Stimulus upon every sound slice detected

Reference	$Time_1$	$Time_2$	$Time_3$	$Time_4$	$Time_5$	$Time_6$	$Time_7$
Time slice 1	0:00:00	2:31:00	3:58:00	4:18:00	5:17:00	6:52:00	7:25:00
Label: iuuu		He moved or was someone in the room, before he had lost focus on the therapist	The book is given	The therapist stood up	Most of the time she is standing behind him hace uuuuu	A vehicle is heard in the background while passing the pages of his book	He got bored reading the book a second time and turning the pages seemed
Time slice 2	1:57:00	2:07:00	2:19:00	2:52:00	3:18:00	7:03:00	8:20:00
Label: aaaa	You won the token (Buenisimoooooooo) + timer beep + eye contact to therapist + gives him yellow toy	Token arrangement	Points to card and the therapist congratulates him + gives him red toy	Buenisimooooooo (good)	He gives him yellow ball, (Biennnn Excellentee)	Wait for the stimulation of the therapist to turn the page	Look at the floor, he gets bored of waiting
…							
…							
Time slice 4	3:40:00	10:30:00	16:25:00	16:30:00	18:10	19:09:00	
Clapping Sound 4	Given the lack of attention to him, the therapist takes a long time in his notes	He asks for silence, takes him a glass and plate and touches his mouth. Seems eager to play and win the lollipop	Sneezing and health	Sneezing and health	They clean his nose	He's bored now, he makes many new movements, he stretches, he wants to finish now	

Table 3.2 ASD typical behavior

Movement/sound	Description
Rhythmic jumps	The body sitting/standing, the patient facing the front, performs vertical stentorian movements with/without arm movements
Clapping	It is the action of clapping the palms of the hands together repeatedly as a sign of both approval (joy) and, sometimes, rejection
Sucking sounds	It is the effect of producing a sound similar to the sucking of the bottle or the mother's breast when drinking milk
Walk on tiptoe	It is the action of walking on the balls of the feet without contact between the heels and the ground. It hits both feet at the same time
Do not look to the face directly	The patient, seen from the front, does not direct his gaze towards us but appears skewed, to the right (more frequently) or to the left
The list continues…	

and Octave libraries Signal and Control. LTFAT provides signal processing algorithms for use with Octave and Python software. The other two completes some basic processing that are part of the current testing, like FFT, modulus of a complex signal, mean, covariance and Fast Walsh-Hadamard Transform (FWHT), wavelets, design of FIR and IIR filters, spectral analysis, etc.

The use of the Fourier transform does not solve the problem of a better localization of a signal. In addition, we will have problems if the events are presented very close to each other since it will be very difficult to differentiate behaviors within the same window width.

The mathematical tool that helps solve these problems is the Wavelet Transform. This transform can detect high frequency phenomena. Once we choose the window size, the frequencies are analyzed with the same time and frequency resolutions, different from what happens in the Wavelet Transform which has a window size adapted to the frequencies.

This work, based on the BIOTECH project, applies also Morphosyntactic Linguistic Wavelets (MLW). An intelligent system capable of providing in-depth information on the reasoning process related to linguistic expressions [34]. It has been applied in the past to interpret WEB content, automatic summarization, dialogue profiles, texts and author profiles, bilingualism effects in young children, etc. [35, 36].

To be able to process with MLW it is mandatory to have linguistic expressions made by the patients. To overcome the lack of language articulations, the project is based on the concept of verbal behavior [37], replacing language production with specific audio and video clues.

Verbal behavior is becoming widely accepted as a crucial perspective for autism spectrum disorder in therapies; Greer [38] expresses that there is more and more evidence on onto-genetics, the sources of language and its development that leads to

a renewed interest in the theory that verbal ability in humans is the result of evolution and onto-genetics of development.

Another important aspect to consider is the particularity of each individual. Experts claim [9] that skills training should be intensive, but also personalized for best results. This article does not cover the MLW production, or the traditional wavelets (like Daubechis 7 or 8), as they are out of the scope of this work.

3.4 Preliminary Evaluation

For this activity the data-set is the audio, obtained according to the steps described in the protocol (see Sect. 3.3). The main goal here is to identify the timestamps corresponding to stimming sounds or any audible activity of the patient, that could be related to any stimulus from the environment.

During the exhaustive revision, 88 instances of interactions were identified. Each one was labeled and then clustered by similarity. For instance, sounds similar to the expression "iuuu" were called Sound 1; the expressions like "aaaa" are Sound 2; "brrrr" is Sound 3, the clapping is Sound 4, and so on. It is important to note that certain sounds, such as the expression of joy, are very complex for a precise identification and not considered in the list.

On several occasions, each background sound is accompanied by a sound or movement of the patient, for example in the order number 4, 5, 14, 21, 28, 29, 33, 36, 43, 46, 47, 57, 58, 59, 74, 77, 79, 80, 81. On other occasions, the patient performs the clapping movement accompanied by one of the sounds registered, for example in the order number 1, 2, 22, 24, 44, 75, 78, 82. In addition, slice of sound 1 is associated with a movement and expression as if it were angry as recorded in the order 15, 61, 69, 83.

3.4.1 Modeling the Problem with Metadata

The model consists of a set of clues, and the variables that best describe them. In this process it is important to find out which is the specific set of clues used by the patient of the use case, taken from Table 3.2. To avoid any biasing in the detection and parameterization of the clues, a sliding windows algorithm on the entire data-set is used to detect when every sound slice is being repeated. Then the modeling needs to determine a very important subset of parameters: the similarity threshold of every variable comparison.

Four main variables were used, A and B, which would be the variables to store the sound file used; AA and AB whose function is to store the number of seconds that each loaded sound consists of.

Also, it uses two for loops from 1 to AA/AB whose main objective is to divide and compare both sounds for every second (The graph will vary based on this). An

if conditional where it is stated that, if both sounds are equal in a given second, they are stored in a variable C. The pseudo-code of the mentioned algorithm is in Fig. 3.1.

Regarding the graphical part, it consists of LTFAT functions, where it is plotted from the scale on the X and Y axes, and the variation of sound with the previously mentioned variable C is displayed (Fig. 3.2 and a zoom-in in Fig. 3.3).

Every sound has 4 sub-slices labeled SS1 to SS4, in order from the first second to the last. Figures have the following color code:

Fig. 3.1 Sliding Windows for matching the pattern slice

```
function retval = SlidingWindows4 ()
B=”TherapyVideo.wav”
A = “Slice.wav”
N = 1e3;
AA = size(A); #Matrix size to make the for in the comparison
AB = size(B);
#Sliding Windows comparison
for i = 1 : AA,
  for j = 1 : AB,
     if metaDataComparison (A(i), B(j))  #meta data are simmilar
       C(i) = A(i)
     endif
  endfor
endfor
endfunction
x = SlidingWindows4()
save salidasw.csv x(1:100)
```

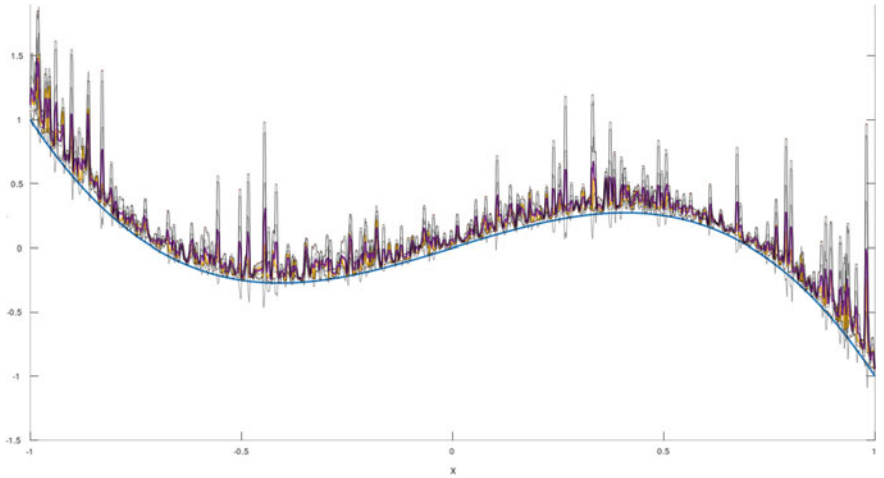


Fig. 3.2 Sliding Windows graphical output

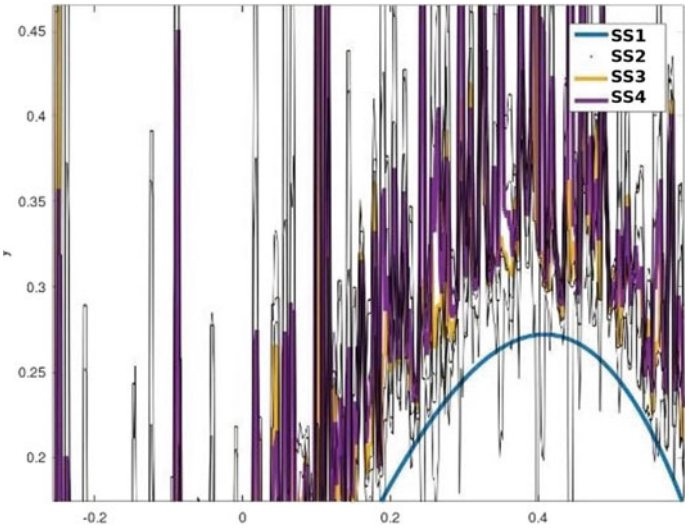


Fig. 3.3 Sliding Windows graphical output with zoom in

- **Blue** - SS1
- **Red dot** - SS2.
- **Yellow** - SS3.
- **Violet** - SS4.
- **Black** – $_{i}$ SS_{i}.

The plot shows the metadata corresponding to each Sub Slice of a pattern.

The plot shows the differences between metadata values for the video sub parts, as being processed for sound pattern. Running the algorithm using the Sliding Windows with two cut pointers algorithm [39] on sounds results in Fig. 3.4.

The plot represents the main curve representing the variations between the current slice and slice 1 is now bi-modal. This is because the comparison is second by second and not the entire slice. Since the sensibility is bigger, the plot shows that behind the global behavior shown in Fig. 3.2, there are moments where metadata representing the sound matches better in some parts.

3.4.2 Advantages and Disadvantages of Using the Proposed Approach

The Sliding Windows technique, was implemented in several versions, with certain pros and odds each one. The first version was created in Python, after an investigation

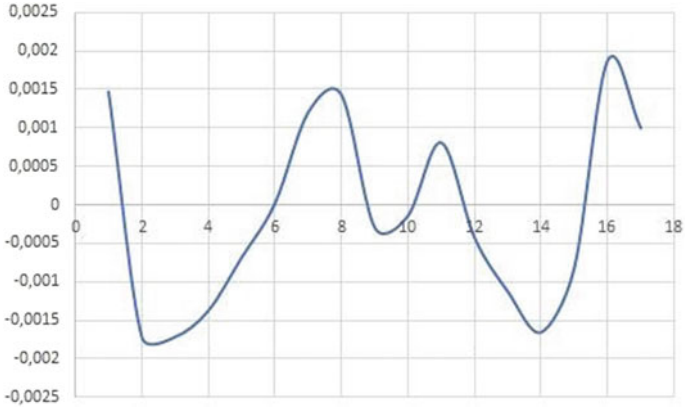


Fig. 3.4 Sliding Windows with 2 cut-points

of the case, the frequency, duration and comparison of similarities were obtained. The biggest downside is that it requires different libraries for it to work.

Then the first version was made in Octave, the result of investigating the operation of the software thoroughly, it is possible to pass an audio strip and that it reflects a graph detailing the different sounds that were detected. The disadvantage is that it is not possible to load more than one audio strip at a time and it is also very impractical to compare different graphics.

The second version was made also in Octave, can pass two different audio strips, obtaining a graph in which it is analyzed in which points there are coincidences.

The remaining versions are in Python, but based on the algorithm in Fig. 3.5, and is able to read a selected audio strip, a number of coefficients, the wavelet's

```
function retval = AplicoWvlet (zPatron, zInput, tipoWavelet, canal, numCoefWv)
f=input1;
[f,fs] = audioread (input1);
J = input2; % number of coefficients
wvNom = input3; % the wavelet to apply
%reduce input to one channel
[c,info] = fwt(f(:,canal),wvNom,J,1); % apply the wavelet and return coefficients
plotwavelets (c, info,fs,'dynrange',90);
## Frequency bands of the transform with x-axis in a long scale and band
## peaks normalized to 1. Only positive frequency band is shown
[g,a] = wfbt2filterbank({wvNom,J,'dwt'});
filterbankfreq(g,a,20*1024,'linabs','posfreq','plot','inf','flog');
% show the information structure of the signal
info
%return coefficients
retval=c;
endfunction
```

Fig. 3.5 Wavelets algorithm

parameters, and the type and number of audio channel. The last version can show analytical graphs and output a file with timestamps, and parameters in csv format.

3.5 The Sounds of the Use-Case

The analysis of every slice provides information on how often certain pattern is being repeated during the interaction with the patient. This type of information intends to make it relevant how every slice acts as a vocabulary for the individual. Following this conception Table 3.2 becomes a reference of that to expect when there is a stimulus, to confirm where is the focus at any moment.

All the selected sound patterns were made to compare each one of them and draw more precise conclusions about the behavior of the data. The comparison of the sound clippings, despite being different in their specific content, apparently have a very similar result between them.

Table 3.3 show the most relevant clippings, the echoic sound (Label), the internal ID (IDU), the order of first instance (FI), the timestamp of the first instance (Time), the threshold defined to determine is the cumulative distances between a metadata in the pattern is similar to metadata in some part of the video (MetaDtThreshold), the threshold defined to determine is the distance between certain metadata in the pattern is similar to metadata in some part of the video (PatternThreshold) and the number of sub-slices of the pattern (NumSubSlices).

Table 3.4 summarizes a selection of the first eight selected clippings considered to be the most repetitive in the recording. The plots show the wavelet description in every case. They are so frequent in the original interaction with the patient that are used as patterns of reactions.

The table makes it clear that every pattern has its own set of values. Something interesting arises when the numeric values of the ten first wavelets are compared. They have exact the same value for half of the code values, for all the patterns (see Table 3.5).

Table 3.3 Summary of patterns detected in the interaction

Label	IDU	FI	Time	MetaDtThreshold	PatternThreshold	NumSubSlices
iuiuiuui.wav	1	1	10:10	0.5	0.4	4
iuiuiuu.wav	2	2	14:22	0.5	0.4	4
iuiuiuiu.wav	3	3	13:30	0.5	0.4	4
iuiuiuiu.wav	4	3	13:30	0.5	0.4	4
iuiuiui.wav	5	1	7:23	0.5	0.4	4
iuiuiu.wav	6	0	0:00	0.5	0.4	4
iuuu.wav	7	4	2:53	0.5	0.4	4
iuiuiu.wav	8	0	10:17	0.5	0.4	4

Table 3.4 Plots of the pattern's wavelet encoding

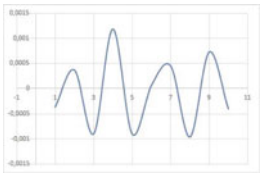
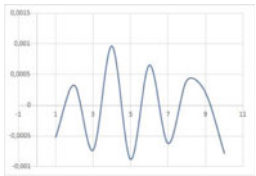
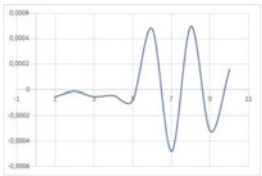
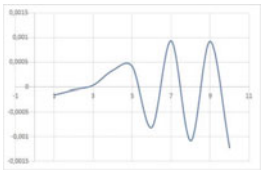
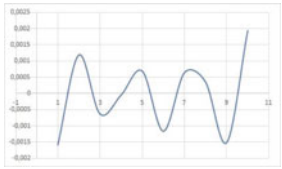
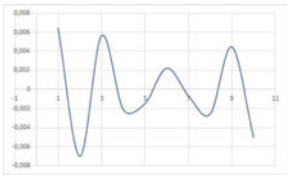
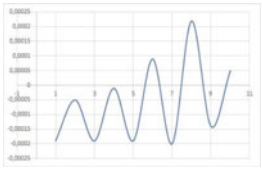
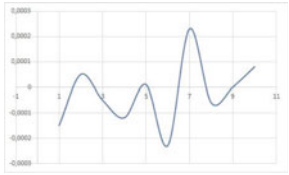
Wavelets 1–10 of patterns		
 IDU 1 **IDU 1**	 IDU 2 **IDU 2**	 IDU 3 **IDU 3**
 IDU 4 **IDU 4**	 IDU 5 **IDU 5**	 IDU 6 **IDU 6**
 IDU 7 **IDU 7**	 IDU 8 **IDU 8**	

Table 3.5 Shared WV-coefficient in patterns

MetaData<x>SubSlice<y>	Pattern IDU	Coefficient
MtDt1SS1	[1…8]	0.0010946527865457112
MtDt3SS1	[1…8]	1321.678276643991
MtDt1SS2	[1…8]	5.6848702944657583e−05
MtDt3SS2	[1…8]	1320.678276643991
MtDt1SS3	[1…8]	−0.00099214605055142672
MtDt3SS3	[1…8]	1319.678276643991
MtSt1SS4	[1…8]	−0.00034117011565242596
MtDt2SS4	[1…8]	1318.678276643991

In the first column **MetaData<x>SubSlice<y>**, represents the meta-data number x of the sub-slice y. For instance MtDt1SS1 is the first meta-data of the first sub-slice. Being the meta-data, the slice selected. Table 3.6 shows the differences between wavelets.

Table 3.6 Peculiar WV-coefficient in patterns

IDU	MtDt2SS1	MtDt2SS2	MtDt2SS3	MtDt2SS4
1	−0.002052857577679	0.000258812900079	−0.0002450727393450	−0.00152047488387144
2	0.0021977747536751	-0.003481925534852	0.0022764760814601	0.000693833371500096
3	−0.0009735533616354	−0.000102468070003	0.0001136690832276	−0.00130344882594886
4	0.0031322958329095	−0.004955072342362	−0.0002328834569618	0.00318271031936332
5	−0.0001635743285072	−0.002160975862761	0.0011700772129727	0.00092281295499667
6	0.0007771241364974	−0.000633479491237	0.0017681726388932	−0.00115926435883696
7	0.001260975846586	0.00403322021435	−0.0018817369033875	0.00483011407221703
8	0.002785511914547	0.0005998036466599	0.0031983855195914	4.3684541036182338

Table 3.7 WV-coefficient in patterns

IDU1	IDU2	IDU3	IDU4	IDU5	IDU6	IDU7	IDU8
−0.00036	−0.00052	−0.00005	−0.00015	−0.00159	0.00645	−0.00019	−0.00015
0.00037	0.00032	−0.00001	−0.00005	0.00118	−0.00702	−0.00005	−0.00005
−0.0009	−0.00073	−0.00005	0.00004	−0.00063	0.00566	−0.00019	−0.00005
0.00119	0.00096	−0.00004	0.00034	−0.00005	−0.00209	−0.00001	−0.00012
−0.0009	−0.00088	−0.00008	0.00042	0.00069	−0.00144	−0.00019	0.00001
0.00006	0.00065	0.00048	−0.00081	−0.00116	0.00222	0.00009	−0.00023
0.00045	−0.00062	−0.00048	0.00094	0.00064	−0.00065	−0.0002	0.00023
−0.00096	0.0004	0.0005	−0.00108	0.00035	−0.00256	0.00022	−0.00006
0.99973	0.0002	−0.00032	0.00093	−0.00151	0.00447	−0.00014	0
−0.0004	−0.00078	0.00016	−0.00122	0.00193	−0.00502	0.00005	0.00008

Finally, Table 3.7 has the wavelet decomposition of the eight patterns as a whole (without Sub-slicing and metadata).

3.6 Test and Evaluation

The previous section describes the main features that allow to determine that the entire audio has only eight sound patterns that may represent a kind of stimming, as they have identical parameters and repeat several times. This section, is the statistical evaluation and validation of the previous findings.

The variables considered for the lightweight version of the detection and their parameters are in Table 3.8.

In order to study how the model works for a single pattern, Pattern4 is considered from the entire list (see Table 3.9).

Sound recording was extracted from the original video into a.wav file. The sampling rate is fs = 44,100. Audio stream has every instances sampled as a tuple of 2 values, here named as X and Y.

The files obtained after Octave processing are labeled SalPattern1.Rami<x>.csv, with x = [1.0.3].

Patterns 9 and 10 are slices that seem to be a distorted version of one of the previous patterns. They were kept since there is still possibility that they are new patterns.

In the table **SalPattern1.Rami<x>.csv** stands for the output file of input **Rami<x>**, with **<x>** being the sequential ID of the input wav. There are up to 3 files, that is x = [1.0.3], since the original video had to be partitioned to adequate it to the computational availability of the project.

This section focuses on Pattern4, selected at random to show the case study. Table 3.10 describes the case of Pattern4.

Table 3.8 Fast detection variables and parameters

Variable	Description	Parameters
AVG-DifX, AVG-DifY	Average difference between first and second coordinate of the wave sample	Coordinate = {1, 2}
minX, minY	Minimum value of the first coordinate	Coordinate = 1
maxX, maxY	Maximum value of the second coordinate	Coordinate = 2
AvgCovX, AvgCovX	Average xcov, which is Compute covariance at various lags [=correlation(x-mean(x), y-mean(y))]	Coordinate = {1, 2}
MaxDifWalsh-HadamardX, MaxDifWalsh-HadamardY	Maximum of the differences in Fast Walsh-Hadamard Transform (FWHT) of every slice against the window of the video	*order = Hadamard, number of elements = 64*
R-CoefFFT	Coefficient R of covariance of FFT harmonics	numCoefFFT = 30
difCoefFFT	Difference between FFT coefficients of the slice versus the window	NumCoefFFT = 30
Threshold for FFT comparison		uFFT = 0.0005
Threshold for accumulated differences		Utot = 3

Table 3.9 Set of patterns detected in the video

Pattern	ID
iu iu iu iu 12,47	Pattern1
iu iuiuiu 3,20	Pattern2
iuiuiu 10,17	Pattern3
iuiuiu min 0	Pattern4
iuiuiui 7,23	Pattern5
iuiuiuiu 8,34	Pattern6
iuiuiuiu 13,30	Pattern7
iuiuiuu 14,22	Pattern8
iuiuiuui 10,10	Pattern9
iuiuuiiu 4,15	Pattern10
iuuu 2,53	Pattern11

Table 3.10 Data processed for Pattern 4

File	Records	Time	Pre-processed file
SalPattern4.Rami1.csv	1,048,576	441”	SalPattern4.set1.csv
SalPattern4.Rami2.csv	1,048,576	441”	SalPattern4.set2.csv
SalPattern4.Rami3.csv	1,046,199	440”	SalPattern4.set3.csv
Total	3,143,351		

The file length corresponds to a rate of 2.378 samples per second, in order to reduce the computational power required. The sampling change is performed after reading in Octave with 44,100 as a sampling rate. The last column is the name after certain pre-processing required for the exploratory analysis with WEKA 3.8.4

3.6.1 Pre-processing

From the previous considerations, the new version of the data-set has to be processed with WEKA to determine the most reduced set of variables to be used for the model of Pattern4. To do that there is the following pre-processing:

Step 1. Files SalPattern4.set<n>.csv, needs to change the regional setting in Octave to the one in LibreOffice Calc:. A script in Octave changes “,” to “;” and “.” to “,”.
Step 2. With LibreOffice Calc every numeric column is changed to numeric format with leading sign, and 8 decimals. When needed until 20 decimals are defined. The file is then exported to CSV format.
Step 3. A script in Octave changes “,” to “.” and “;” to “,”
Step 4. Import data in WEKA (c).

It is important to note that all the processing is being made with open source, standard in the field: Octave and WEKA. Both of them very used by its compatibility with other tools (like python and Java), its flexibility, cross platform installation, and deep Learning abilities. Furthermore, every graphic in this paper was performed in these tools.

3.6.2 Variable Selection

This section explains the variable selection procedure and justification. Some are based on practical criteria, others required more sophisticated approaches like Subset Evaluation algorithms based on information provided of the variables and PCA.

Variables ***hit***, and ***tira*** are discarded, since in this version of the project, ***hit*** depends only on ***ScoreTotal***; and ***tira*** is the name of the pattern’s use case, which is the same

for every instance under the scope of present study. For this process there is not wavelet or any other extra feature considered, so ***ScoreTotal*** and ***ScoreParcial*** are equal (the accumulated scoring and the scoring due to basic features), thus ***ScoreTotal*** is also discarded. Features considered as basic features are in Table 3.11.

The data-set presents no missing values by construction, and all (except ***tira***, which is the pattern name) are numbers. The distribution of ***ScoreParcial*** (see Fig. 3.6) shows that a few instances are close to 0.1, the value of the highest approximation of the current slice to the pattern. Figures 3.7, 3.8 and 3.9 show similar behavior.

Zones with highest R are the ones candidates to be the pattern, since coefficients trend to follow a similar sequence. It is important to note that peaks are not located in

Table 3.11 Features considered for processing

Feature	Description
AVG-DifX	Average of the difference in value X between Pattern and current slice of the audio
AVG-DifY	Average of the difference in value y between Pattern and current slice of the audio
AvgCovX	Average of the covariance in value X between Pattern and current slice of the audio
AvgCovY	Average of the covariance in value Y between Pattern and current slice of the audio
minX, minY	Minimum of difference in value X and Y respectively
maxX, maxY	Maximum of difference in value X and Y respectively
MaxDifWalsh-HadamardX	Maximum of difference in Walsh-Hadamard transform coefficients in value X
MaxDifWalsh-HadamardY	Maximum of difference in Walsh-Hadamard transform coefficients in value Y
difCoefFFT-1 … 30	Difference between coefficients FFT1… FFT30 between pattern and slice
R-CoefFFT’	Coefficient R of FFT covariance between pattern and Slice

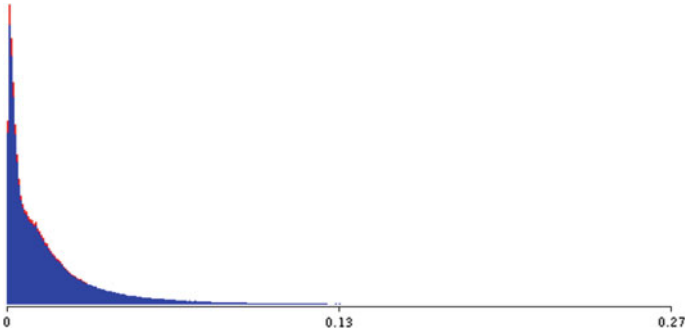


Fig. 3.6 Dispersion of ***ScoreParcial***

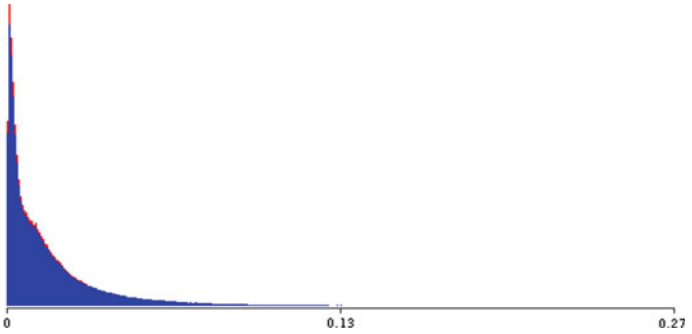


Fig. 3.7 Dispersion ***MaxDifWalsh-HadamardX***

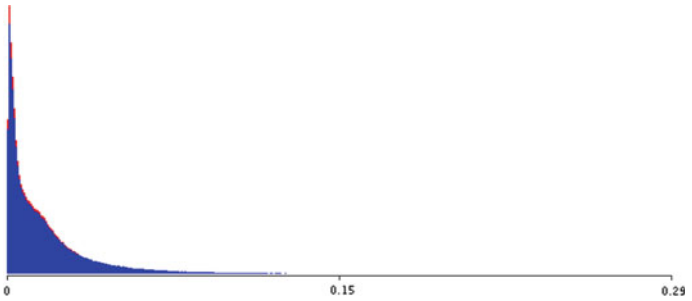


Fig. 3.8 Dispersion ***MaxDifWalsh-HadamardY***

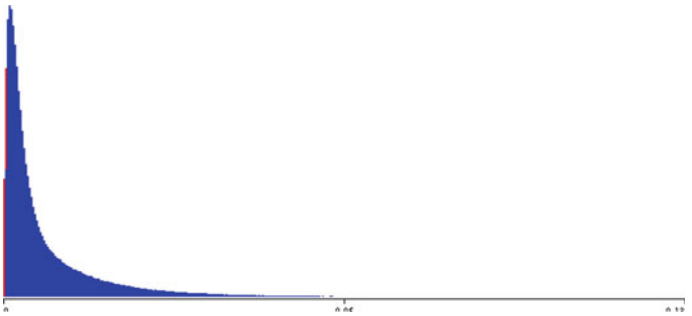


Fig. 3.9 Dispersion Pattern4 ***RcoefFFT***

a single timestamp starting the pattern, because of the sampling rate—A good model will be able to determine if this corresponds to the Pattern4 or a similar sound.

The next step in WEKA, for attribute selection process starts with all attributes except those discarded previously.

(i) **Subset considering class *ScoreParcial***

The algorithm for attribute selection requires an Attribute selector and a search method. As an standard approach the ***selector*** recommended by WEKA by default is ***CfsSubsetEval***, that evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.

The *Search* method selected is ***BestFirst***. This way, the algorithm Searches the space of attribute subsets by greedy hill-climbing augmented with a backtracking facility. Setting the number of consecutive non-improving nodes allowed controls the level of backtracking done. Best first may start with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or start at any point and search in both directions (by considering all possible single attribute additions and deletions at a given point).

The test runs with 39 attributes and 100,000 records. The evaluation mode is "evaluate on all training data". The attribute Subset Evaluator runs supervised, considering the *class* of type numeric ***ScoreParcial***. The CFS Subset Evaluator includes locally predictive attributes.

The selected attribute with 9: 1 scoring is ***R-CoefFFT***.

The result can be explained as currently, scoring depends on FFT differences between FFT in X and Y, so the algorithm is detecting the deep relationship between both measures and not information dependency with the entire set of features.

The same test is performed with *Selector* ***CfsSubsetEval*** and *Search* method ***Greedy Stepwise***, a greedy forward (or backward) search through the space of attribute subsets. May start with no/all attributes or from an arbitrary point in the space. Stops when the addition/deletion of any remaining attributes results in a decrease in evaluation. Can also produce a ranked list of attributes by traversing the space from one side to the other and recording the order that attributes are selected. The parameters selected are for forward search, starting with no attribute.

As in the previous test, the selected attribute with 9:1 scoring is ***R-CoefFFT***, Merit of best subset found: 0.159 also identical to ***BestFirst***.

From all this analysis, the ***ScoreTotal*** depends mainly on the covariance coefficient of the FFT values of the patterns. The rest of the thresholds determined for closeness between pattern and slice features contribute less to it.

The next section evaluates, beyond ***ScoreTotal***, how features contribute to ***R-CoefFFT***.

(ii) **Subset considering class *R-CoefFFT***

In order to explore how features contribute to values of FTT covariance, ***ScoreTotal*** y ignored and the algorithm runs with *Evaluator* ***CfsSubsetEval*** and *Search* ***BestFirst***. Class is ***R-CoefFFT***.

Results Merit of the best subset found rises up to 0.44 selecting only ***AVG-DifX*** attribute, as relevant. Scoring depends on FFT differences between FFT in X and Y, so the algorithm shows the deep relation between both measures. The same Evaluator with ***GreedyStepwise*** results in the same variable and merit (see Table 3.12).

Table 3.12 Ranked attributes with GreedyStepwise

Order	ID	Attribute		Order	ID	Attribute
1	19	DifCoefFFT-11		17	31	DifCoefFFT-23
2	16	DifCoefFFT-8		18	32	DifCoefFFT-24
3	17	DifCoefFFT-9		19	34	DifCoefFFT-26
4	11	DifCoefFFT-3		20	21	DifCoefFFT-13
5	10	DifCoefFFT-2		21	35	DifCoefFFT-27
6	9	DifCoefFFT-1		22	36	DifCoefFFT-28
7	7	MaxDifWalsh-HadamardY		23	30	DifCoefFFT-22
8	2	AVG-DifX		24	29	DifCoefFFT-21
9	3	AVG-DifY		25	28	DifCoefFFT-20
10	4	AvgCovX		26	27	DifCoefFFT-19
11	5	AvgCovY		27	22	DifCoefFFT-14
12	6	MaxDifWalsh-HadamardX		28	23	DifCoefFFT-15
13	18	DifCoefFFT-10		29	24	DifCoefFFT-16
14	20	DifCoefFFT-12		30	25	DifCoefFFT-17
15	37	DifCoefFFT-29		31	26	DifCoefFFT-18
16	33	DifCoefFFT-25		32	1	Timespamp

The *Evaluator* **ClassifierAttributeEval** evaluates the worth of an attribute by using a user-specified classifier. This time a simple ZEROR is selected (a tree with just one node). The Search method is leaded by a *RANKER*, that ranks attributes by their individual evaluations. Works in conjunction with attribute evaluators (ReliefF, GainRatio, Entropy etc.). In the case of ***ZeroR***, it is GainRatio. Table 3.13 shows the ranking.

Relevance of attributes are: 38, 14, 12, 13, 15, 19, 16, 17, 11, 10, 9, 7, 2, 3, 4, 5, 6, 18, 20, 37, 33, 31, 32, 34, 21, 35, 36, 30, 29, 28, 27, 22, 23, 24, 25, 26, 1. A few FFT coefficients characterize the shape of the sound wave of Pattern4. Also the difference in fast transform Walsh Hadamard. They can be used to perform a fast testing in real time instead of all the 39 features.

(iii) **PCA**

Since all nominal attributes were removed in pre-processing, Principal Components Analysis (PCA) can be useful to identify and avoid inter-dependencies among features. Ir works in conjunction with a Ranker search. Dimension reduction is accomplished by choosing enough eigenvectors to account for some percentage of the variance in the original data-default 0.95 (95%). Attribute noise can be filtered by transforming to the PC space, eliminating some of the worst eigenvectors, and then transforming back to the original space. But it is not considered for obtaining a new rotated set of axis but to evaluate the information provided by the attributes.

Table 3.13 Ranked attributes **ClassifierAttributeEval** for R-coefFFT

Order	Score	Attribute		Order	Score	Attribute
1	38	DifCoefFFT-30		19	20	DifCoefFFT-12
2	14	DifCoefFFT-6		20	37	DifCoefFFT-29
3	12	DifCoefFFT-4		21	33	DifCoefFFT-25
4	13	DifCoefFFT-5		22	31	DifCoefFFT-23
5	15	DifCoefFFT-7		23	32	DifCoefFFT-24
6	19	DifCoefFFT-11		24	34	DifCoefFFT-26
7	16	DifCoefFFT-8		25	21	DifCoefFFT-13
8	17	DifCoefFFT-9		26	35	DifCoefFFT-27
9	11	DifCoefFFT-3		27	36	DifCoefFFT-28
10	10	DifCoefFFT-2		28	30	DifCoefFFT-22
11	9	DifCoefFFT-1		29	29	DifCoefFFT-21
12	7	MaxDifWalsh-HadamardY		30	28	DifCoefFFT-20
13	2	AVG-DifX		31	27	DifCoefFFT-19
14	3	AVG-DifY		32	22	DifCoefFFT-14
15	4	AvgCovX		33	23	DifCoefFFT-15
16	5	AvgCovY		34	24	DifCoefFFT-16
17	6	MaxDifWalsh-HadamardX		35	25	DifCoefFFT-17
18	18	DifCoefFFT-10		36	26	DifCoefFFT-18
				37	1	Timespamp

As ***ScoreParcial*** is a derived formula provided it depends mainly on ***R-CoefFFT***, it is not considered for this analysis. Table 3.12 shows the output of PCA, denoting that every variable and the 25 first FFT coefficients are relevant.

Selected attributes are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25: a total of 25 of 38. Although just ***R-cofFFT*** describes properly the pattern detection criteria, the entire set of features until ***coefFFT-18*** provide relevant information about the signal, and therefore must be kept in the data-set analysis. The rest of the FFT coefficients are not relevant.

3.6.3 Automatic Timestamp Detection

To validate the timestamp of the automatically processed slices, a comparison against hand made table of occurrences is performed. Since the previous analysis detected that there are "hit ranges" of timestamps instead of single matches, a clustering is performed to find the scope of the regions. The same clustering with the entire set, serves to evaluate the matching performance. Hit ranges are in the reduced set

Table 3.14 Timestamp for hand made matching of Pattern4

Time (min)	Time (s)	Timestamp	Test set	Stimulus
4.16	256	608,697	1	4.14′
7.23	443	1,053,331	2	2.22′
10.10	610	1,450,412	2	10.08′
14.10	850	2,021,065	2	14.09′
16.35	995	2,365,835	3	6.33′

Table 3.15 Matching of Pattern4

Timestamp	Time	MaxDifWalsh-HadamardX	MaxDifWalsh-HadamardY	R-coefFFT
621,804	4′ 20″	0.000655651	0.000399113	0.320047
729,992	5′ 7″	0.000473976	0.00031662	0.449992

containing only registers marked with highest ***ScoreTotal*** (or ***R-coefFFT***, as shown previously, works a the metric corresponding to the ***ScoreParcial*** indicator).

(i) **Validation with hand made registering**

Table 3.1 shows part of a hand made Time-sheet of Stimulus—sound detected. For Pattern 4, the corresponding entry details are in Table 3.14.

Every entry has a stimulus associated. In the case of Pattern 4, the stimulus is in the background sound: birdsong. The conversion to timestamp is using the reduced sampling rate determined during pre-processing.

Table 3.15 shows the corresponding timestamps of hits detected in the first test set (from timestamp 1 to 1,048,576), representing the first portion of the video, with a length of 7′ 21″.

It can be seen that there is one more matching not detected previously, as those timestamps correspond to a fast and loud aggregation of sounds. In both cases MaxDifWalsh-HadamardX, MaxDifWalsh-HadamardY and R-coefFFT have values different from the rest. The converted timestamp to minutes also has certain differences probably due to the ability of the evaluator to calculate the time. Figures 3.10 and 3.11 shows the dispersion of MaxDifWalsh-HadamardX and R-coefFFT, highlighted the timestamps with best matching of Pattern4. As the case of the rest of the figures, they are plots from WEKA.

Figure 3.10 shows detected matchings as minimal values of the MaxDifWalsh-HadamardX (maximum of the difference for Walsh Hadamard coefficients in X range).

For R-CoefFFT (Fig. 3.11), there are some extra values around the optimum of the selected timestamp, due to the sampling rate, the granularity is big. In the following analysis (clustering with the entire set) there are two timestamps ranges of hits (that also have ScoreParcial, ScoreTotal, and R-coefFFT hight) instead of one (see portions in red of Fig. 3.12). Despite this is an improvement to the hand-made registration

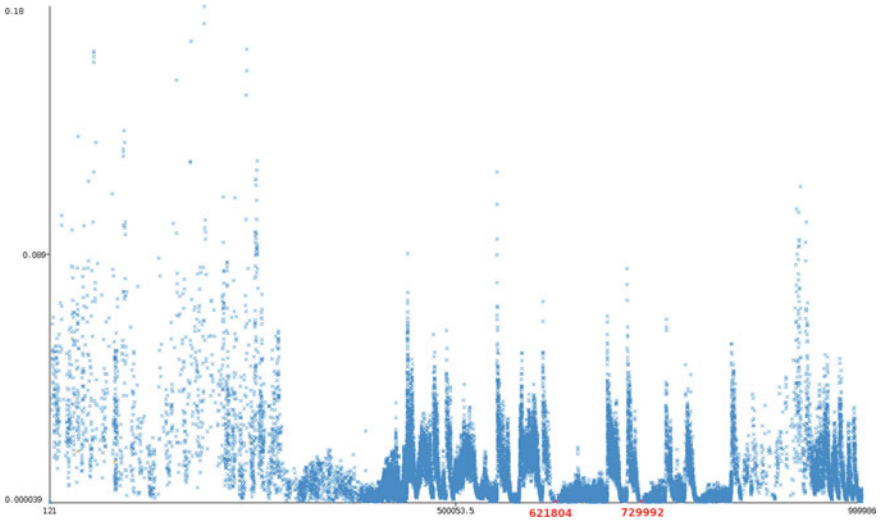


Fig. 3.10 Timestamp versus MaxDifWalsh-HadamardX

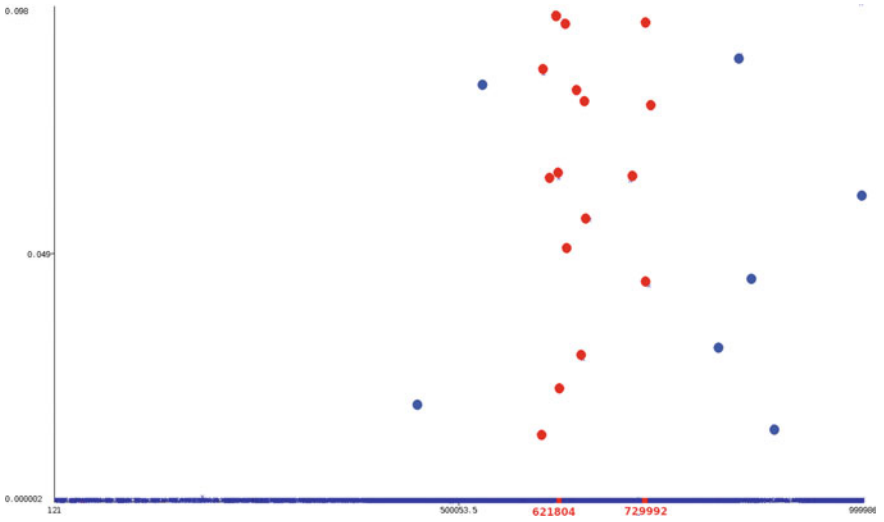


Fig. 3.11 Timestamp versus R-CoefFFT

(with only one match), it is not good enough to precise when the pattern starts. The model with NN fixes this problem (see section "Model for timestamp detection").

(ii) **Clustering with the entire set**

To cluster instances Simple K-Means is the fastest algorithm to process the set 1. Parameter K = 2 clusters, for detecting if the model is able to split features related

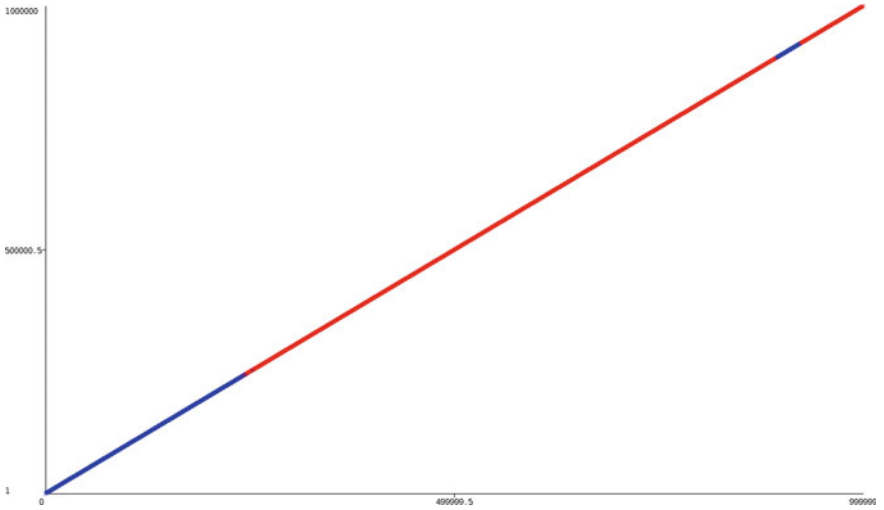


Fig. 3.12 Matching of Pattern4: Timestamp versus instance number

to Pattern4, all the features selected in the previous section are used. Other parameters are: distance function Euclidean Distance, maxIterations (maximum number of iterations) 500, numExecutionSlots (number of execution threads to use. Must be equal to the number of available cpu/cores) 1, dontReplaceMissingValues (Replace missing values globally with mean/mode) false, preserveInstancesOrder (Preserve order of instances) false, InitializationMethod (The initialization method to use) Random, fastDistanceCalc (Use cut-off values for speeding up distance calculation, but suppressing the calculation and output of the within cluster sum of squared errors/sum of distances) false, and Seed value of 10. It takes 9 s to build the clusters for 1,000,000 records, Within cluster sum of squared errors: 271,479.03. Table 3.16 has the final cluster centroids.

It detects that there are 2 clusters one with low R-CoefFFT and the other with high values. It corresponds to MaxDifWalsh values. A total of 272,799 Clustered Instances (27%) belong to cluster 1 (lower differences) and 727,201 (73%) are in cluster 0. Values of difCoedFFT-1 to difCoedFFT-18 are fewer in cluster 1 than in cluster 0, due to the proximity of those instances to the pattern. Several sounds that aren't Pattern 4 are in the cluster. This implies that K-means is not good enough to detect with precision where the pattern matches. Figure 3.12 shows the cluster distribution of Pattern4.

The cluster detects the two main time periods when the pattern is present (portions in red), but the model lacks of enough specificity.

(iii) **Clustering with reduced set**

The reduced set of attributes are 25 (the original data-set has 39 attributes), and 42,682 registers (drifted from the starting 1,048,576 instances). The clustering process now takes only 0.81 s and clustered instances are organized as follows: cluster 1 with

Table 3.16 Clustering details

		Cluster#	
Attribute	Full data	0	1
	(1,000,000.0)	(727,201.0)	(272,799.0)
AVG-DifX	0.0387	0.0131	0.0482
AVG-DifY	0.0338	0.0114	0.0422
MaxDifWalsh-HadamardX	0.0173	0.0374	0.0097
MaxDifWalsh-HadamardY	0.0166	0.0353	0.0096
R-CoefFFT	0.0075	0.0047	0.015
difCoefFFT-1	2.95	3.5736	2.7161
difCoefFFT-2	2.5176	3.35	2.2054
difCoefFFT-3	2.1955	2.9226	1.9227
difCoefFFT-4	2.6231	3.6645	2.2324
difCoefFFT-5	2.4688	3.5241	2.0729
difCoefFFT-6	2.595	3.4923	2.2584
difCoefFFT-7	4.8701	5.4453	4.6543
difCoefFFT-8	3.6354	4.3074	3.3834
difCoefFFT-9	2.529	3.3567	2.2185
difCoefFFT-10	2.0781	3.1077	1.6918
difCoefFFT-11	4.0816	4.8391	3.7975
difCoefFFT-12	3.3608	4.2408	3.0307
difCoefFFT-13	3.0082	3.9483	2.6555
difCoefFFT-14	2.0709	3.7964	1.4237
difCoefFFT-15	2.1553	3.0641	1.8144
difCoefFFT-16	2.0514	3.1202	1.6505
difCoefFFT-17	2.4121	3.4797	2.0117
difCoefFFT-18	2.1623	3.1289	1.7998

8272 (19%), and cluster 0 with 34,410 (81%). Although there are fewer timestamps classified as matching zone, the precision is still poor.

The hit found by using the main features have matching ranges, because the sampling rate is high (44,100 samples per second) compared to the time granularity for speech instances in language processing (seconds).

The initial hit rate, just considering R-Coeff reduced set 1 from 1,000,000 to 42,682 records. (The adjusted hit with a combination of MaxDifWalsh-Hadamard, AVG-Dif, AVG-Cov and difCoefFFT-1 to difCoefFFT-18 improved fitness but it needs to be even better. I the following sections Neural Model is able to perform a better classification of the true matching.

Table 3.17 NN model with test sets

Set	instances	Correctly classified instances (%)	Incorrectly classified instances
1	1,048,576	98.62	1
2	1,048,576	91.98	5
3	1,046,199	99.14	2

3.6.4 Model for Timestamp Detection

This section describes the simple neural model for timestamp detection of Pattern4. As precision of AvgCovX and have scientific notation, AvgCovXe15 instead which is 10^{15} times smaller. data-sets are considered in a reduced version with only relevant attributes in the previous analysis: Timespamp, AVG-DifX, AVG-DifY, MaxDifWalsh-HadamardX, MaxDifWalsh-HadamardY, R-CoefFFT, AvgCovXe15, AvgCovYe15 and HIT.

Table 3.17 is the summary of the models obtained, using Multilayer Perceptron of 9 input nodes, one hidden layer of 5 units and one output unit. In all cases Test mode is split using 66.0% for training and remainder for testing.

Kappa statistics (0.98) and classification results suggest that the model is good for detecting Pattern4.

3.6.5 Model Findings and Results Analysis

Table 3.18 is Table 3.14 extended with the NN model findings: timestamp found with the Perceptron network, and the number of records representing the hit (column Length derived).

There are two new rows with extra matchings detected automatically: timestamp 1,224,527 and 1,930,712. From these results it can be said that matching timestamp is more accurate when the model finds a matching, and that the number is higher than in hand-made evaluation of the recording.

Table 3.18 Timestamp for hand made matching of Pattern4

Time (min)	Time (s)	Timestamp	Test set	Timestamp derived	Length derived	Stimulus
4.16	256	608,697	1	608,697	2	4.14′
7.23	443	1,053,331	2	1,053,332	5	2.22′
8.35	515	1,224,527	2	1,224,504	2	8.34′
10.10	610	1,450,412	2	1,449,996	3	10.08′
13.32	812	1,930,712	2	1,928,767	2	13.29′
14.10	850	2,021,065	2	1,998,262	1	14.09′
16.35	995	2,365,835	3	2,365,876	4	16.33′

3.7 Conclusions and Future Work

This paper presents an introduction of main characteristics of patients suffering Autistic spectrum Disorder. Those that has not preserved verbal behavior have severe problems to communicate. They can't articulate any language in the sense that has been traditionally defined. After analyzing the recording of a patient's interaction, it is possible to determine several repetitive movements and sounds, that are present upon certain stimuli. The sounds can be considered as part of a pattern to evaluate when it is a verbal behavior during the interaction. All the patterns have certain basic characteristics. FFT coefficients are not enough to detect similarities between a pattern and the sound track. Considering the covariance of the difference between pattern and wav file, of coefficients in fast transform Walsh Hadamard and FFT are the best option. They can be used to perform a fast testing in real time. Several parameters and coefficients were considered but are not relevant. The model derived in this chapter is a neural Network, using as parameters the covariance of differences between FFT coefficients and Walsh Hadamard. The test made with the use case is able to detect one matching not detected previously, and more accurate timestamps. The findings in this paper might be considered as a first step to evaluate other patient's interactions and their features. Other pending activities are: implement a prototype that makes it simpler to drift off sound patterns and evaluate them automatically, apply similar considerations to video recordings, track the evolution in time of the communication of the individual.

References

1. Filipek, P.A., Accardo, P.J., y Ashwal, S.: Parameter practice: detection and autistic diagnose. Neurology **55**, 468479 (2000)
2. Maenner, M.J., Shaw, K.A., Baio, J., et al.: Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 Sites, United States. In: MMWR Surveillance Summaries 2020, vol. 69(No. SS-4), pp. 1–12 (2016)
3. Zablotsky, B., Black, L.I., Maenner, M.J., Schieve, L.A., Danielson, M.L., Bitsko, R.H., Blumberg, S.J., Kogan, M.D., Boyle, C.A.: Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. Pediatrics **144**(4), e20190811 (2019)
4. Mayada, et al.: Global prevalence of autism and other pervasive developmental disorders. Autism Res. **5**(3), 160–179 (2012)
5. Sevilla, M.d.S.F., Bermúdez, M.O.E., Sánchez, J.J.C.: How many people have autism? Theor. Rev. Int. J. Dev. Educ. Psychol. INFAD. N1 V1, 769–786 (2013). ISSN 0241-9877
6. Greer, R.D.: The wide application of behavior analysis to schooling behavior (CABAS®). Behav. Soc. Matter
7. Krans, B.: Small humanoid robot helps children with ASD. Healthline News (2013). Obtained from http://www.healthline.com/healthnews
8. Greer, R.D.: The comprehensive application of behavior analysis to schooling. (CABAS®). Behav. Soc. Issues **7**(1) (1997)
9. Bustamante, P., Lafalla, A., Coria, N., Agüero, M., López de Luise, D., Parra, C., Azor, R., Moya, J., Cuesta, J.: Protocol for evaluation of somatic and oral behavior to sound stimuli in patients with autism spectrum. CI2S Lab. Pringles 10- P2. University of Burgos (2016)

10. Kuiper, M.W., Verhoeven, E.W., Geurts, H.M.: Stop making noise! auditory sensitivity in adults with an autism spectrum disorder diagnosis: physiological habituation and subjective detection thresholds. J. Autism Dev. Disord. **49**(5), 2116–2128 (2019). https://doi.org/10.1007/s10803-019-03890-9
11. Summers, J., Shahrami, A., Cali, S., et al.: Self-injury in autism spectrum disorder and intellectual disability: exploring the role of reactivity to pain and sensory input. Brain Sci. **7**(11), 140 (2017). Published 26 Oct 2017. https://doi.org/10.3390/brainsci7110140
12. Marsh, K.L, Isenhower, R.W., Richardson, M.J., et al.: Autism and social disconnection in interpersonal rocking. Front. Integr. Neurosci. **7**, 4 (2013). Published 18 Feb 2013. https://doi.org/10.3389/fnint.2013.00004
13. Mazefsky, C.A., Herrington, J., Siegel, M., et al.: The role of emotion regulation in autism spectrum disorder. J. Am. Acad. Child. Adolesc. Psychiatry. **52**(7), 679–688 (2013). https://doi.org/10.1016/j.jaac.2013.05.006
14. Mazefsky, C.A., White, S.W.: Emotion regulation: concepts & practice in autism spectrum disorder. Child. Adolesc. Psychiatr. Clin. N. Am. **23**(1), 15–24 (2014). https://doi.org/10.1016/j.chc.2013.07.002
15. van Santen, J.P., Sproat, R.W., Hill, A.P.: Quantifying repetitive speech in autism spectrum disorders and language impairment. Autism Res. Off. J. Int. Soc. Autism Res. **6**(5), 372–383 (2013). https://doi.org/10.1002/aur.1301
16. De Luise, D.L., Azor, R., Párraga, C.: Autistic verbal behavior. Automated model of the patient profile—audio. Edit. Académica Española (2015). ISBN 978-3-659-07183-6
17. Cheol-Hong, N., Ahmed, H.: Novel pattern detection in children with autism spectrum disorder using iterative subspace identification. Department of Electrical and Computer Engineering University of Minnesota, USA (2010). 978-1-4244-4296-6/10 ©2010 IEEE ICASSP 2010. PPS 2266/2269
18. Mower, E., Black, M.P., Flores, E., Williams, M., Narayanan, S.: Design of an emotionally targeted interactive agent for children with autism. University of Southern California (USC), 2011. IEEE (2011). 978-1-61284-350-6/11
19. De Luise, D.L.: Autistic verbal behavior parameters. Paper presented in: 9th International Workshop on Soft Computing Applications (SOFA 2020). Arad-Romania, 27–29 November 2020. Hosted by "Aurel Vlaicu" University of Arad (2020)
20. DSM-V Encoding update®. Supplement to the Diagnostic and Statistical Manual of Mental Disorders, 5th edn. American Psychiatric Association (2014)
21. Ministry of Science and Innovation Spain. Clinical practice guide for the management of patients with autism spectrum disorders in primary care. Nipo: 477-09-052-8. ISBN: 978-84-451-3244-9: Estilo Estugraf Impresores, S.L. Pol. Ind. Los Huertecillos, Nave 13 - 28350 Ciempozuelos (Madrid) (2009)
22. Skinner, B.F.: Verbal Conduct (Trans.: R. Ardilla, Edit. T. Mexico) (1981) ISBN 968-24-0987-X
23. Kanner, L. (1943). Autistic disturbances of affective contact. Nervous Child. N2, (217250) (Trans: T.S. Vicario). Published in the Spanish Journal of Intellectual Disability Siglo Cero. www.feaps.org
24. Wing, L.: The definition and prevalence of autism: a review. Eur. Child Adolesc. Psychiatry **2**, 61–74 (1993)
25. Riviere, A.M.: The little boy with autism (2000). ISBN: 9788460702610
26. Python in practice: create better programs using concurrency, libraries, and patterns (Developer's Library), 1st Edición. Mark Summerfield. ISBN-13: 978-0-321-90563-5
27. Fluent Python: Clear, concise, and effective. Ramalho, Luciano. ISBN-13: 978-1491946008
28. Beazley, D., Jones, B., Jones, K.: Python Cookbook, 3rd edn. O'Reilly Media, Inc. ISBN: 9781449340377
29. Automate the Boring Stuff with Python. Al Sweigart. Free to read under a Creative Commons License. ISBN-13: 9781593279929
30. Summerfield, M.: Python 3. Editorial: ANAYA MULTIMEDIA. (INFORMATICA). Año de edición: 2009 (2009). ISBN: 978-84-415-2613-6

31. Baghai-Ravary, L., Beet, S.: Automatic speech signal analysis for clinical diagnosis and assessment of speech disorders (2013). ISBN 978-1-4614-4574-6
32. Biing-Hwang, J., Rabiner, L., Wilpon, J.: On the use of bandpass liftering in speech recognition. IEEE Trans. Acoust. Speech Sig. Process. **35**(7), 947–954 (1987). https://doi.org/10.1109/TASSP.1987.1165237
33. De Luise, D.L., Hisgen, D., Cabrera, A., y Morales Rins, M.: Modeling dialogs with linguistic wavelets. Theory Pract. Mod. Comput. **1**, 11–13 (2012)
34. De Luise, D.L.: Morphosyntactic linguistic wavelets for knowledge management. InTechOpen **8**, 167–189 (2012). https://doi.org/10.5772/35438
35. De Luise, D.L., Hisgen, D.: MLW and bilingualism: case study and definition of basic techniques. IGI global: Advanced Research and Trends in New Technologies, Software, Human-Computer Interaction, and Communicability, pp. 568–600 (2013)
36. Sundberg, M.L.: A brief overview of a behavioral approach to language assessment and intervention for children with autism association for behavior analysis newsletter, 30(3) (2007)
37. Greer, D.: The ontogenetic selection of verbal capabilities: contributions of skinner's verbal behavior theory to a more comprehensive understanding of language. Int. J. Psychol. Psychol. Ther. **8**(3), 363–386 (2008)
38. Two Pointer with Sliding Window Algorithm to Compute the Longest Substring with at most K Distinct Characters (2020) Taken from https://helloacm.com/
39. Fumarola, F., Ciampi, A., Annalisa, A., Malerba, D.: A sliding window algorithm for relational frequent patterns mining from data streams. In: 12th International Conference on Discovery Science (2009). http://dx.doi.org/10.1007/978-3-642-04747-3_30