

Uso de la IA para proporcionar retroalimentación al docente a través del análisis de las grabaciones de clases

Federico Pardo García, Óscar Cánovas, Felix J. García Clemente y Pilar González-Férez

Departamento de Ingeniería y Tecnología de Computadores

Universidad de Murcia

30100 Murcia

{federico.pardog, ocanovas, fgarcia, pilargf}*@um.es

Resumen

Este trabajo presenta el despliegue práctico de una plataforma de análisis del habla basada en inteligencia artificial en el contexto de la actividad docente. Utilizando grabaciones de clases, el objetivo central es desarrollar elementos de información significativos para los docentes, proporcionándoles retroalimentación detallada sobre el desarrollo de sus clases. El análisis de audio aborda aspectos como el momento de intervención tanto del docente como de los estudiantes, los patrones de discurso del docente, la duración de las intervenciones y las pausas, además de la identificación de los métodos docentes empleados. Esta información se presenta a través de informes diseñados en colaboración con los docentes, con el objetivo de resaltar información clave, disponerla de manera adecuada, emplear elementos de diseño apropiados y facilitar el análisis de la evolución temporal. En este artículo, se detallan las características principales de los informes de retroalimentación, su implementación en contextos docentes reales y las perspectivas que ofrecen para futuros escenarios de análisis docente. Además, se incluyen las principales valoraciones cualitativas realizadas por un conjunto de docentes que participaron en el piloto.

Abstract

This work presents the practical deployment of an artificial intelligence-based speech analysis platform within the context of educational activities. Utilizing classroom recordings, the primary objective is to generate meaningful informational components for educators, furnishing them with detailed feedback on the progression of their classes. The audio analysis encompasses factors such as the timing of intervention from both the educator and students, speech patterns of the educator, the duration of interventions and pauses, along with the identification of teaching methods. This information is conveyed through reports collabo-

ratively designed with educators, with the intention of emphasizing crucial information, arranging it appropriately, employing fitting design elements, and facilitating the analysis of temporal evolution. This article delineates the principal features of feedback reports, their implementation in authentic educational settings, and the prospects they offer for future educational analysis scenarios. Additionally, it incorporates the primary qualitative assessments conducted by a group of educators who participated in the pilot study.

Palabras clave

Análisis del habla, inteligencia artificial, retroalimentación docente.

1. Introducción

Para una enseñanza significativa, es aconsejable que los docentes se comprometan en un aprendizaje continuo, analizando sus prácticas de enseñanza y adaptando sus métodos pedagógicos para mejorar su metodología. En este sentido, resulta fundamental que el profesorado reciba retroalimentación oportuna e inmediata sobre sus sesiones de enseñanza. Sin embargo, la práctica deliberada, guiada por formadores y acompañada de retroalimentación constructiva, a menudo es un proceso que demanda una considerable cantidad de tiempo y puede resultar poco eficiente [1].

Conscientes de la importancia de la retroalimentación en el aprendizaje del docente, creemos que es clave proporcionar una comprensión clara y profunda de su propio desempeño y progreso. El objetivo de este trabajo es abordar la falta de retroalimentación inmediata y objetiva para los docentes. El estudio realizado por Wang *et.al.* [17] sugiere que cuando a los docentes se les proporciona retroalimentación automatizada sobre la proporción de diálogo entre el docente y el estudiante, hay un aumento significativo en las intervenciones de los estudiantes. Incluso información básica

sobre los patrones de discurso en el aula puede llevar a cambios positivos en la dirección deseada.

Este artículo expone una experiencia de uso de una plataforma software de análisis de audio basada en inteligencia artificial para generar informes visuales que proporcionan información adecuada al docente de forma sencilla y comprensible sobre sus clases. Dichos informes han sido diseñados en colaboración con docentes y se han entregado periódicamente, cada semana, a los docentes de 3 materias distintas. Al finalizar la fase piloto se ha solicitado a los participantes que realicen una valoración cualitativa que se incluye como parte de este trabajo.

La plataforma software realiza un procesamiento del audio que se caracteriza por llevar a cabo la extracción de características derivadas del proceso de diarización [13]. La diarización permite la partición de las grabaciones en segmentos específicos para cada hablante, lo que permite un análisis más profundo de señales no verbales como la duración del turno del hablante o la superposición de voces. Con 12 características no verbales (descritas en la Sección 4.3), obtenemos información valiosa sobre la dinámica del discurso del docente. Las señales no verbales proporcionan una fuente rica de información sobre la interacción y dinámica dentro del aula. Además, nuestra propuesta hace uso de un sistema de clasificación automatizado de los métodos docentes en base a las características extraídas en el proceso de diarización. Específicamente, el sistema es capaz de identificar tres métodos docentes principales: clase magistral, trabajo en grupo de estudiantes y el uso de un sistema de respuesta para la audiencia.

Los informes proporcionan retroalimentación visual al docente, para que la información sea comprensible de un vistazo. Estos informes aportan una visión general del comportamiento tanto del docente como de los estudiantes durante la sesión. Mediante el uso de gráficos se puede comprender, por ejemplo, el equilibrio en las intervenciones de cada parte en momentos específicos de la docencia. Veremos más adelante que estos gráficos han permitido a los docentes obtener conclusiones interesantes acerca de su estilo de enseñanza. Creemos, por tanto, que estos informes realizados mediante nuestra plataforma software de análisis de audio suponen una contribución relevante en el ámbito de las analíticas del proceso de enseñanza-aprendizaje.

2. Trabajos previos

En los últimos años, se han realizado varios estudios para analizar el ambiente en el aula [9]. Aunque la práctica de grabar sesiones de clase no es nueva, la transición hacia el análisis automático de estas grabaciones ha ganado impulso recientemente [6]. La mayoría de las propuestas recientes aplican técnicas de

aprendizaje automático o redes neuronales para examinar diferentes prácticas y estilos de enseñanza. El análisis de estas grabaciones puede implicar la extracción de características no verbales o la aplicación de técnicas de procesamiento del lenguaje natural.

En relación con el discurso en el aula, varios equipos de investigación han dedicado sus esfuerzos al desarrollo y validación de sistemas automatizados destinados a clasificar prácticas docentes. Por ejemplo, Donnelly et al. [8] entrenaron modelos de aprendizaje automático supervisado para clasificar grabaciones. Otros trabajos basados en el reconocimiento automático del habla se han centrado en segmentar el discurso en el aula entre docentes y estudiantes [7] y aprovechar características acústicas de bajo nivel. La mayoría de las propuestas prestan una atención particular al papel del profesor [14] para clasificar tareas de aprendizaje activo [11, 12, 15].

Sin embargo, es importante señalar que estos trabajos se centran predominantemente en lograr una alta precisión de clasificación, a menudo pasando por alto la definición de características del discurso que pueden servir como datos descriptivos e informativos [12, 17]. Analizar diferentes estilos de enseñanza y ofrecer recomendaciones valiosas a los docentes requiere no solo una clasificación precisa, sino también una comprensión integral de la dinámica subyacente del discurso. Los datos informativos son cruciales para capturar los matices de las prácticas docentes y proporcionar ideas significativas. Creemos que el enfoque presentado en este artículo aborda adecuadamente este fin, puesto que utiliza características informativas para realizar no solo la clasificación de prácticas docentes, sino también un análisis de estilos de enseñanza dentro de cada método.

Actualmente, existen productos comerciales disponibles, como la start-up TeachFX¹, que se ha centrado principalmente en cuantificar la proporción de habla del profesor y del estudiante, al proporcionar gráficos relacionados con la participación e interacción. Sin embargo, su enfoque es limitado en términos del número de características analizadas, lo que impide un análisis temporal detallado y completo.

Además del enfoque no verbal, los métodos tradicionales de análisis automatizado del discurso del profesor han confiado en transcripciones de reconocimiento automático del habla (ASR, por sus siglas en inglés). Este método ampliamente adoptado implica la extracción de características lingüísticas de alto nivel que abarcan palabras, oraciones y niveles de discurso. En algunos estudios [16], se han empleado técnicas de aprendizaje profundo, por ejemplo, para identificar estrategias dialógicas particulares dentro de las aulas de matemáticas. Aunque actualmente estamos trabajando en un sistema híbrido para el análisis del aula basado

¹<https://teachfx.com>

en características no verbales y técnicas de Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés), este último enfoque está fuera del alcance de los informes presentados en este artículo.

3. Contexto de la experiencia

3.1. Materias y participantes

Este trabajo presenta la experiencia docente desarrollada durante el curso académico 2023/24 en tres asignaturas del Grado en Ingeniería Informática de la Facultad de Informática de la Universidad de Murcia. Una de las asignaturas es Fundamentos de Computadores que se cursa en primero, otra es Redes de Computadores que se cursa en segundo, y la tercera asignatura se cursa en tercero y es Ampliación de Sistemas Operativos. Las tres asignaturas son de 6 créditos ECTS repartidos en 3 créditos de teoría y otros 3 de prácticas.

Respecto a la asignatura Fundamentos de Computadores, en el grupo que se ha realizado el estudio hay un total de 88 estudiantes matriculados, de los cuales han asistido a clase regularmente alrededor de 70 estudiantes. El grupo que ha sido objeto de estudio para la asignatura de Redes de computadores está formado por 75 estudiantes matriculados, y han acudido regularmente a clase unos 65 estudiantes. En la asignatura Ampliación de Sistemas Operativos, el grupo que ha sido objeto de estudio tiene un total de 78 matriculados, aunque regularmente a clase han asistido unos 60 estudiantes. El estudio se ha centrado exclusivamente en la grabación de las clases de teoría, con el fin de tener un entorno lo más homogéneo posible.

En las tres asignaturas se usa una metodología similar en la parte de teoría. Los conceptos se introducen usando clases magistrales. Para afianzar dichos conceptos, en muchas de estas clases el docente resuelve ejercicios o problemas en pizarra, fomentando en todo momento la participación del estudiante. Para todos los temas, se utiliza la herramienta Wooclap con cuestiones específicas del tema correspondiente, lo que permite resolver dudas y comprobar si los estudiantes han entendido y asimilado los conceptos. Por último, en algunas de las sesiones teóricas del curso se realizan seminarios en los que los estudiantes trabajan en grupo los ejercicios propuestos para la sesión por el docente.

3.2. Procedimiento de obtención de datos

Antes de proceder a la grabación de las clases, los docentes informaron a todos los estudiantes acerca del procedimiento (aprobado por la institución y el comité ético en el marco de un proyecto de innovación docen-

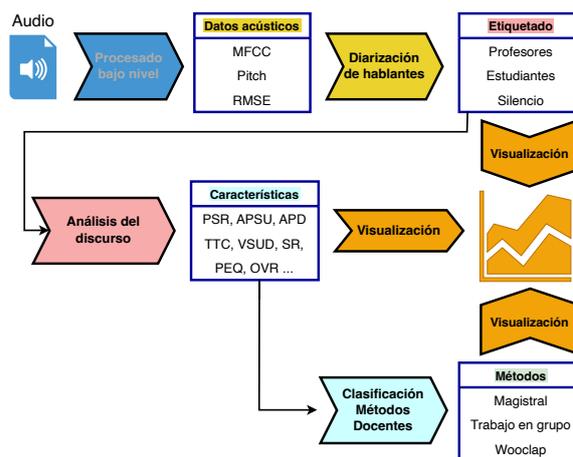


Figura 1: Secuencia de etapas de procesamiento y elementos de datos

te), la finalidad del estudio y el compromiso de privacidad.

Las grabaciones se realizaron con una grabadora digital de mano, modelo TASCAM DR-07X, situada a una distancia mínima de 1.5 metros del docente y de los estudiantes de la primera fila. Posteriormente se procesaron manualmente los audios para seleccionar únicamente los fragmentos que abarcaban la clase, eliminando un posible descanso realizado a mitad de la clase, o algunos minutos finales debidos a no parar la grabación justo al terminar la clase. Posteriormente dichas grabaciones fueron procesadas por el sistema de análisis de audios [4] para generar la diarización de las mismas y la extracción de características, datos que se explican y analizan en detalle en las siguientes secciones.

4. Sistema de análisis de audio basado en IA

Nuestra plataforma de análisis de audio, de desarrollo propio, consta de un pipeline de cinco etapas, como se muestra en la Figura 1. Cada etapa proporciona información de entrada a la siguiente para generar información de más alto nivel. La etapa de “Visualización” ofrece, entre otros elementos, los informes analizados específicamente en la Sección 5.

4.1. Procesamiento de bajo nivel

La etapa inicial del procesamiento de audio implica la extracción de características de bajo nivel de las grabaciones de audio. Algunas características importantes incluyen los Coeficientes Cepstrales de Frecuencia Mel (MFCC), el tono y la energía promedio para cada

muestra de voz. Estas características se emplean ampliamente en la literatura científica como entradas para técnicas de diarización de hablantes. En nuestro desarrollo este proceso se realiza mediante la librería librosa² en Python.

4.2. Diarización de hablantes

La diarización de hablantes es el proceso de dividir las grabaciones de audio en segmentos y asignar etiquetas de hablantes para determinar “Quién Habla Cuándo”. Un sistema de diarización consta de un modelo de Detección de Actividad Vocal (VAD) que identifica intervalos de tiempo en el audio donde hay habla, ignorando el ruido de fondo. También incluye un modelo de codificación de los hablantes que extrae características de audio únicas del hablante de los segmentos de habla identificados por el VAD. En nuestro enfoque, podemos diferenciar los discursos de los docentes. Sin embargo, no distinguimos entre estudiantes individuales, por lo que etiquetamos al resto de hablantes como “Estudiantes”. Como herramienta de diarización utilizamos la librería Pyannote-audio³, basada en una red neuronal end-to-end [3].

4.3. Cálculo de características

Tomando en consideración varios trabajos previos [2, 10], hemos definido varias características no verbales extraídas de la diarización que son útiles para la caracterización de discursos. Específicamente, definimos cuatro características de discurso por rol (profesor y estudiantes):

- Participant Speaking Ratio (PSR). La proporción de participación de cada rol durante el segmento de grabación.
- Participant Speaking Utterances (PSU). El número de intervenciones del rol en el segmento.
- Participant Speaking Utterances Ratio (PSUR). La proporción de participaciones de cada rol durante el segmento de grabación.
- Average Participant Speaking Utterance Duration (APSUD). Duración promedio de las intervenciones de cada rol en el segmento.

Adicionalmente, también hemos definido ocho características globales del discurso:

- Average Lapse Duration (ALD). La duración promedio de los periodos de silencio.
- Silence Ratio (SR). La proporción de periodos de silencio durante el segmento de grabación.

- Average Pause Duration (APD). La duración promedio del intervalo de silencio entre participaciones del mismo rol.
- Participation Equality (PEQ). Un indicador que evalúa el equilibrio de participación entre diferentes roles. Se calcula siguiendo la metodología descrita en [10]. Valores cercanos a 1 indican una distribución equitativa de la participación.
- Turn Taking Count (TTC). El número de cambios de turno ocurridos en el diálogo entre los estudiantes y el profesor.
- Very Short Utterances Ratio (VSUR). La proporción de intervenciones de duración muy corta (menos de 2 segundos) sobre el total.
- Overlapping Rate (OVR). La tasa de intervenciones superpuestas de diferentes participantes.
- Overlapping Utterances Rate (OVUR). La proporción de intervenciones que están superpuestas.

Como se verá, estas características proporcionan gran detalle de información acerca del transcurso de las clases, y pueden ser representadas tanto de forma global, es decir, con valores promedio para el conjunto de toda la grabación, como mediante secuencia de valores ordenados temporalmente utilizando ventanas de tiempo.

4.4. Clasificación de métodos docentes

Una vez que ya se han obtenido las características del discurso anteriormente descritas, el siguiente paso consiste en la utilización de dicha información a la hora de clasificar el método de enseñanza empleando por el docente en cada instante. Dicha clasificación permitirá enriquecer los informes proporcionados mediante la identificación de las distintas metodologías empleadas en cada instante de la grabación.

Para ello realizamos una tarea de clasificación multi-clase que involucra tres clases: Lección Magistral, Trabajo en Grupo y Wooclap. Nuestro enfoque utilizó algoritmos de aprendizaje automático supervisado para predecir la etiqueta adecuada para un segmento de grabación de entrada. Exploramos una variedad de modelos, incluidas Máquinas de Soporte Vectorial (SVM), k-Nearest Neighbor (kNN), Random Forest, Naive Bayes, Regresión Logística y Gradient Boosting, para identificar el algoritmo más efectivo para nuestra tarea [5]. Como se indica en el trabajo citado, entre los modelos evaluados, el modelo SVM exhibió el mejor rendimiento, logrando un F1 cercano al 94 % y AUC (Area Under the Curve) de 0.98 en la validación cruzada y en la clasificación de audios de test. Esto indica la robustez y efectividad del modelo SVM en clasificar con precisión los métodos docentes.

Dado que un profesor puede utilizar varios métodos de enseñanza en la misma clase, nuestra etapa de cla-

²<https://librosa.org/>

³<https://github.com/pyannote/pyannote-audio>

sificación tiene como objetivo identificar los puntos de inicio y fin de cada método. Para lograr esto, el proceso de clasificación implica analizar pequeños segmentos de las grabaciones. Hemos examinado tamaños de ventana superpuestos que van desde 60 segundos hasta 300 segundos. Nuestros resultados preliminares muestran que el mejor equilibrio entre granularidad y precisión se obtiene utilizando ventanas de 120 o 180 segundos.

4.5. Visualización de datos

El último paso es realizar la visualización de los datos. Es importante una visualización efectiva de datos para proporcionar a los profesores información valiosa sobre su actividad docente y dotarlos de analíticas informativas para mejorar sus prácticas de enseñanza. La Sección 5 ahonda en los detalles de visualización.

5. Informes generados

Cada informe refleja la dinámica de una clase. El objetivo es proporcionar al docente información detallada sobre la clase, como por ejemplo sus intervenciones, las de los estudiantes o los momentos de silencio, así como representar los intervalos de los distintos métodos docentes detectados. Las Figuras 2 y 3 muestran un ejemplo del contenido de un informe elaborado para una grabación real concreta.

La cabecera del informe incluye el nombre del docente, la asignatura a la que pertenece el informe, las metodologías utilizadas durante la clase y la fecha de la clase. A la hora de subir la grabación del audio de la clase, el profesor introduce toda esta información con el propósito de generar los metadatos que facilitarán la organización y la generación de los informes.

La primera página del informe muestra las características PSR, APSUD y PSU del discurso por rol (teacher y student). Se decidió adoptar la decisión de diseño de presentar la información global de cada característica extraída del conjunto de la grabación en la mitad izquierda del informe, con un valor medio por rol en un diagrama de columnas. A su derecha, los diagramas de líneas representan la evolución temporal de la característica correspondiente durante la clase, con una granularidad de un minuto. Dentro de cada característica, la gráfica superior representa la evolución de la característica para el docente y la inferior la evolución para los estudiantes. La disposición vertical y alineada de estas gráficas permite comparar fácilmente el valor de la característica en un punto concreto para ambos roles. Como detalle adicional, se ha coloreado el fondo de las gráficas en tres tonos de grises para representar el método docente que se ha detectado en base a

las características en esa ventana de tiempo. Los métodos docentes que se predicen utilizando el sistema de clasificación SVM mencionado anteriormente son, de más claro a más oscuro: “Lecture” (lección magistral), “Wooclap” y “Group Work” (trabajo en grupo). Esto puede ayudar al docente a identificar momentos claves de la clase de una forma más sencilla.

La segunda página del informe incluye la representación de las características restantes. En concreto, un diagrama de barras con los valores medios de PSUR y una representación temporal de la característica ALD. La distribución de los silencios aporta información valiosa sobre el desarrollo de la actividad docente en diferentes puntos de la misma. Las características VSUR, PEQ y SR se representan en un diagrama radial, en un rango de 0 a 1. Hemos considerado representar estas tres variables de forma conjunta puesto que están muy relacionadas entre sí. Si el diagrama está más cerca de 1 en las tres características, la actividad docente ha estado repartida entre los alumnos y el docente y ambos han intervenido de forma continua y equitativa. Finalmente, la característica TTC, representada en una línea temporal, permite visualizar como se reparten los turnos de intervención entre estudiantes y docentes a lo largo del tiempo.

Analizando específicamente el ejemplo de informe mostrado en la Figura 2, podemos visualizar el progreso de una clase. A primera vista, se ve que la clase se divide en una primera sección de trabajo en grupo, seguido de un periodo de lección magistral y finaliza con un Wooclap. Hasta el minuto 00:48 el sistema detecta que se está realizando un trabajo en grupo. A continuación, el profesor hace la corrección y explicación del ejercicio, lo que el sistema detecta como una clase magistral. Finalmente, los últimos 20 minutos de clase son un Wooclap de repaso de los contenidos. Se observan dos imprecisiones en el sistema de detección de métodos docentes debido a que el sistema tiene un porcentaje de acierto del 95 %, en dos ventanas temporales que pueden identificarse fácilmente como errores de predicción difíciles de subsanar porque en ocasiones los matices para distinguir entre metodologías son muy sutiles.

6. Valoraciones de los informes

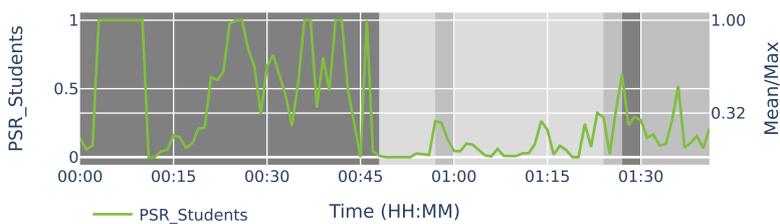
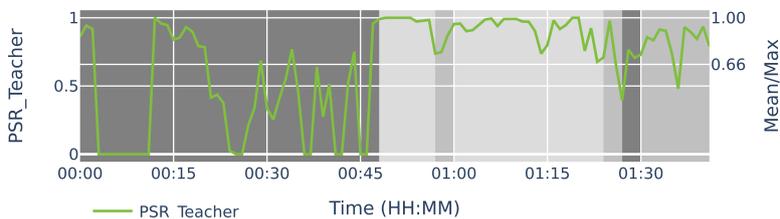
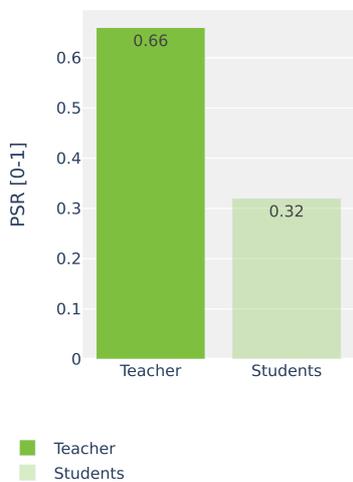
Para verificar la eficacia de estos informes en un ámbito real, se ha pedido a tres docentes implicados en el proyecto piloto que den su opinión sobre los informes generados en sus clases. Para ello, se les ha solicitado una evaluación cualitativa de los informes articulada a través de las siguientes cinco preguntas:

1. ¿Consideras que las características analizadas reflejan de manera precisa tu desempeño en el aula?

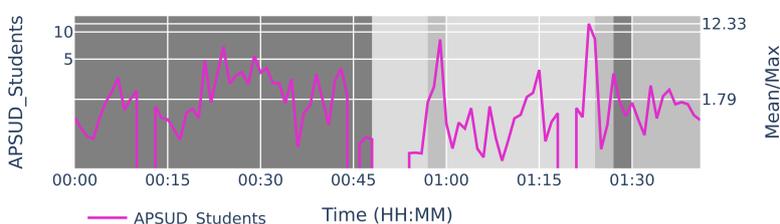
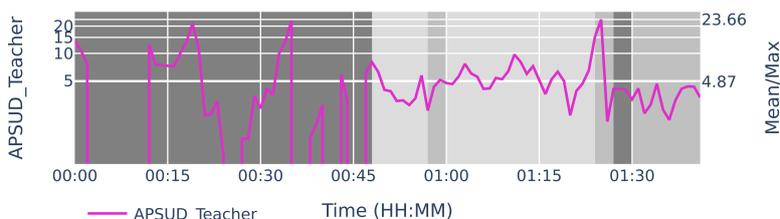
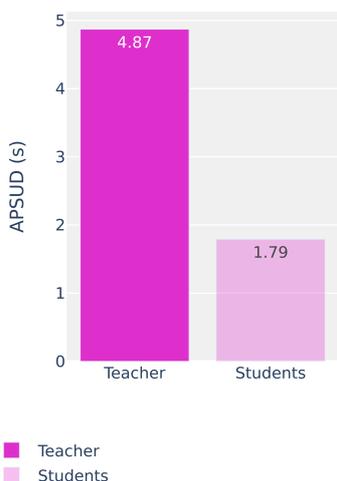
Classroom Interaction Report

PSR: Participation's ratio for each role.

Lecture Wooclap Group Work



APSUD: Utterances's average duration for each role in seconds.



PSU: Number of utterances in the current recording segment.

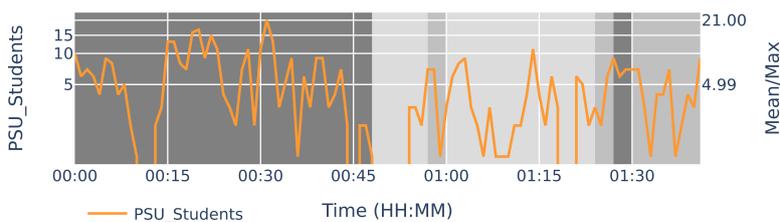
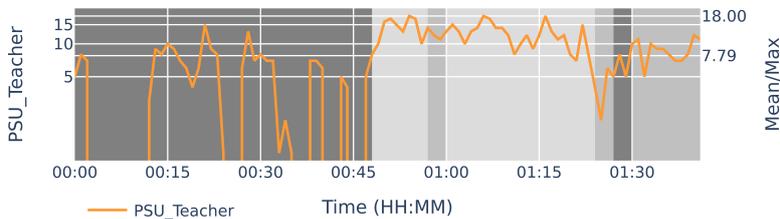
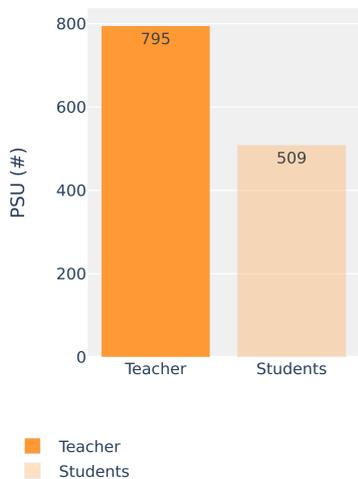
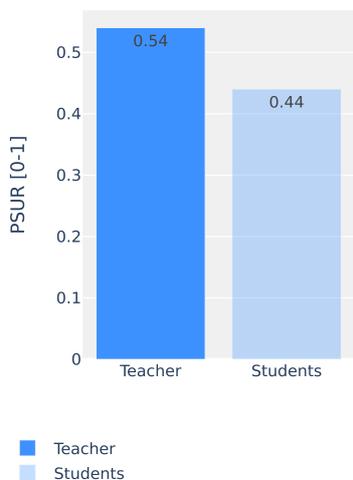
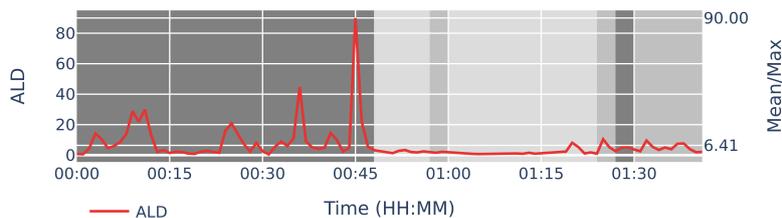


Figura 2: Primera página del informe generado para el docente.

PSUR: Utterances's ratio for each role.



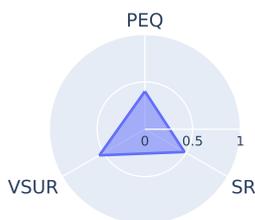
ALD: 6.41, Silences's average duration in seconds.



VSUR: The ratio of very short speech utterances (less than 2 seconds) over the total.

PEQ: An indicator that assesses the balance of participation among different roles. Values close to 1 indicate an equal distribution of participation.

SR: The ratio of participation of each role during the recording segment.



TTC: 469.0, Number of turns changes.

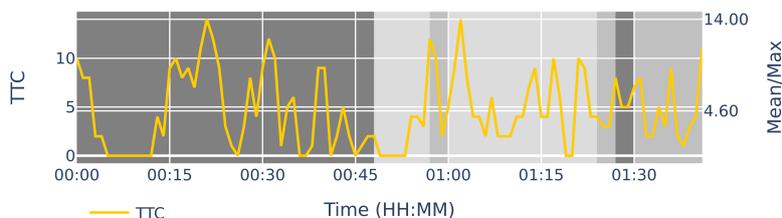


Figura 3: Segunda página del informe generado para el docente.

2. ¿Qué opinas sobre la clasificación automática del tipo de actividad docente?
3. ¿Hay información específica que consideres crucial y no esté incluida en los informes actuales?
4. ¿Cómo evalúas la utilidad general de estos informes en términos de proporcionar feedback sobre tu actividad docente?
5. ¿En qué medida crees que estos informes podrían contribuir a tu desarrollo profesional y a la mejora de tus prácticas pedagógicas?

Los docentes evaluados coinciden en la precisión y utilidad de los informes generados para reflejar las dinámicas del aula, destacando la adecuada clasificación de actividades pedagógicas y la inclusión de métricas específicas. Valoran la capacidad de estos informes para facilitar un análisis detallado de la distribución del tiempo entre los diferentes métodos docentes, la planificación de clases, y la interacción estudiantil. En particular destacan la utilidad del uso de las métricas de participación de estudiantes frente a las suyas, así como la longitud de las intervenciones de los estudiantes, apoyados en el contexto del método docente que se esté utilizando en ese momento. Todos coinciden en que

pueden ser de interés para detectar comportamientos anómalos. Algunos docentes también consideran que pueden ser útiles para identificar si han interactuado suficiente con los estudiantes durante la lección, además de detectar qué temas han producido un mayor interés en el alumnado. También consideran que, en el caso de los seminarios de problemas, puede servir para determinar si hubo que hacer muchas aclaraciones durante el trabajo en grupo de los estudiantes, lo que podría indicar cierta dificultad del problema o la necesidad de mejorar la claridad del enunciado.

Sin embargo, también opinan que algunos aspectos de los informes podrían mejorarse, por ejemplo la introducción de la identificación de múltiples estudiantes, en lugar de agruparlos como un único ente. Esto permitiría conocer si las interacciones de los estudiantes proceden de un único estudiante, de varios o es generalizada en el grupo. Además, se sugiere que pueda mostrarse información acerca del grado de variación tonal del docente para estudiar la correlación que pueda tener con el nivel de participación de los estudiantes. Se cuestionan si un docente monótono a la hora de hablar propicia una baja participación de los estudiantes.

7. Discusión y conclusiones

En este artículo hemos introducido un sistema innovador de análisis automático de clases mediante el procesamiento de audio, con el objetivo de generar informes detallados sobre la actividad docente. Se ha llevado a una puesta en marcha del sistema en un contexto real y se ha proporcionado una retroalimentación continua al profesorado. Esto ha permitido tener en cuenta sus opiniones a la hora de iterar en el diseño del informe. La validez de estos informes como herramienta de análisis se ha evaluado con opiniones adicionales de los docentes entrevistados al final del proceso.

El sistema ha demostrado una capacidad consistente para recopilar datos de manera precisa y objetiva a partir del audio de las clases. La confiabilidad en la recopilación de datos es esencial para garantizar la validez de los informes generados. Es importante volver a resaltar que los informes se basan en métricas cuidadosamente seleccionadas para evaluar la actividad docente, incluyendo el grado de participación de docentes y estudiantes, el número de intervenciones, la duración de las mismas o la duración de los silencios entre otros. La relevancia de estas métricas se respalda mediante la revisión de la literatura [2, 10], lo que fortalece la validez de los resultados obtenidos.

Otro aspecto clave de la propuesta es que el sistema no está restringido a un único contexto educativo y tiene capacidad para adaptarse a diversas metodologías docentes. La validez de los informes se ve reforzada al considerar la diversidad de metodologías docentes, lo que permite su aplicación desde aulas tradicionales hasta entornos de educación a distancia.

En el corto y medio plazo, nuestro objetivo es hacer crecer esta herramienta mediante la incorporación de nuevas características basadas en el procesamiento del lenguaje natural, el etiquetado automático de segmentos de interés y elementos para el análisis de la evolución del docente a lo largo de un periodo de tiempo. También estudiaremos el impacto que podría tener en el rendimiento de los estudiantes el hecho de que estos informes están disponibles para el profesorado y el alumnado de forma inmediata.

Referencias

- [1] Jeff Archer, Steven Cantrell, Steven L Holtzman, Jillian N Joe, Cynthia M Tocci y Jess Wood. *Better feedback for better teaching: A practical guide to improving classroom observations*. John Wiley & Sons, 2016.
- [2] Indrani Bhattacharya, Michael Foley, Ni Zhang, Tongtao Zhang, Christine Ku, Cameron Mine, Heng Ji, Christoph Riedl, Brooke Foucault Welles y Richard J Radke. A multimodal-sensor-enabled room for unobtrusive group meeting analysis. En *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 347–355, 2018.
- [3] Hervé Bredin, Ruiqing Yin, Juan Manuel Coria, Gregory Gelly, Pavel Korshunov, Marvin Lavachin, Diego Fustes, Hadrien Titeux, Wassim Bouaziz y Marie-Philippe Gill. Pyannote. audio: neural building blocks for speaker diarization. En *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7124–7128. IEEE, 2020.
- [4] Óscar Cánovas y Félix J. García. Analysis of classroom interaction using speaker diarization and discourse features from audio recordings. En *Learning in the Age of Digital and Green Transition*, pp. 67–74, 2023.
- [5] Oscar Canovas, Felix J Garcia-Clemente y Federico Pardo. Ai-driven teacher analytics: Informative insights on classroom activities. En *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pp. 1–8. IEEE, 2023.
- [6] Meghan E Dale, Amanda J Godley, Sarah A Cappello, Patrick J Donnelly, Sidney K D’Mello y Sean P Kelly. Toward the automated analysis of teacher talk in secondary ELA classrooms. *Teaching and Teacher Education*, 110:103584, 2022.
- [7] Sidney K D’Mello, Andrew M Olney, Nathan Blanchard, Borhan Samei, Xiaoyi Sun, Brooke Ward y Sean Kelly. Multimodal capture of teacher-student interactions for automated dialogic analysis in live classrooms. En *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 557–566, 2015.
- [8] Patrick J Donnelly, Nathan Blanchard, Borhan Samei, Andrew M Olney, Xiaoyi Sun, Brooke Ward, Sean Kelly, Martin Nystran y Sidney K D’Mello. Automatic teacher modeling from live classroom audio. En *Proceedings of the 2016 conference on user modeling adaptation and personalization*, pp. 45–53, 2016.
- [9] Anusha James, Yi Han Victoria Chua, Tomasz Maszczyk, Ana Moreno Núñez, Rebecca Bull, Kerry Lee y Justin Dauwels. Automated classification of classroom climate by audio analysis. En *9th International Workshop on Spoken Dialogue System Technology*, pp. 41–49. Springer, 2019.
- [10] Catherine Lai, Jean Carletta y Steve Renals. Modelling participant affect in meetings with turn-taking features. En *Proc. Workshop of Affective Social Speech Signals*, 2013.
- [11] Hang Li, Yu Kang, Wenbiao Ding, Song Yang, Songfan Yang, Gale Yan Huang y Zitao Liu. Multimodal learning for classroom activity detection.

- En *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 9234–9238. IEEE, 2020.
- [12] Owens, Melinda T. and others. Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences of the United States of America*, 114(12):3085–3090, 2017.
- [13] Tae Jin Park, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J. Han, Shinji Watanabe y Shrikanth Narayanan. A review of speaker diarization: Recent advances with deep learning. *Computer Speech and Language*, 72, 2022.
- [14] Danner Schlotterbeck, Pablo Uribe, Roberto Araya, Abelino Jimenez y Daniela Caballero. What classroom audio tells about teaching: a cost-effective approach for detection of teaching practices using spectral audio features. En *LAK21: 11th International Learning Analytics and Knowledge Conference*, pp. 132–140, 2021.
- [15] Hang Su, Borislav Dzodzo, Xixin Wu, Xunying Liu y Helen Meng. Unsupervised Methods for Audio Classification from Lecture Discussion Recordings. En *INTERSPEECH*, pp. 3347–3351, 2019.
- [16] Abhijit Suresh, Tamara Sumner, Jennifer Jacobs, Bill Foland y Wayne Ward. Automating analysis and feedback to improve mathematics teachers' classroom discourse. En *Proceedings of the AAAI conference on artificial intelligence*, volumen 33, pp. 9721–9728, 2019.
- [17] Zuowei Wang, Xingyu Pan, Kevin F Miller y Kai S Cortina. Automatic classification of activities in classroom discourse. *Computers & Education*, 78:115–123, 2014.