

Procesamiento automático y manejo de Conducta Verbal Autista*

Daniela López De Luise**

Ben Raúl Saad, Pablo Pescio, Tiago Ibacache, Lucas Soria***

Resumen

Los pacientes con severo Trastorno del Espectro Autista (TEA) presentan severas alteraciones en la producción y entendimiento del lenguaje natural, y en consecuencia en la socialización. Este trabajo presenta una estrategia novedosa de procesamiento automático de la conducta verbal registrada en grabaciones de video durante la terapia. El trabajo comprende un estudio de caso, halla conductas repetidas típicas del paciente estudiado, y su relación con los estímulos del entorno mediante técnicas de Aprendizaje Automático (AA). De los parámetros que describen las señales, se establecen métricas de evaluación de manera sistemática y objetiva, para evidenciar las causas de las reacciones del paciente, asociándolas con patrones de comportamiento y así comprendiendo el estado emocional e intenciones del individuo.

Palabras clave: Trastorno Espectro Autista. Procesamiento de Lenguaje Natural. Lingüística. Procesamiento de sonido.

Abstract

Patients with severe Autistic Spectrum Disorder (ASD) present several alterations in the production and understanding of natural language, and consequently in socialization. This work

* Enviado: 14-02-21. Aceptado: 08-04-21.

** Especialista en Ingeniería De Sistemas Expertos, Doctora en Ciencias Informáticas y Especialista en Comunicación Pública de la Ciencia y la Tecnología. Directora del CI2S Lab. Directora del laboratorio de investigación IDTI y de la Diplomatura en Gestión de Calidad de Software en la Universidad Autónoma de Entre Ríos (UADER). Fundadora de la rama local y directora del IEEE Game Technical Committee Argentina. Fundadora del capítulo argentino del IEEE Computational Intelligence Society. Coordinadora técnica y fundadora del CETI en la Academia Nacional de Ciencias de Buenos Aires. Coordinadora Outreach de proyectos STEAM. Docente en Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Universidad en la Abierta Interamericana (UAI) y la UADER. Miembro del Grupo de investigación GIBD (Grupo de Investigación de Bases de Datos). Consultora independiente en sistemas inteligentes (Computational Intelligence) y soft computing. Correo electrónico: daniela_idl@ieee.org.

*** IDTI Lab - Facultad de Ciencia y Tecnología – UADER. Concepción del Uruguay.

presents a novel strategy for the automatic processing of verbal behavior recorded in video recordings during therapy. The work includes a case study, finds typical repeated behaviors of the patient studied, and their relationship with the stimuli of the environment through Machine Learning (ML) techniques. From the parameters that describe the signals, metrics are evaluated in a systematic and objective way, to show the causes of the patient's reactions, associating them with behavior patterns and thus understanding the emotional state and intentions of the individual.

Keywords: Autistic Spectrum Disorder. Natural Language Processing. Linguistics. Sound Processing.

Resumo

Pacientes com Transtorno do Espectro Autista (TEA) grave apresentam várias alterações na produção e compreensão da linguagem natural e, conseqüentemente, na socialização. Este trabalho apresenta uma nova estratégia para o processamento automático do comportamento verbal registrado em gravações de vídeo durante a terapia. O trabalho inclui um estudo de caso, encontra comportamentos típicos repetidos do paciente estudado e sua relação com os estímulos do ambiente por meio de técnicas de Aprendizado de Máquina (AM). Dos parâmetros que descrevem os sinais, as métricas são avaliadas de forma sistemática e objetiva, para mostrar as causas das reações do paciente, associando-as aos padrões de comportamento e, assim, compreender o estado emocional e as intenções do indivíduo.

Palavras-chave: Transtorno do Espectro Autístico. Processamento de Linguagem Natural. Lingüística. Processamento de Som.

Introducción

El Trastorno del Espectro Autista (TEA) es un desorden caracterizado por una serie de síntomas que pueden resumirse en alteraciones del lenguaje, conductas estereotipadas y desordenes en la interacción social (Filipek, Accardo y Ashwal, 2000). Existen distintas estadísticas sobre el impacto TEA en el mundo. El rango puede ir desde 1 en 80 niños hasta 1 en 160 (Elsabbagh, Divan, Koh, Kim, Kauchali, Marcín, Montiel-Nava, Patel, Paula, Wang, Yasamy y Fombonn, 2012), pero el número depende de los métodos de análisis, regiones y años, debido a diferencias en métodos, criterios diagnósticos, diseño de muestras, franjas

etarias, etc. Pero en la comunidad existe el consenso de que hay una incidencia global creciente (Fortea Sevilla, Escandell, Bermúdez y Castro Sánchez, 2013). Los pacientes autistas tienen especial dificultad en la identificación e interpretación de información social, la que les es compleja. Estas alteraciones suelen acompañarse de otros síntomas cognitivos y físicos, según el grado de severidad (Greer, 1997).

De acuerdo al manual *Diagnostic and Statistical Manual of Mental Disorders* (DSM) de la asociación de psiquiatría de Estados Unidos de Norteamérica, en el grado más severo el paciente sufre de la imposibilidad completa de articular lenguaje (Bustamante y col., 2016). También se asocia a desórdenes de procesamiento sensorial (Sensory processing disorder, SPD), condición por la que una persona no responde con normalidad a sonidos, aromas, texturas u otros estímulos. Otras alteraciones son las conductas estereotipadas y repetitivas de índole auto-estimuladorio. Algunas de ellas son agitación de manos, balanceos, giros, repetición de palabras o ruidos, conductas que pueden verse amenazantes o atemorizantes (Kuiper, Verhoeven y Geurts, 2019) (Summers, Shahrami, Cali, Summers, Shahrami, D’Mello, Kako, Palikucin-Reljin, Savage, Shaw y Lunskey, 2017) (van Santen, Sproat y Hill, 2013). Es típico que el individuo no pueda controlar estas manifestaciones debido a que no puede manejar la información sensorial (Marsh, Isenhower, Richardson, Helt, Verbalis, Schmidt y Fein, 2013) (Mazefsky, Herrington, Siegel, Scarpa, Maddox, Scahill y White, 2013). No está establecida aún la relación entre las emociones y dichas conductas, denominadas “*stimming*” (Mazefsky y White, 2014), que son las que conforman la conducta verbal del autista y por lo tanto son centro de estudio en este trabajo. Entre otras cuestiones, el objetivo es discriminar si estas actitudes también son una suerte de comunicación rudimentaria, una mera distracción, o representa otra situación.

La correcta discriminación del origen y características representa una poderosa información que permite comprender lo que sucede internamente al paciente. Es importante destacar que cada paciente tiene su propio conjunto de manifestaciones; a diferencia de otros trabajos, en esta eventualidad se toma como una característica diferencial del individuo y no como algo azaroso y caótico. La hipótesis de trabajo fundamental aquí es que el conjunto de *stimmings*, pueden modelarse con mecanismos de aprendizaje automático y pueden establecer una luz sobre cómo el paciente TEA se relaciona con su entorno.

Si bien este trabajo cuenta con algunas publicaciones previas (López De Luise, Azor y Párraga, 2015) (López De Luise, 2020) ahora se establece de manera automática el tipo de relación causal en un caso de estudio. Dada la complejidad del estudio, el alcance se ha limitado

al análisis del sonido extraído de las grabaciones de video generadas durante una sesión de terapia. Existen otros trabajos como los de Cheol-Hong, N. y Ahmed H. (2010), que emplean un acelerómetro y colocan en un espacio ortogonal la clasificación de sus resultados. Pero estos dispositivos no solo proveen información muy limitada, sino que también debe recordarse que el paciente es totalmente sensible al tacto y colocación de sensores en su cuerpo o vestidura. Lo que hace impracticable esta lectura en casos severos. Lamentablemente muchos de los trabajos en campo se basan en la colocación de sensores, dejando fuera al rango de individuos que más necesitan este tipo de estudios.

La Universidad del Sur de California (Estados Unidos de Norteamérica) trabaja en otra línea basada en agentes interactivos (Mower, Black, Flores, Williams y Narayanan, 2011). Su prototipo, llamado Rachel, conversa con el paciente y su padres, generando escenarios vinculados a emociones para valorar las habilidades comunicativas del TEA. Esta otra vertiente, además de no ser diagnóstica ni terapéutica, tiene el problema de basarse en la vocalización y administración de la lengua natural, lo que deja nuevamente fuera a los casos intermedios y severos.

Al día de hoy, mucho de la evaluación de progreso se deja registrado en tablas manuales y su evaluación es más bien subjetiva. Este trabajo provee junto con los anteriores, las bases para que eso cambie. De hecho en muchos de los casos los progresos detectados son subjetivos y que el paciente, evaluado correctamente no presenta en realidad una mejoría sino un cambio de comportamiento. Otra de las ventajas del modelo automático que se propone en este trabajo es que, con la adecuada adaptación, puede ser una herramienta de trabajo para el terapeuta, y de registro de progresos comparados (López De Luise, D., 2020).

Lo que sigue de este trabajo presenta algunas consideraciones básicas de trabajo (sección II), presenta los fundamentos del proyecto denominado BIOTECH (sección III), el caso de estudio (sección IV), conclusiones y trabajos futuros (sección V).

Consideraciones Iniciales del Proyecto BIOTECH

A fin de contextualizar el proyecto, es menester introducir algunas precisiones referentes al trastorno autista. TEA comprende un espectro de alteraciones que pueden presentarse en diverso grado. La *American Psychiatric Association* (APA), ya mencionada en la introducción, mantiene actualizado un manual que detalla cuestiones como estas. En su quinta edición

llamada DSM-V (American Psychiatric Association, 2014) establece una guía de clasificación y diagnóstico en tres niveles de severidad. Hoy día se considera esta una guía básica en la comunidad (Ministerio de Ciencia e innovación de España, 2009), a pesar de lo cual, los especialistas concuerdan en que cada paciente tiene características únicas.

De los grados definidos por la APA, el más leve es el grado 1 donde prevalece algún tipo de comunicación social y limitados intereses. A pesar de ello pueden comunicarse en frases completas aunque con serias deficiencias de tipo organizativo, rigidez y comprensión real del habla. Por caso el doble sentido o la sátira. En el otro extremo el grado 3 tiene severas dificultades en focalizar su atención, imposibilidad de desarrollo independiente y comunicación. La rigidez en su conducta hace extremadamente difícil su funcionamiento social e individual.

Dado que el TEA severo no puede expresarse a través del lenguaje vocal, los autores de este trabajo remiten a la conducta verbal propuesta por Skinner B.F. (1981). Sus estudios proponen que el lenguaje vocal, que se conoce como lenguaje natural, es solo una parte del lenguaje que el individuo emplea para comunicarse. Según Skinner la emisión de sonidos o ciertas acciones, gestos y cambios corporales acompañan al lenguaje organizado del habla puesto que generan reacciones en el interlocutor.

En el contexto autista, las acciones repetitivas y estereotipadas reconocidas desde los primeros estudios de Leo Kanner como típicas del TEA (Kanner, 1943) (Wing, 1993) conforman la composición completa de comunicación verbal no vocal.

El Proyecto BIOTECH

Esta sección introduce las decisiones principales y características del proyecto BIOTECH en el contexto del que se desarrollan las investigaciones base de este trabajo.

La Figura 1 muestra un esquema general de cómo se organiza el trabajo del proyecto. El estudio comienza con el registro video-grabado de una interacción con el paciente TEA.

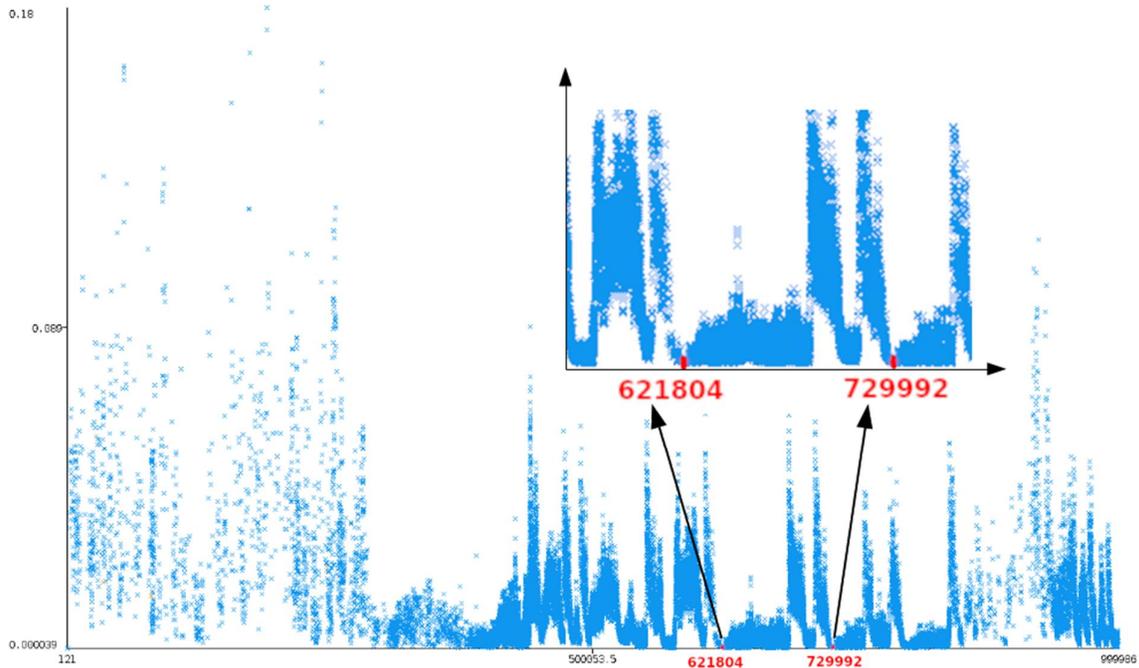


Figura 1. Estrategia global de BIOTECH

A partir de la misma se extrae el audio y el video. De cada uno, se genera el modelo por medio de técnicas de aprendizaje automático. Es de notar que el modelo se compone para cada individuo por las razones indicadas en la introducción y en la sección previa. El modelo evidencia ciertos parámetros y permite extraer metadatos por procesos de inferencia heurística. De los parámetros se pueden asociar métricas e indicadores que constituyen evaluaciones objetivas para monitorear la evolución del paciente.

Con base en los metadatos y el estudio de *timestamp* (momento en que se registra el metadato) se dispone de suficiente información como para generar un modelo de asociación estímulo-conducta verbal, y con el mismo alimentar al modelo de *Wavelet* Morfosintáctica (MLW por sus siglas en inglés) del Lenguaje TEA (López De Luise, Hisgen, Cabrera y Morales Rins, 2012a). El alcance de este trabajo cubre el proceso de generación de metadatos y del modelo de asociación para los registros de audio, dejando de lado todo el resto del proceso. Tampoco se incluye el proceso análogo de la parte visual que aún permanece en estado manual.

Materiales y métodos

Para la realización del proyecto así como los casos de uso se emplean tecnologías de acceso simple y económico dado que la intención es que los prototipos queden a disposición de la comunidad. El equipo empleado es una cámara Sony Handycam HDR-CX250, con 5 canales

de sonido, colocado convenientemente para que el paciente no la viera. Gran parte del procesamiento se realiza en un procesador de 5 núcleos, Intel Core i7 con procesador de séptima generación, RAM de 8Gb, 64 bits con sistema operativo Ubuntu 20.04.2 LTS. Para algunos otros procesos se emplea un procesador AMD E-450, Dual Core de segunda generación, RAM de 8Gb, también de 64 bits con sistema operativo Ubuntu 20.04.2 LTS.

Los módulos del proyecto actualmente constituyen un prototipo, que está en su sexta versión. Se implementan versiones combinadas de Python y Octave, dado que ambos son estándar (Beazley, Jones, y Jones, 2013) en el ámbito de procesamiento de señales, análisis estadístico y ciertos procesos de minería básica. La versión actual recurre a librerías Octave (Eaton, Bateman, Hauberg y Wehbring, 2017) como LTFAT, Señal y Control para procesamientos relacionados con la Transformada rápida de Fourier (en adelante FFT), covarianzas, wavelets (en adelante WV) y transformada rápida de Walsh-Hadamard (FWHT) (Ahmed, N. y Rao K.R., 1975).

Las limitaciones de FFT son compensadas con WV, especialmente las relacionadas con la segmentación del audio en tiras o patrones, que representan los *stimmings* (Spilsbury y Euceda, 2019). Pero el estudio especial de estas circunstancias, así como las bondades de WV por encima de FFT caen fuera del estudio de este trabajo. El proyecto BIOTECH también aplica *Morphosyntactic Linguistic Wavelets* (MLW), una estrategia de descomposición lingüística que infiere información cruzada y estratégica del proceso de razonamiento detrás de expresiones de la lengua (López De Luise, Hisgen, Cabrera y Morales Rins, 2012a). Fue aplicada en el pasado para interpretación automática de contenidos WEB, resúmenes automáticos, perfilado de diálogos, textos y tipos de autores, efectos del bilingüismo precoz, etc. (López De Luise, 2012b) (López De Luise y Hisgen, 2013).

Si bien la presentación detallada de MLW trasciende los alcances de esta publicación, sí conviene destacar que la misma requiere de expresiones del lenguaje vocal oral o escrito. Dado el contexto TEA grado 3, no se cuenta con articulaciones lingüísticas. Para compensar esta situación se extiende el concepto de vocalización remitiendo al de conducta verbal propuesta por J. Skinner (Sundberg, 2007), reemplazando las producciones del lenguaje con patrones de audio y video procesadas automáticamente por técnicas de Aprendizaje Automático.

Dicha estrategia, basada en comportamiento verbal ha sido previamente adoptada para terapias TEA. Según Greer (Greer, 2008) las evidencias apuntan a que los orígenes del lenguaje apoyan esta postura, puesto que la habilidad humana del habla sería el resultado evolutivo y del desarrollo onto-genético. Es importante reiterar que los modelos obtenidos (ver Fig. 1) son

personalizados (Greer, R.D., 1997) y no universales. En ello recae la necesidad de contar con un prototipo que lo derive de manera transparente, objetiva y automática.

Constitución de la base de datos

El caso en estudio se basa en la sesión video-grabada de un paciente con grado 3 de autismo, de 13 años de edad, nativo de Argentina, con autorización de los padres. La actividad desarrollada en la sesión comprende las tareas típicas de identificación de imágenes, señalamiento y atención focalizada. El video tiene un total de 22:00 minutos, del cual se extrajo con el aplicativo gratuito VLC un audio formato wav de 233.3 Mb. El archivo de sonido tiene una tasa de muestreo $fs=44100$, con dos componentes que se denotan como X e Y. Dado el peso de los archivos csv asociados al convertirlos en Octave, se fraccionan en 3 partes.

Estos archivos se procesan para generar salidas con 39 metadatos asociados, que se hallan en la tabla 1, resumen de los archivos salientes. SalPattern4.Rami<x>.csv denota la salida para el patrón 4, del preproceso del sonido original Rami<x>, con <x> el identificador de cada archivo wav. Los datos se fraccionan para poder trabajar con los equipos disponibles por el equipo de proyecto. Dado que los resultados alimentan secuencialmente al modelo, dicho fraccionamiento no impacta en el sesgo final del modelo.

Archivo	Registros	Tiempo	Archivo procesado
SalPattern4.Rami1.csv	1048576	441"	SalPattern4.set1.csv
SalPattern4.Rami2.csv	1048576	441"	SalPattern4.set2.csv
SalPattern4.Rami3.csv	1046199	440"	SalPattern4.set3.csv
Total	3143351		

Tabla 1. Organización de archivos procesados.

Los archivos corresponden a una tasa reducida de 2.378 muestras por segundo, para poder trabajar más cómodamente sin perder información de detalle. La frecuencia se reduce luego de la lectura original a 44100Hz en Octave. La última columna muestra el archivo saliente del análisis-procesamiento de esta sección (inclusive selección de variables) realizado también con WEKA 3.8.4

Todo el procedimiento se registra paralelamente en formularios en papel, así como los *stimmings* con su *timestamp*, a fin de servir de contraste con lo hallado automáticamente. Los

tiempos, tipo de actividad, y estímulos referentes se vuelcan en una planilla denominada *videos.xls*. El registro manual se compone de las columnas: *timestamp*, actividad, características y comentarios. Esta información se toma sólo a los efectos de validación, y se extiende para considerar estímulos hasta 90 segundos antes y después del evento registrado. Este intervalo de tiempo se determina según hallazgos en estudios previos del mismo proyecto (ver López De Luise, D., 2020).

También se realiza una verificación manual exhaustiva para determinar el conjunto más relevante de *stimmings* del paciente; aquellos que se reiteran ante el mismo estímulo es considerado de interés. De acuerdo con dicho criterio la porción de audio trabajada se obtienen 8 comportamientos repetitivos bien diferenciados y otros 2 que podrían ser adicionales pero no han podido ser confirmados aún. Este primer paso sirve de control siendo parte de las conductas estudiadas en la comunidad y reportadas por especialistas como Ángel Riviere (2000).

Caso de estudio

El objetivo de este estudio de caso es evidenciar la dinámica y evaluar la efectividad de la estrategia descrita en la Figura 1, para la etapa de generación de metadatos y la asociación de estímulo-conducta verbal. Se destaca nuevamente que en ambos casos se trabaja solo con las producciones sonoras, dejando de lado las visuales (movimientos) y las asociaciones cruzadas entre ambos tipos de información.

El estudio toma el instante de detección del *stimming*, al que se refiere en este contexto como *timestamp*. La actividad se detecta como sonido asociado a *stimming* solo cuando un especialista la ha identificado como estereotipada y repetitiva. Adicionalmente se valida que las actividades sean algunas de las detectadas en los registros manuales por parte de los especialistas. Esto es así a fin de validar indirectamente los resultados obtenidos con las opiniones expertas. Es de esperar que algunas actividades adicionales sean *stimming* y no sean consideradas. Se espera que en el futuro, una vez establecido el nivel de confiabilidad del proceso completo, se pueda dar paso a la atención de casos no detectados manualmente y se evalúen con un protocolo adecuado.

A fin de conformar la base de datos, se realizaron los siguientes pasos:

A. Análisis de datos por especialistas

Los especialistas en TEA que colaboran con el proyecto de manera externa han detectado 88 instancias de interacciones de interés.

B. Tratamiento estadístico y heurístico de los casos manuales

Cada tira patrón-audio se etiqueta y se agrupa con técnicas de agrupamiento o *clustering*, usando para ello metadatos clásicos de sonidos tales como FFT, *wavelets*, espectro-poder, etc. Resultan 9 agrupaciones, siendo 8 de esas agrupaciones altamente cohesivas y la restante con 2 actividades que podrían tomarse como outliers o casos especiales.

Se determinan los elementos característicos (eigenvectores), mediante heurístico *Expectation Maximization*, de los 8 grupos cohesivos y se los considera como patrones o tiras modelo del *stimming* detectado. Dos casos adicionales son tratados cada uno como patrones a pesar de que no toman cada uno una cantidad proporcional de casos respecto a los anteriores.

C. Determinación manual de estímulos asociados

Los 10 patrones obtenidos se relacionan manualmente a situaciones visuales y sonoras que puedan identificar estímulos.

D. Asociación experta de estímulos a patrones o tiras

Los estímulos son evaluados y se identifica cuáles se asocian sistemáticamente a todas las instancias de cada patrón. Se descartan aquellos estímulos casuales. De este proceso se genera una tabla exhaustiva.

E. Derivación de meta-datos relacionados a cada patrón

El análisis matemático de cada patrón genera información que lo describe de manera objetiva, permite compararlos y principalmente diferenciarlos. Además de una etiqueta única (IDU) se asocia un nombre nemotécnico (Label), el timestamp de primera instancia (FI) el timestamp de la instancia actual (Time) en el cuerpo del actual registro que está siendo analizado, el umbral de similitud por debajo del cual se considera que el metadato de la instancia actual es similar al patrón (MetaDtThreshold), y otros descriptores adicionales. Por ejemplo la diferencia entre los valores de coeficientes WV del patrón y de la instancia actual,

la diferencia de coeficientes FFT, coeficientes de correlación, covariancias, etc. Todos estos resultados han sido publicados previamente en (López De Luise, D., 2020).

F. Reducción de metadatos

Para los fines de este paper, la muestra debe reducir la cantidad de variables requeridas manteniendo la capacidad de generar metadatos y el modelo de asociación con la mínima cantidad de metadatos posible. Eso se hace necesario para que el módulo pueda funcionar en tiempo real respecto a la captura de sonidos de una interacción con el paciente, lo que se relaciona a la celda denotada en la figura 1 como MLW.

La Tabla 2 muestra el total de metadatos considerados a los fines de este paper, que describen la relación matemática y estadística entre un patrón de sonido y una porción determinada de audio registrada.

Feature	Description
AVG-DifX	Promedio de diferencia en X entre Patrón y audio
AVG-DifY	Promedio de diferencia en Y entre Patrón y audio
AvgCovX	Promedio covariancia en X entre Patrón y audio
AvgCovY	Promedio covariancia en Y entre Patrón y audio
minX, minY	Mínimo de diferencia en X e Y
maxX, maxY	Máximo de diferencia en X e Y
MaxDifWalsh-HadamardX	Máximo de diferencia coeficientes Walsh-Hadamard sobre X
MaxDifWalsh-HadamardY	Máximo de diferencia coeficientes Walsh-Hadamard sobre Y
difCoefFFT-1 ... 30	Diferencia entre coeficientes FFT1 .. FFT30 entre Patrón y audio
R-CoeffFFT	Coficiente R de covariancia FFT versus audio

Tabla 2. Metadatos considerados.

Como se puede apreciar, los metadatos se centran en las diferencias entre un determinado patrón y una porción de audio (en el caso de uso se trata del Patrón 4, seleccionado al azar). Las características de las señales de cada uno de estos patrones fueron explicadas en publicaciones anteriores (ver López De Luise, D., 2020) y carecen de relevancia para el trabajo actual. El estudio se realiza con un *Sliding Window* que comprara progresivamente cada porción de patrón contra una ventana del audio actual, desplazando dicha ventana hasta cubrir todo el registro en estudio.

Las comparaciones tienen un umbral de tolerancia de divergencia por encima del cual se considera que el patrón no es igual a la porción de audio en estudio. Estos umbrales se verifican por cada tipo de metadato y para el total de metadatos en su conjunto. El resultado de estas comparaciones constituye el *ScoreParcial* de cada tramo de audio comparado, y es

registrado como parte del proceso de *sliding window* junto con el identificador de patrón y *timestamp*.

Los metadatos se estudian con el selector de las variables de la plataforma WEKA. Se emplean los siguientes métodos:

1. Selector de atributos **CfsSubsetEval**, evaluando la importancia de cierto subconjunto según su habilidad predictiva de un patrón, manteniendo redundancia. Se conjuga con el método selector **BestFirst**, que busca en el espacio de atributos con un método *hill-climbing* tradicional con el agregado de *backtracking*). La clase es **ScoreParcial**. El atributo ganador es R-CoeffFFT.
2. Selector de atributos **CfsSubsetEval**, búsqueda por **Greedy Stepwise**, que busca *forward* (o *backward*) agregando variables progresivamente hasta que nuevas incorporaciones reducen la función objetivo. Se empleó con opción *forward* iniciando desde agrupación vacía de elementos. El atributo ganador nuevamente es R-CoeffFFT.
3. Se repite el selector anterior con **BestFirst**, dando el mismo resultado.
4. Dado que las variables seleccionadas para este trabajo son todas numéricas, se realiza un Análisis de Componentes Principales (**PCA** por sus siglas en inglés) para identificar las variables interdependientes principales. El resultado arroja que todas las variables salvo difCoefFFT-19 ... difCoefFFT-30 son relevantes para explicar el 95% de los comportamientos en datos. El mayor grado de explicación lo da R-CoeffFFT.
5. El conjunto de variables se evalúa ahora no por su contribución a datos sino por su poder de determinar la presencia de patrones. Se realiza con **Evaluator ClassifierAttributeEval** con el *classifier* ZEROR (equivale a la construcción de un árbol de inducción de único nodo). El método de búsqueda es **RANKER**, que evalúa contribuciones individuales solamente y el evaluador de distancias **ZeroR**, basado en la estrategia tipo *GainRatio*. La secuencia de variables en orden de mayor a menor relevancia son las de la Tabla 3.

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

Tabla 3. Metadatos con poder clasificador.

De lo anterior, se puede resumir que:

- ScoreParcial depende del coeficiente R de covarianza entre FFT de patrón y tramo actual del registro.
- El valor ScoreTotal representa el total acumulado de valores ScoreParcial para distintos tipos de metadatos, pero en este trabajo se aplican solo los metadatos primarios, por lo que ScoreParcial es igual a ScoreTotal y se puede prescindir de uno de los dos.
- R-CoeffFFT es una variable esencial en la detección de patrones
- PCA denota que 25 de los 38 atributos interactúan de manera significativa con mayor o menor grado de participación. Los atributos de coeficientes FFT 19 en adelante lo hacen en escaso grado.
- El *ranking* de atributos por su capacidad de determinación de la existencia o no de patrón en determinada parte del audio permite trabajar con: R-CoeffFFT, coeficientes Walsh Hadamard y los primeros 15 coeficientes FFT.

En resumen, es posible reducir a 25 la cantidad de atributos conjugando el criterio de aporte de información y capacidad de determinación de existencia de un patrón. El conjunto resultante se incluye en la Tabla 4.

Variable	Descripción
AVG-DifX, AVG-DifY	Diferencia promedio entre las componentes x e y de cada muestra de sonido
minX, minY	Valor mínimo de componente x
maxX, maxY	Valor máximo de componente y
AvgCovX, AvgCovY	Promedio de covariancia de cada componente contra su media: $[x - \text{media}(x), y - \text{media}(y)]$
MaxDifWalsh-Hadamard	Máximo de diferencias en coeficientes Walsh-Hadamard Transform (FWHT)
R-CoeffFT	Coefficiente R de covariancia FFT
DifCoeffFT1..DifCoeffFT15	Diferencia entre coeficientes FFT 1 a 15

Tabla 4. Variables seleccionadas.

G. Detección automática de timestamp para Patrón 4

El prototipo BIOTECH cuenta con un *script* de *Sliding Window* que tiene parametrizado el tamaño de su ventana al del patrón de trabajo. A tal fin se corre una función que calcula dicho parámetro para cada patrón. La Tabla 5 muestra para cada caso la duración en segundos y la cantidad de registros.

Patrón	Duración	Tamaño
1	2.28"	100665
2	2.88"	127008
3	1.64"	72441
4	1.39"	61152
5	2.77"	122304
6	3.05"	134535
7	2.65"	116660
8	1.94"	85613
9	2.15"	95021
10	3.41"	150528

Tabla 5. Características de cada patrón.

La clasificación de los registros se hace en tiempo real con una red neuronal Multi Layered Perceptron (MLP) de 3 capas, previa normalización de rangos. Dada la limitada capacidad de cómputo de los equipos, solo se consideraron de los 25 metadatos los siguientes: Timespamp, AVG-DifX, AVG-DifY, MaxDifWalsh-HadamardX, MaxDifWalsh-HadamardY, R-CoeffFT, AvgCovXe15, AvgCovYe15 y HIT (variable de clasificación obtenida por el proceso de *clustering* K-Means) que son los que estadísticamente han demostrado ser más

relevantes en estudios previos de los autores (López De Luise, D., 2020). Este último, dado que refleja la agrupación natural considerando la información vertida de los 25 campos, se considera suficiente para una detección rápida de patrones.

La arquitectura neuronal cuenta con 9 nodos de entrada, una capa oculta de 5 unidades y una de salida con 1 unidad. La función de transferencia elegida es la sigmoidea y la inicialización es al azar. Los coeficientes de retención y aprendizaje son 0.05 y 0.01 respectivamente. Las métricas de resultados son *Kappa Statistics*, y porcentaje de aciertos, calculados considerando estrategia *split information*, con 66.0% para entrenamiento y el resto por *testing*, lo que constituye una validación estándar en la materia (ni pesimista como el *cross-validation*, ni optimista como el *supplied-test-set*).

La red MLP arroja altos niveles de precisión: 98.62 %, 91.98 %, 99.14 % para los *Set* 1, 2 y 3 respectivamente, con 1048576 registros en los primeros dos casos y 1046199 para el último. En todos los casos el *Kappa statistics* (0.98) indica que el modelo neuronal es apto para la tarea de detección de patrones usando los metadatos seleccionados.

H. Validación de timestamp hallados para Patrón 4

A fin de validar los *timestamp* obtenidos automáticamente con el proceso anterior contra el procedimiento manual de los especialistas, es necesario realizar una adaptación de los valores obtenidos. El proceso automático, al contar con mucho mayor sensibilidad detecta la ocurrencia de patrones no en un único *timestamp* sino en una secuencia, que son percibidos por el experto como un único segundo.

Dado que la intención es que el proceso no tenga intervención manual salvo para correr módulos automáticos, se realiza un procedimiento de agrupamiento en base a dichas regiones. El heurístico seleccionado es Simple K-Means con 2 clases tomando ScoreTotal como valor de referencia de calidad.

La Tabla 6 muestra la comparación entre lo obtenido manualmente y el resultado automático, para el patrón 4 que ocupa este caso de estudio.

Test set	Estímulo	Proceso manual		Proceso automático	
		Min	timestamp	Min	timestamp
1	4.14'	4.16	608697	4.20	621804
1	5.00'	--	--	5.72	729992
2	7.22'	7.23	1053331	7.38	1053339
2	10.08'	10.10	1450412	10.10	1450412
2	14.05'	14.10	2021065	14.16	2021074
3	16.33'	16.35	236589	16.58	2365838
3	16.57'	--	--	17.01	2427655

Tabla 6. Timestamp manual y automático.

La tabla muestra para cada entrada manual un estímulo asociado. Cada patrón obedece a un estímulo diferente (López De Luise D., 2012a) y para el Patrón 4 se trata del canto de pájaros. Los *timestamps* han sido reducidos a su correspondiente momento en la grabación, mediante simple cálculo matemático basado en la velocidad del muestreo reducido.

Es de notar que el proceso automático determina sistemáticamente un *timestamp* superior, posiblemente debido a que el tiempo en reaccionar por parte del especialista durante el procedimiento manual, se ha descontado aproximadamente.

Adicionalmente se producen hallazgos adicionales a los registrados en la tabla manual. Esto lleva a rastrear el estímulo de canto de pájaros en dichas circunstancias.

Otro hallazgo interesante es que sistemáticamente los valores MaxDifWalsh-HadamardX, MaxDifWalsh-HadamardY y R-coeffFFT son los que más significativamente difieren en estos *timestamp*. Para el test set 1, los valores para la instancia en minuto 4.20' son 0.000655651, 0.000399113 y 0.320047 respectivamente, al menos 1 orden de magnitud de diferencia respecto al resto del data-set. Para el minuto 5.72' los valores son 0.00473976, 0.00031662 y 0.449992.

La Figura 2 muestra en rojo las regiones de *timestamp* para *test-set* 1, donde se produce el patrón 4.

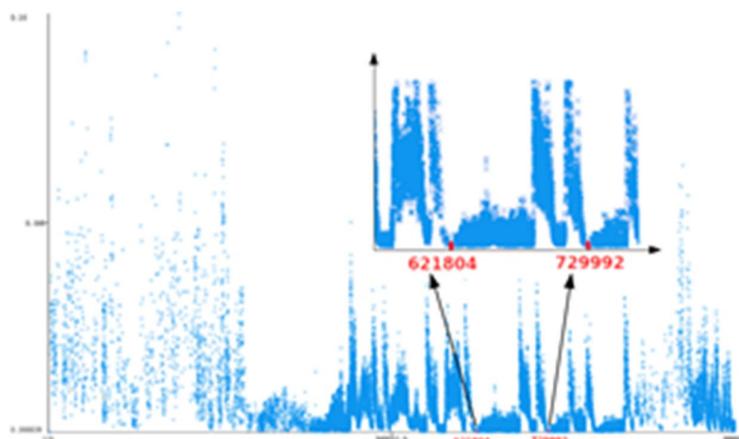


Figura 2. Timestamp vs. MaxDifWalsh-Hadamard

I. Análisis de los resultados

La Tabla 7 extiende a Tabla 6 con los resultados de MLP: el minuto hallado y el número de registro (que equivale en este contexto a *timestamp*).

Set	Estím.	Proceso manual		Proceso Automático		Modelo Neuronal	
		Min.	timestamp	Min.	timestamp	Min.	timestamp
1	4.14'	4.16	608697	4.20	621804	4.16	618208
1	5.00'	--	--	5.72	729992	--	--
2	7.22'	7.23	1053331	7.38	1088997	7.36	1084241
2	8.34'	--	--	--	--	8.35	1224527
2	10.08'	10.10	1450412	10.18	1452789	10.16	1449996
2	13.29'	--	--	--	--	13.52	1930711
2	14.05'	14.10	2021065	14.16	2021074	14.10	2021065
3	16.33'	16.35	236589	16.58	2365838	16.35	2365834
3	16.57'	--	--	17.01	2427655	17.01	2427668

Tabla 7. Timestamp manual y automático.

El proceso neuronal, confirma los hallazgos automáticos y agrega otros dos en los minutos 8.34' y 13.29'. Se observa que los *timestamp* no son exactamente iguales entre el proceso automático y el modelo neuronal. Esto sucede porque en el primer caso se trata de una aproximación con el valor central en las secuencias de concordancia patrón-audio, mientras que la red neuronal se basa en su aprendizaje en base a los metadatos. Por lo anterior, es de suponer que la MLP sea más precisa.

Conclusiones

Este trabajo introdujo algunas de las características de los pacientes con TEA, y su relación con el problema comunicacional. El proyecto BIOTECH, diseña y desarrolla una serie de metadatos basados de procesamientos matemáticos y heurísticos de ciertas tiras de sonido correspondientes a patrones detectados previamente por especialistas en el tema. Se ha evaluado uno de los patrones y seguido el proceso desde el registro grabado hasta la obtención de los momentos donde el patrón se verifica durante un audio correspondiente a una sesión con el paciente seleccionado para el caso de estudio. De lo presentado, considerando el sesgo heurístico y por el excelente valor de la métrica de validación *kappa statistics*, se puede inferir que los metadatos más eficaces para una detección rápida de patrones en un audio son el coeficiente de correlación entre componentes FFT, y la máxima diferencia entre componentes Walsh Hadamard. Los metadatos fueron procesados con una red neuronal MLP con parámetros ajustados para poder detectar rápidamente en una situación real.

Los resultados de este trabajo también confirman los hallazgos anteriores del grupo, en cuanto a que existen patrones de sonido con los que el paciente responde selectivamente ante ciertos estímulos. También muestra las bondades del proceso automático en cuanto a precisión y repetitividad, así como mayor potencia respecto al proceso manual para hallar momentos en los que el paciente se comunica. Es de remarcar que el sobreajuste no existe ya que se evalúa por el valor de Kappa y que no existe sesgo por confusión ya que se considera la matriz de confusión como parte del estudio de convergencia de la MLP.

Entre otras actividades resta aún la integración de los módulos en un prototipo unificado con una interfaz simple para un usuario no técnico, el hallazgo automático de los patrones, extendiendo los módulos de detección actuales. También resta la realización y validación de procesamientos análogos para registros visuales.

En el proceso de integración de ambas vertientes, visual y auditiva, es de esperar que algunas actividades adicionales sean *stimming* y no sean consideradas. Se espera que en el futuro, una vez establecido el nivel de confiabilidad del proceso completo, se pueda dar paso a la atención de casos no detectados y se evalúen con un protocolo adecuado. Otros detalles que se están trabajando actualmente incluyen: cambiar el parámetro tamaño de ventana de *Sliding-Window* a cálculo automático a partir del patrón, el *subsampling* automático y no manual, y que los patrones se detecten automáticamente: teniendo un rango de menor a mayor tamaño, luego

del caso de uso, es posible determinar cuáles son las distintas ventanas posibles para realizar dicha tarea.

Bibliografía

- Ahmed, N. y Rao, K. R. (1975). *Walsh-Hadamard Transform*. En: *Orthogonal Transforms for Digital Signal Processing*. Berlin: Springer.
- American Psychiatric Association (2014). *DSM-V Encoding update®. Supplement to the Diagnostic and Statistical Manual of Mental Disorders*.
- Beazley, D., Jones, B. y Jones, K. (2013) *Python Cookbook*. Ciudad: O'Reilly Media, Inc.
- Bustamante, P., Lafalla, A., Coria, N., Agüero M., López de Luise, D., Parra C, Azor R, Moya, J. y Cuesta, J. (2016) *Protocol for evaluation of somatic and oral behavior to sound stimuli in patients with autism spectrum. CI2S Lab. Pringles 10- P2*. University of Burgos.
- Cheol-Hong, N., Ahmed H. (2010). *Novel pattern detection in children with autism spectrum disorder using iterative subspace identification*. Department of Electrical and Computer Engineering, University of Minnesota.
- Eaton, J. W., Bateman, D., Hauberg, S. y Wehbring, R. (2017). *Octave Manual. Edition 4 for Octave version 4.2.1*. Free Software Foundation, Inc.
- Elsabbagh, M., Divan G., Koh, Y., Kim, Y.S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M.T. y Fombonne, E. (2012). “Global prevalence of autism and other pervasive developmental disorders”. *Autism Research*, 5 (3): 160–179.
- Filipek, P. A., Accardo, P. J., y Ashwal, S. (2000). “Parameter practice: Detection and autistic diagnose”. *Neurology*, 55: 468-479.
- Fortea Sevilla, M. S., Escandell, M. O., Bermúdez, J. y Castro Sánchez, J. (2013). “How Many People have Autism? A Theoretical Review”. *International Journal of Developmental and Educational Psychology*, 1 (1): 769-786.
- Greer, R. D. (1997). “The comprehensive application of behavior analysis to schooling. (CABAS®)”. *Behavior and Social Issues*, 7 (1).
- _____ (2008). “The ontogenetic selection of verbal capabilities: contributions of skinner’s verbal behavior theory to a more comprehensive understanding of language”. *International Journal of Psychology. Psychology Therapy*, 8 (3): 363-386.
- Kuiper, M. W., Verhoeven, E. W., Geurts, H. M. (2019). “Stop Making Noise! Auditory Sensitivity in Adults with an Autism Spectrum Disorder Diagnosis: Physiological Habituation and Subjective Detection Thresholds”. *Journal Autism Development Disorder*, 49 (5): 2116–2128.
- Kanner, L. (1943). “Autistic disturbances of affective contact”. *Nervous Child*, 2: 217-250.
- López De Luise, D. (2012). “Morphosyntactic linguistic wavelets for knowledge management”. *TechOpen*, 8: 167–189.
- López De Luise, D., Hisgen, D., Cabrera, A. y Morales Rins, M. (2012). “Modeling dialogs with linguistic Wavelets”. *Theory and Practice in Modern Computing*, 1, 11-13.
- López De Luise, D., Hisgen, D. (2013). “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*: 568–600.
- López De Luise, D., Azor, R., Párraga, C. (2015). “Autistic verbal behavior. Automated model of the patient profile – audio”. Editorial Académica Española.
- López De Luise, D. (2020). *Autistic Verbal Behavior Parameters (paper)*. 9th International Workshop On Soft Computing Applications (SOFA 2020). Arad-Romania, University of Arad.

- Marsh, K. L., Isenhower, R. W., Richardson, M. J., Helt, M., Verbalis, A. D., Schmidt, D. C. y Fein D. (2013). "Autism and social disconnection in interpersonal rocking". *Frontiers in Integrative Neuroscience*, 7: 4.
- Mazefsky, C. A., Herrington, J., Siegel, M., Scarpa, A., Maddox, B. B., Scahill, L. y White, S. W. (2013). "The role of emotion regulation in autism spectrum disorder". *Journal of the American Academy of Child and Adolescent Psychiatry*, 52 (7): 679-688.
- Mazefsky, C. A. y White, S. W. (2014). "Emotion regulation: concepts & practice in autism spectrum disorder". *Child and Adolescent Psychiatric Clinics of North America*, 23 (1): 15-24.
- Ministerio de Ciencia e Innovación de España (2009). *Clinical practice guide for the management of patients with autism spectrum disorders in primary care*. Madrid: Estilo Estugraf Impresores.
- Mower, E., Black, M. P., Flores, E., Williams, M. y Narayanan, S. (2011). *Design of an emotionally targeted interactive agent for children with autism*. California: University of Southern California.
- Riviere, A. M. (2000). *The little boy with autism*. Free Software Foundation, Inc.
- Skinner, B. F. (1981). *Verbal Conduct*. México: Trillas.
- Spilisbury, M. J. y Euceda, A. (2019). "Transformada Rápida de Fourier". *Revista de la Escuela de Física*, 4 (2): 45-52.
- Summers, J., Shahrami, A., Cali, S., D'Mello, C., Kako, M., Palikucin-Reljin, A., Savage, M., Shaw, O. y Lunskey, Y. (2017). "Self-Injury in Autism Spectrum Disorder and Intellectual Disability: Exploring the Role of Reactivity to Pain and Sensory Input". *Brain Science*, 7 (11): 140.
- Sundberg, M. L. (2007). "A brief overview of a behavioral approach to language assessment and intervention for children with autism". *Association for Behavior Analysis Newsletter*, 30 (3).
- van Santen, J. P., Sproat, R. W. y Hill, A. P. (2013). "Quantifying repetitive speech in autism spectrum disorders and language impairment". *Autism research: Official Journal of the International Society for Autism Research*. 6 (5): 372-383.
- Wing, L. (1993). "The definition and prevalence of autism: a review". *European Child and Adolescent Psychiatry*, 2: 61-74.